Image Plane Interaction Techniques In 3D Immersive Environments

Jeffrey S. Pierce[1], Andrew Forsberg[2], Matthew J. Conway[1],
Seung Hong[2], Robert Zeleznik[2], Mark R. Mine[3]

University Of Virginia[1], Brown University[2], University Of North Carolina[3]

Abstract

This paper presents a set of interaction techniques for use in head-tracked immersive virtual environments. With these techniques, the user interacts with the 2D projections that 3D objects in the scene make on his image plane. The desktop analog is the use of a mouse to interact with objects in a 3D scene based on their projections on the monitor screen. Participants in an immersive environment can use the techniques we discuss for object selection, object manipulation, and user navigation in virtual environments.

CR Categories and Subject Descriptors: I.3.6 [Computer Graphics]: Methodology and Techniques - Interaction Techniques; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - Virtual Reality.
Additional Keywords: virtual worlds, virtual environments, navigation, selection, manipulation.

1 INTRODUCTION

The 2D image plane for desktop interfaces is the monitor screen that displays the 3D scene. An application projects objects in the scene onto the image plane for display to the user. The user can interact with the objects in the 3D scene using the image plane by positioning the mouse cursor over the projection of the object in the image plane and clicking to select the object.

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

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

Department of Computer Science, University of North Carolina
Chapel Hill, NC 27599-3175
mine@cs.unc.edu

techniques that use the 2D image plane concept for selection, manipulation, and navigation in virtual environments.

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 his image plane are positioned directly above and below the projected image of the chair (see Figure 1). The user does not need information about the actual size or distance of the object to interact with it.

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

Physical intersection and laser pointing are two of the first techniques that researchers used in virtual environments for selection [6]. Physical intersection has the drawback that only objects within reach can be selected. Laser pointing solves this problem, but a user can have difficulty selecting small objects at a distance with laser pointing because small hand motions result in large angular displacements for the laser selection spot when the objects pointed at are far away. The spotlight technique [10] uses a conic selection volume to address this problem. However, the spotlight selection technique requires some method for disambiguating which object the user wants to select when multiple objects fall within the selection volume.
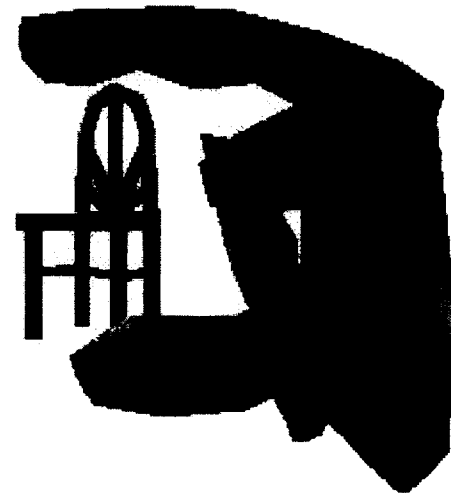

Figure 1: The Head Crusher Technique

A variety of other systems have used image plane techniques. 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 [5] on the desktop to constrain the position of an object on the image plane.

Myron Krueger discussed how the VIDEOPLACE system [9] can use the user's hands to select and manipulate objects in a 3D scene. The user can interact with objects by changing the position of his hands relative to other objects in the image displayed on a large 2D screen (the image plane). For example, the user can select an object by touching it with his hand in the image or move it around by pushing its image with his hand.

The aperture based selection technique developed by Forsberg et al [4] performs the selection task through a hand held aperture. Although 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 [11] includes some innovative techniques for directed navigation. However, this system requires the user to directly touch his desired destination point in order to navigate to it. While there is a mechanism for scaling the user's size up or, equivalently, scaling the size of the environment down, the system limits the user 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" [7], 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 character 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 pick ray into the scene.

