Image Plane Interaction Techniques In 3D Immersive Environments

Jeffrey S. Pierce[1], Andrew Forsberg[2], Matthew J. Conway[1], Seung Hong[2], and Robert Zeleznik[2]

User Interface Group
Department of Computer Science
University of Virginia
Charlottesville, VA 22903
{pierce, conway}@cs.virginia.edu

Brown University site of the
NSF Science and Technology Center for
Computer Graphics and Scientific Visualization
Providence, RI 02912
(401) 863-7693
{asf, sph, bcz}@cs.brown.edu

ABSTRACT

This paper presents a set of interaction techniques for use in head-tracked immersive virtual environments. These techniques allow the user to operate on the 2D projections of 3D objects on his image plane in much the same way a user uses a mouse to interact with a 3D scene on a desktop monitor. These techniques can be used for selection, navigation, and manipulation in virtual environments.

1 INTRODUCTION

The 2D image plane for desktop interfaces is the monitor screen that displays the 3D scene. A typical use of this image plane to interact with a 3D scene is to use the mouse cursor to select an object by clicking on its projection on the 2D plane.

We believe that the use of the 2D image plane for interaction in 3D scenes extends beyond desktop interfaces to immersive environments. In this paper, we will demonstrate a group of techniques that use the 2D image plane concept for selection, manipulation, and navigation in virtual environments.

The 3D analog to the desktop 2D plane, the monitor screen, is the surface of the retina. Objects in the environment project onto the retina to form a 2D image. Normally we use depth cues such as occlusion and binocular disparity to construct a mental model of the 3D scene that we use to interact with the scene. However, using the techniques we present in this paper we can use the 2D image directly to interact with the environment.

For example, consider the task of selecting a chair on the other side of the room in a virtual environment. Using one of our techniques, the user selects the chair by positioning his hand in the 3D scene so that the projection of his thumb and index finger on the image plane are positioned directly above and below the projected image of the chair.

In the next section, we will discuss some of the previous work that has been done on interaction techniques for virtual environments We will then present our techniques and discuss some of their advantages and disadvantages, and then close with a discussion of future work.

2 PREVIOUS WORK

Two of the first techniques that were used in virtual environments for selection were physical intersection and laser pointing [1]. The drawback of the former is that only objects within reach can be selected, and the drawback of the latter is that it's difficult to select small objects or objects at a distance. The spotlight technique [2] tried to address this problem using a conic selection area. A problem with spotlight selection is that it can be difficult to select one of a number of closely grouped small objects.

Figure #1: The Head Crusher Technique

The NASA Ames VIEW system [3] demonstrated one of the first uses of the image plane to interact with a 3D scene while immersed. The VIEW system divided up the image plane into views of different spaces within the virtual environment. More recently, Michael Gleicher discussed the use of Through-the-Lens controls [4] to constrain the position of an object on the image plane.

The aperture based selection technique developed by Forsberg et al [5] performs the selection task through a hand held aperture. Although the their work did not explicitly discuss the idea of using the 2D image plane in an immersive environment, their aperture selection technique implicitly makes use of this idea by selecting an object whose projection on the 2D image plane falls within the aperture's projection when viewed from the user's eye-point.

The Smartscene system [6] presented some innovative techniques for directed navigation. However, in their system the user is required to directly touch his desired destination point in order to navigate to it. While there is a mechanism for scaling your size up or, equivalently, scaling the size of the environment down, the user is still limited to navigating only to points that can be reached with a single gesture.

3 THE TECHNIQUES

The techniques we present have their roots in "The Kids in the Hall" [6], a comedy show from the late 1980s. This show featured a sketch with a character who would pretend to "crush people's heads" by positioning his victim's head between the index finger and thumb of his outstretched hand (see Figure #2). In technical terms, this man is operating on 3D objects at a distance by interacting with their 2D projections on his image plane. We extend this idea through several loosely-related interaction techniques. Each of these techniques has the common theme of finding the position of the user's hand on his image plane and using that point as a 2D cursor to cast a picking ray into the scene.