Figure 2: A scene from the "Head Crusher" sketch from the comedy show "The Kids in the Hall"

In the **Head Crusher** technique, the user positions his thumb and forefinger around the desired object in the 2D image. Figure 3 shows the user selecting a chair in the scene using this technique. We determine which object is between the user's fingers in the image plane by casting a pick ray into the scene from the user's eye-point through the point midway between the user's forefinger and thumb.
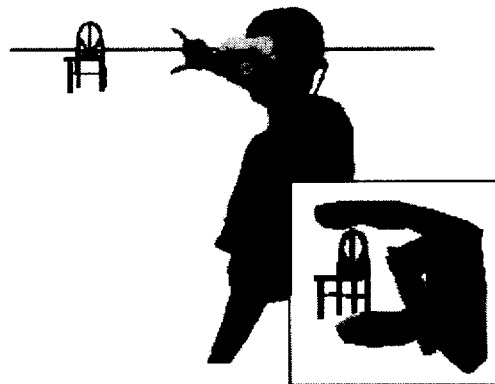

Figure 3: A third person view of the Head Crusher technique. The inset shows the first person view.

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 Finger technique. To determine which object the user has selected, we cast a ray into the scene from the user's eye-point through the location of the tip of the user's index finger in the scene. Objects intersecting this ray are beneath the user's fingertip in his image plane.
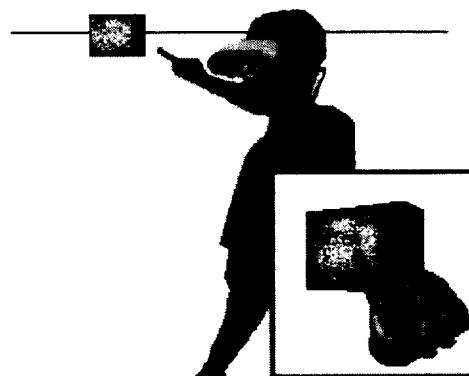

Figure 4: A third person point of view of the Sticky Finger technique. The inset shows the first person view.

We extend the notion of using the image plane for selection with the **Lifting Palm** technique. We borrowed the idea for the Lifting Palm technique 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 which object the user has selected by finding the current

location of his palm in the scene and imposing a slight offset on it. This ensures that we check for the object that lies above the user's palm in the image plane. We then cast a pick ray into the scene from the eye-point through that position.
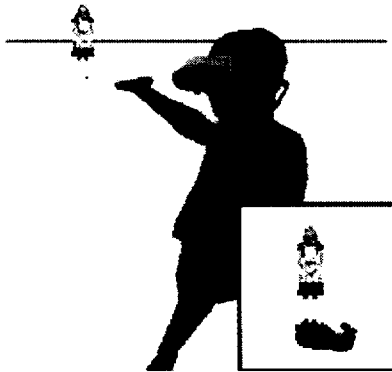

Figure 5: A third person point of view of the Lifting Palm technique. The inset shows the first person view.

Our last technique, **Framing Hands**, uses both hands to select objects. Using this technique, the user positions his hands to form the two corners of a frame in the 2D image. The user then positions this frame 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. We again cast a ray into the scene from the user's eye-point through that midpoint to select an object.
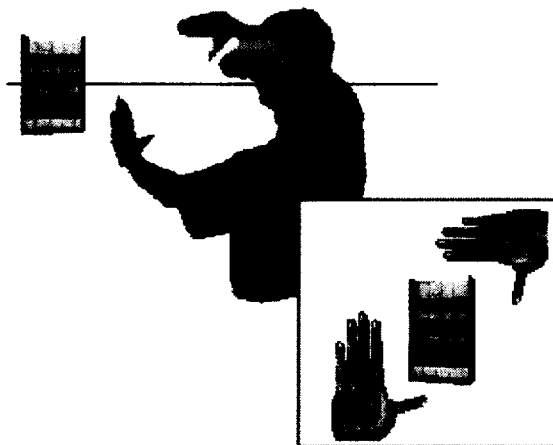

Figure 6: A third person point of view of the Framing Hands technique. The inset shows the first person view.

The user can also use the Framing Hands technique to select a group of objects by selecting all of the objects that lie within the frame formed by the user's hands. 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 simultaneously to specify the desired area. The user can arbitrarily rotate or resize the frame formed by his hands with a motion of his hands. We can draw the frame explicitly on the image plane to provide additional feedback to the user for which objects will be selected.