Figure #2: The Kids in the Hall Head Crusher

In the **Head Crusher** technique, the user positions his thumb and forefinger around the desired object in the 2D image as shown in Figure #3. We determine which object is desired by finding the 3D midpoint between the user's forefinger and thumb, projecting that point on the user's image plane, and casting a pick ray into the scene from that point. An advantage of this technique is that the user's fingers provide an orientation that can be used to disambiguate the user's selection when there are a number of candidate objects. As suggested by Forsberg et al [5], the object with the closest matching orientation can be chosen.

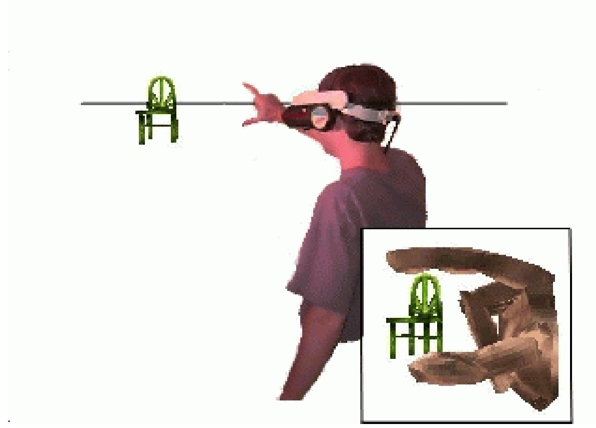

Figure #3: A third person view of the Head Crusher technique. The Inset shows the first person view.

The **Sticky Finger** technique provides an easier gesture when picking very large or close objects by using a single outstretched finger to select objects. The object underneath the user's finger in the 2D image is the object that is selected. In Figure #4, we show the user selecting a distant television using the Sticky Fingers technique. To determine which object the user has selected, we find the current location of the tip of the user's index finger in the scene and then convert this into a position on the user's image plane. We then cast a ray into the scene through that point to determine what object was behind the finger.

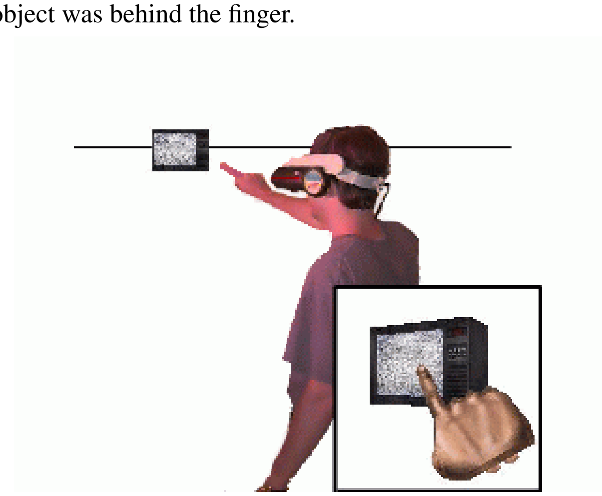

Figure #4: A third person point of view of the Sticky Fingers technique. The inset shows the first person view.

The idea for the **Lifting Palm** interaction technique is borrowed from the famous optical illusion of a man in the background of a photo apparently standing on the palm of a man in the foreground. The user selects an object by flattening his outstretched hand and positioning his palm so that it appears to lie below the desired object on his image plane (see Figure #5). We determine what object the user has selected by finding the current location of his palm in the scene, converting it to a position on the image plane, and then casting a ray from the eye-point through that position.