There are a few general notes about these techniques. First, the system should provide explicit feedback to the user about what object will be selected when these techniques are used. The system can provide this feedback by highlighting or showing the bounding box of the object that is the current candidate for selection. The user can use this feedback to confirm that he has positioned his hand correctly for a desired object before issuing a selection command.

These techniques also provide an orientation that can be used to disambiguate the user's selection when there are a number of candidate objects with identifiable orientations. As suggested by Forsberg et al [4], the object with the closest matching orientation in the user's image plane can be chosen. The user's finger(s) provide this orientation for the Head Crusher, Sticky Finger, and Framing Hands techniques. The normal to the user's palm is the disambiguating orientation for the Lifting Palm technique.

4 MANIPULATION

One of the primary motivations for using any of these techniques is to allow the user to act on objects beyond arm's reach. If the user wishes to perform a close-in operation (e.g. direct manipulation scaling) on a distant object, we need to provide mechanisms that move the object to the user in a non-jarring, sensible manner.

One 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.

This leaves the user with a tiny model of the selected object that might be too small for the user to manipulate conveniently. We considered several scaling options that the user could employ. The first option is to resize the object automatically to fill a "convenient" working volume, which we defined as about one foot in diameter for our implementation. We animate this resize operation over an arbitrary amount of time (one second in our implementation), starting when the object is translated to the selection point.

Another option is to resize the object dynamically based on the distance from the user's hand to his hand's natural working position (for most people this is slightly in front of their waist). As the user moves his hand to this position the object scales to a convenient working size. This has the advantage of reducing arm fatigue because it encourages the user to bring objects to a more natural working position before manipulating them.

Note that both of these options scale the object to a convenient working size instead of its original size. Although the latter might work for small objects, it causes problems if the user has selected a skyscraper on the horizon with a small projection on his image plane.

Finally, we considered the option of leaving the object small and allowing the user to scale the object explicitly by making a two handed gesture indicating the desired size. This gives the user direct control over the size of the object after selection.

Rather than scaling down and translating just the selected object, we can scale down the entire world and translate it so that the selected object is moved to the technique's selection point. This approach has the advantage of bringing the world within reach so that the user can translate the object by directly manipulating the object's position in the scaled world. When the object is released, the world scales back to its original size and translates to its original

41

position.

As an alternative to translating the object or the world 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. We compute the distance between the object and the user as a function of the distance between the user's hand and his eyepoint (e.g. using a linear function, moving the hand half the distance to the user's eyepoint moves the object half the distance to the user).

Another option is to use object associations [2]. 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, the system would place a selected object with a pseudo-gravity association that is on top of a table in the 2D image on top of the table in the 3D scene.

5 NAVIGATION

Our techniques can also use the image plane for navigation relative to a selected object. We place a constraint on the positions of the user's 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 the user selects a distant building with 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, we maintain this constraint 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.

We can hold the distance of the user from the object constant to let the user orbit the object. Koller et al [8] have implemented orbital viewing and flying techniques that 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, we can vary the user's distance from the object 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, if the user moves his hand half the distance to his eye-point he moves half the distance to the selected object.

The user can also use the manipulation technique of scaling down the entire world and translating it for gross navigation. When the user changes the position of his head relative to the selected object and releases the object, his position in the scene will be different when the world scales back up and translates. The size and position of the object on his image plane before the object is released determine his position after release so that the resized object will occupy the same size and position on his image plane.

6 DISCUSSION

To use these image plane techniques, the user must be able to interpret what he sees as both a "realistic" 3D environment and a "magical" 2D picture [12], 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 informal tests with six users to determine whether or not people have problems using these techniques. No user has had any trouble understanding how the techniques work. Every user 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 a few problems with these image plane techniques that must be considered. The first problem is choosing the left or right eye's image plane for use with these techniques when the scene is rendered in **stereo**. Holding your finger six inches in front of your eyes and alternately closing each eye demonstrates the importance of this problem. 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 the system must chose one of the two.