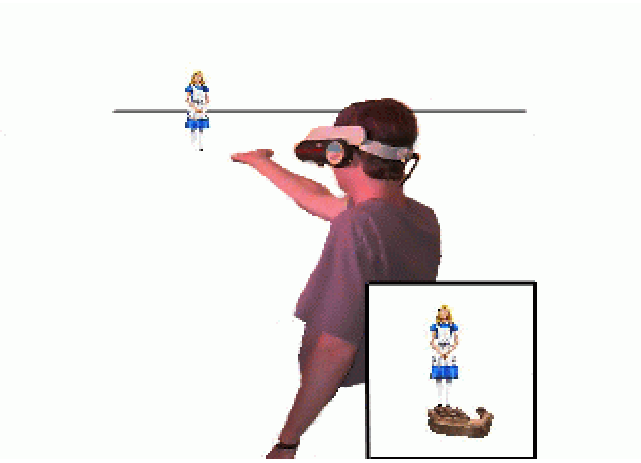

Figure #5: A third person point of view of the Lifting Palm technique. The inset shows the first person view.

Our last technique, **Framing Hands**, uses both hands to select objects. Using this technique, the user's hands are positioned to form the two corners of a frame in the 2D image. This frame can then be positioned to surround the object to be selected (see Figure #6). The implementation for this selection is similar to the Head Crusher's implementation. We determine the location of the user's hands, and then calculate the midpoint between the two in the scene's coordinate system. By translating this point to image plane coordinates, we can again cast a ray into the scene to select an object. This technique can also be used to select a group of objects by selecting all the objects that lie within the frame. This is similar to the 2D rectangle selection technique used in many desktop applications, except that this technique allows the user to use both hands to specify the desired area.

4 MANIPULATION

Once an object has been selected, the user may wish to manipulate the object directly, in which case we need to "break the illusion" and decide where to place the object with respect to the user's fingers.

Perhaps the most straightforward option is to instantaneously translate the desired object to the selection point (e.g. between the user's fingers when using the Head Crusher technique) and to scale the object down so that the object's projection on the image plane remains unchanged.

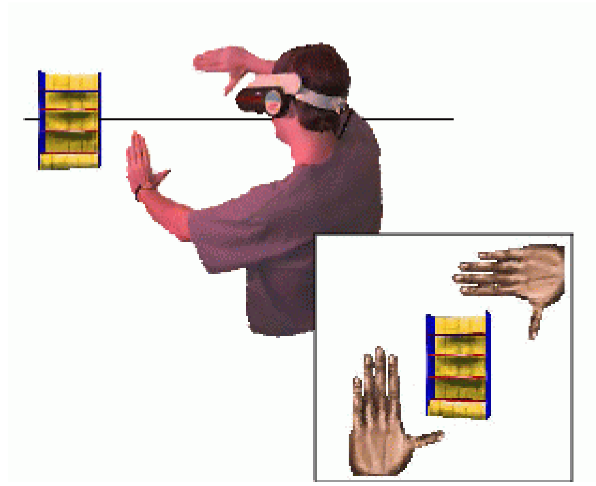

Figure #6: A third person point of view of the Framing Hands technique. The inset shows the first person view.

This leaves the user with a tiny model of the selected object. This object may not be of a useful size for manipulation, so we considered several scaling options that the user could employ. The first option is to scale the object automatically to fill a "convenient" working volume, which we defined as about one foot in diameter for our implementation. The scale takes one second and begins after the object is translated to the selection point. Another option would be to scale the object dynamically based on the distance from the user's hand to his eyepoint. As the user draws his hand in, the object scales up or down to a convenient working size. We also considered the option of leaving the object tiny, but allowing the user to scale the object explicitly by making a two handed gesture indicating the desired size.
Instead of translating the object to the user's hands, we can allow the user to interact with the object at a distance. One way of doing this is to constrain the object to lie somewhere along a line that runs from the user's eyepoint through the selection point. The distance between the object and the user is computed as a function of the distance between the user's hand and his eyepoint (i.e. moving the hand closer moves the object closer in a scaled motion).

Alternatively, we note that the distance can also be determined by using object associations [8]. Because the 2D image plane displays the same image that a monitor on the desktop would, associations can be implemented as they would be in desktop applications. For example, a selected object with a pseudo-gravity association that is on top of a table in the 2D image would be placed on top of the table in the 3D scene.