A number of possible solutions exist for systems using stereo. 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 close the non-dominant eye when selecting an object to avoid confusion regarding which object will be selected.

Another solution for a stereo system is to render the hand used for selection in different positions for each eye, or to render the hand only into the image presented to the dominant eye. However, this can cause problems if the user's hand is used for other tasks where correct stereo information is important.

The simplest solution to this problem, which our current implementation uses, is to render the scene monocularly. With this solution only one image plane is available for these techniques.

A second problem is **arm fatigue** from the user working with his arms constantly extended to his eye-height. Our techniques address this problem by allowing the user to move the object to a natural working position after the selection task has been completed. If the user is seated, he can use a desk or table 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.

A final problem is that the user's hands may actually **obscure** the object if the object is far away or very small. To address this problem, the system may render the user's hands translucently so that the object can be seen through the user's hands. We have also considered as an alternate solution replacing the user's hands with a more abstract screen representation (such as cursors) that would occupy less of the user's image plane.

7 FUTURE WORK

We believe that formal usability testing for these techniques is required. Although we have performed an informal evaluation, we need more rigorous testing to determine how the speed and accuracy of these techniques compare to more established techniques like laser pointing and spotlight selection.

We are currently examining all of the techniques we have presented 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 understanding which techniques work better for certain tasks, and how the user should indicate the technique he wants to use.

The appearance of these interaction techniques to other users in a shared virtual environment is another open question for investigation. If the user selects an object on his image plane and brings it into his natural working volume, we must decide how this appears to an observer standing nearby or at a distance. Possibilities include showing the object floating to the user's hand or having it disappearing in a puff of smoke.

We have also considered using this technique in conjunction with portals and mirrors [1]. 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 could use them to interact with the objects they display using these same 2D image plane techniques. Therefore, the user could select an object seen through a portal or reflected in a mirror and bring it into his scene.

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 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 reify the 2D image plane and interact with it as a separate object in the 3D scene.
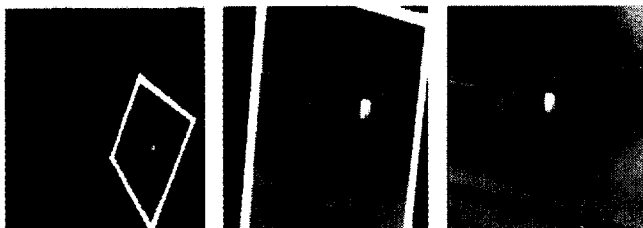

Figure 7: A first person point of view of the user bringing an image plane up to fill his view, a shot of the plane nearly filling his view, and the user's new location after the plane has reached the his eye-point.

8 ACKNOWLEDGEMENTS

REFERENCES

[1] 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.

[2] Richard Bukowski and Carlo Sequin. Object Associations. *1995 Symposium on Interactive 3D Graphics*, pages 131-138., 1995

[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] Andrew Forsberg, Kenneth Herndon, and Robert Zeleznik. Aperture Based Selection For Immersive Virtual Environments. *Proceedings of UIST '96*, pages 95-96, November 1996

[5] Michael Gleicher and Andrew Witkin. Through-The-Lens Camera Control. *Proceedings of SIGGRAPH 1992*, pages 331-340, July 1992.

[6] 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.

[7] The Kids in the Hall. Created by The Kids in the Hall comedy troupe. Prod. Lorne Michaels, CBC, 1988-1994.

[8] David Koller, Mark Mine, and Scott Hudson. Head-Tracked Orbital Viewing: An Interaction Technique For Immersive Virtual Environments. *Proceedings of UIST '96*, pages 81-82, November 1996.

[9] Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. *Proceedings of SIGCHI '85*, pages 35-40, April 1985.

[10] Jiandong Liang and Mark Green. JDCAD: A Highly Interactive 3D Modeling System. *Computers and Graphics*, v18 n4 pages 499-506, July/August 1994.

[11] SmartScene™ is a product of MultiGen Inc. More information on SmartScene™ is available from MultiGen's website at http://www.multigen.com/smart.htm.

[12] 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.