5 NAVIGATION

Our techniques can also take advantage of the image plane for navigation relative to a selected object. A constraint is placed on the positions of the hand and selected object in the 2D image so that they remain in the same position relative to each other on the image plane [4]. For example, if a distant building is selected using the Head Crusher technique the constraint will keep the building between the user's fingers in the 2D image. Because the position of the selected object is held constant in the 3D scene, this constraint is maintained by positioning the user along the vector from the user's eye-point to the object through the current selection point of the technique being used.

The distance of the user from the object can be held constant to orbit the object. Koller et al [9] have implemented orbital viewing and flying techniques which allow a user to quickly move his viewpoint around an object. However, while their implementation requires the user to move his head to orbit the object, our implementation uses the movement of the user's hands to change his point of view.

Alternatively, the user's distance from the object can vary as a function of the user's hand motion. Currently we are using a linear function that scales changes in the hand's distance from the user to changes in the user's distance from the object. In this implementation, moving the hand halfway toward the eye-point moves the user half the distance to the selected object.

6 DISCUSSION

To use these image plane techniques, a user must be able to interpret what he sees as both a "realistic" 3D environment and a "magical" 2D picture [10], and to move freely between these interpretations. We feel that this transition is not nearly as cognitively taxing as it would first appear. Prior to implementing these techniques, we observed a user in a virtual environment trying to grab a distant object with his hand. The user was unaware of our work and the object he wanted was clearly out of his reach, yet he fully expected that he could grab the object because his hand was on top of it on the image plane. This suggests that he was trying to select the object using its projection on his 2D image plane. The classic "man in palm" photo is another example of this phenomenon. The very fact that this photo is an optical illusion lends credence to our observation that users can and do operate in both realms.

We have done some initial tests with users to determine whether or not they have problems using these techniques. So far, no user has had any trouble understanding how the techniques work. Everyone has been able to select and manipulate objects and to navigate around the scene. Although our informal tests have been positive, we feel the need for more definitive user studies to determine how well these new techniques work in relation to previous techniques like laser pointing and spotlight selection.

There are two problems with these image plane techniques that must be considered. The first is the problem of choosing the left or right eye's image plane for use with these techniques when the scene is rendered in stereo. The importance of this problem can be demonstrated by holding your finger one foot in front of your eyes alternately closing each eye. The position of your finger is different in the 2D image for each eye. If different objects are beneath the finger in each image then one of the two must be chosen. If stereo is used, a number of possible solutions exist. The user can designate their dominant eye prior to using the system, and then have all image plane operations use that eye as the eye-point. The user can then close the non-dominant eye when selecting an object to avoid confusion, the hand used for selection can be rendered in different positions for each eye, or the hand can be rendered only into the image presented to the dominant eye. The simplest solution to this problem, which our current implementation uses, is to render the scene monocularly.

The second problem is that having the user work with his arms constantly extended to his eye-height is very fatiguing. Our techniques address this problem by allowing the user to move the object to a natural working position before interacting with it after the selection task has been completed. If the user is seated, then a desk or table could be used as a support surface when using these techniques. In addition, if we are correct in our belief that it requires less time to select an arbitrary object with this type of selection (versus laser pointing or spotlight selection), then the user will also need his hand extended for a shorter period of time when choosing an object compared to traditional ray casting techniques.

7 FUTURE WORK

All the techniques we have presented are currently being examined in isolation; there is no means to dynamically switch between techniques. Therefore there is room to experiment with how well these techniques work together, and how they can be incorporated into a single system. Mark Mine at UNC and Carlo Sequin at Berkeley are currently working on which techniques work better for certain tasks, and how the user should indicate which technique he wants to use.

The appearance of these interaction techniques to other users in a shared virtual environment is an open question for investigation. For example, if a user selects an object on his image plane and brings it into his natural working environment, a decision must be made about how this appears to an observer standing nearby or at a distance.

We have also considered using this technique in conjunction with portals and mirrors [11]. We define a portal as a static or dynamic view into another scene or world. Portals and mirrors are just 2D image planes that display the projections of 3D objects, so the user should be able to use them to interact with the objects they display using these same 2D image plane techniques.

Finally, we have completed a proof of concept implementation where the user can step between worlds by explicitly changing the 2D image that he perceives. The user holds in his hand a flat square with a texture showing another world. By moving the square to fill his image plane, the user instantly transports himself to that location, so that when he pulls the square away he is looking at the world from the texture (see Figure #7). This technique, and the portals and mirrors technique, both reify the 2D image plane and interact with it as a separate object in the 3D scene.

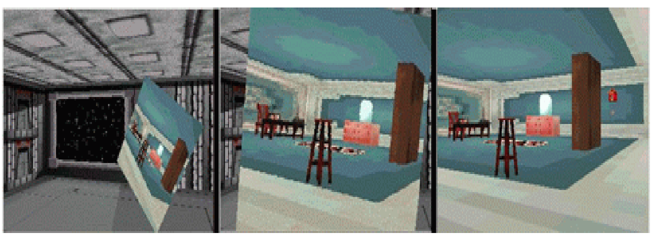

Figure #7: A first person point of view of an image plane being brought up to fill the user's view, a shot of the plane nearly filling the view, and user's new location after the plane has reached the user's eye-point.

8 ACKNOWLEDGEMENTS

We would like to thank Randy Pausch and Andries Van Dam for their help, support, and advice. We also owe thanks to David Staack and Steven Audia, who generated the models used in this paper, Kristen Monkaitis and Chris Sturgill, who provided the hand painted textures, and the User Interface Group at the University of Virginia, who provided Alice, the 3D Prototyping System. A thank you as well to the members of the Brown Graphics group who contributed ideas.

REFERENCES

1. Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassel. A Survey of Design Issues in Spatial Input. *Proceedings of UIST '94*, pages 213-222, November 1994.
2. Jiandong Liang and Mark Green. JDCAD: A Highly Interactive 3D Modeling System. *Computers and Graphics*, v18 n4 pages 499-506, July/August 1994.
3. Scott Fisher, M. McGreevy, J. Humphries, and W. Robinett. Virtual Environment Display System. *1986 Workshop on Interactive 3D Graphics*, pages 77-87, 1986.
4. Michael Gleicher and Andrew Witkin. Through-the-Lens Camera Control. *Proceedings of SIGGRAPH 1992*, pages 331-340, July 1992.
5. Andrew Forsberg, Kenneth Herndon, and Robert Zeleznik. Aperture Based Selection for Immersive Virtual Environments. To be presented at *UIST '96*, November 1996
6. Dan Mapes and Mike Moshell. A Two Handed Interface for Object Manipulation in Virtual Environments. *Presence*, v4 n4, pages 403-416, 1995.
7. The Kids in the Hall. Created by The Kids in the Hall comedy troupe. Prod. Lorne Michaels, CBC, 1988-1994.
8. Richard Bukowski and Carlo Sequin. Object Associations. *1995 Symposium on Interactive 3D Graphics*, pages 131-138., 1995
9. David Koller, Mark Mine, and Scott Hudson. Head-Tracked Orbital Viewing: An Interaction Technique for Immersive Virtual Environments. To be presented at *UIST '96*, November 1996.
10. Randall B. Smith. Experiences With the Alternate Reality Kit: An Example of the Tension Between Literalism and Magic. *CHI+ GI 1987 Proceedings*, pages 311-317, 1987.
11. E. A. Bier, M. Stone, K. Pier, W. Buxton, and T. DeRose. Toolglass and Magic Lenses: The See Through Interface. *Proceedings of SIGGRAPH 1993*, pages 73-80, 1993.