Interactive Viewpoint Control and Three-Dimensional Operations

Michael McKenna

Computer Graphics and Animation Group
The Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

Techniques are discussed for creating a rendered view into a 3D scene, interactively based on the locations and orientations of the observer's head and the display surface. Stereoscopic head-mounted displays (HMDs) demonstrate a simplified, special case of these techniques, because the eyes and monitors move in unison. A largely overlooked class of interactive displays uses the relative positions between the eyes and monitor as input. These displays can be stereo or monoscopic, fixed or mobile, and the rendering process should incorporate the correct perspective distortion, which depends on the locations of the viewpoint(s) and the display monitor.

Three real-time graphics display systems were prototyped and examined: a high-resolution display which corrects the perspective projection based on the location of the observer's eye; the same display, extended to modify the view as the monitor is tilted and swiveled; and a handheld LCD display which can be freely moved and rotated as it displays a view based on the eye and monitor positions.

A simple experiment indicates that tracking the head and providing the appropriate view improves the ability to pick specific 3D locations in space using a 2D display, when compared to a fixed view and a mouse-controlled view.

1. INTRODUCTION

In the everyday world, we continually shift our visual attention from place to place. We rotate the eyes and head, scanning different regions of our field of view. In addition, we move our heads to different locations in space, changing our *viewpoints*. As an observer changes his or her viewpoint, objects at different relative depths appear to move with respect to each other. This effect is known as *motion parallax*, a powerful depth cue [5;7]. Changing one's viewpoint also allows an observer to "look around" objects, and to see the different sides of objects, obtaining multiple perspective views. Perspective and motion parallax are both *monocular* depth cues; the sensation of depth we derive from them requires only one eye, and thus, requires only a 2D display.

Motion parallax can be used to increase the visual correspondence between an operator and a remote or synthetic telerobotic manipulator. An important aspect in the design of displays and controls is creating *isomorphisms* between the local and remote operations [8]. (See Figure 1.) For example, the movement of a control should create a movement of the corresponding manipulator in the same direction, of the same apparent magnitude, on the display. An intelligent display should provide the operator with a view "corrected" for his or her relative position to the display, so that the displayed manipulator movements always appear isomorphic with her or his own movements. An uncorrected view requires that the operator remain exactly centered in front of the display, in order to remain isomorphic. One way to provide the correct view is through the use of a "true" 3D display— i.e. an *autostereoscopic* display, which does not require viewing aids such as glasses [7]. Real-time autostereoscopic displays are problematic, especially concerning bandwidth and computational requirements. For teleoperations, a more difficult problem is the development of the camera required to record the spatial information for an autostereoscopic display. An alternate means of supplying the correct view is to track the locations of the eyes, and then provide the appropriate imagery. For a teleoperator, this requires that the remote camera be servoed to the operator's head movements. In addition, views in which the operator moves off-axis from the center of the monitor require that the displayed image be distorted, either by translating the receptors on the image-focus plane of the camera, providing a sub-image from a wide field-of-view, or by approximating the distortion in hardware/software. The use of head-mounted display systems bypasses the problem of distortion, since the eyes do not move relative to the displays.

The modification to the rendering process to generate off-axis perspective projections is straightforward, using parameters already built into most rendering systems. This can easily be implemented on today's real-time rendering workstations, through the addition of any number of tracking methods. Unfortunately, this technique has been largely overlooked, despite its ease of implementation and perceptual benefits. It is important that the

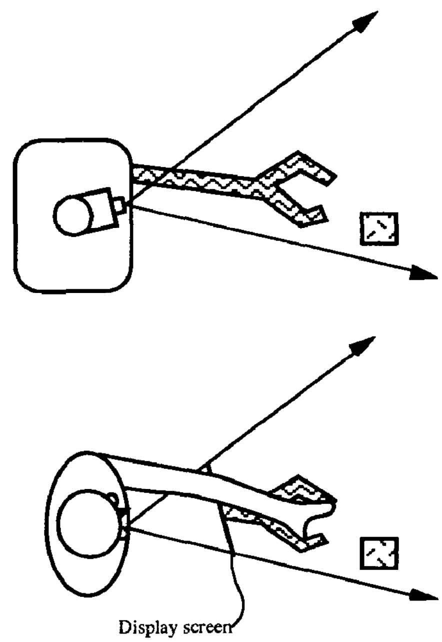


Figure 1: Isomorphisms between a remote robotic manipulator and human operator. Measures that appear equal between the two diagrams are, in fact, equal. The operator cannot put his or her hand through the display, obviously. However, the use of a head-mounted or a flat panel display allows the optical image of the display to share the same space as the operator's hand. Adapted from Sheridan [8].

correct perspective distortion be incorporated in the rendering process. This is not a type of "eye-in-hand" or "eye-on-head" camera control paradigm, in which only the eyepoint and viewing direction are modified [13]. Instead, it is an accurate way of modeling the visual characteristics of a 3D scene.

Prototype display systems were developed by the author to examine the use of tracking techniques to provide an accurate perspective projection, based on the relative positions of the viewer's eyes, the display surface, and the "real," inertial reference frame. Qualitatively, these displays add a great deal of depth perception via motion parallax. The ease of "look around" by moving the head is also a very attractive feature. Providing for a mobile display creates even greater flexibility for "look around" and the exploration of 3D scenes.

A simple experiment was conducted in order to explore the importance of isomorphic imaging on perceiving and interacting with three-dimensional information. Specifically, the experiment tested how many times a subject could move a three-dimensional cursor to a three-dimensional target within a given time period while viewing a 2D display. Different phases of the experiment tested the subject's responses when the view was fixed, when the view could be interactively changed using a mouse, and when the view could be interactively changed by moving the head.

By adding tracked objects in real space which have matching computer representations, important applications can be developed. For example, for medical examination and surgical planning and assist, computer models and scanned data of internal body features can be isomorphically displayed in the "patient space," along with tracked surgical instruments. Similarly, for training and repair, real world objects can be augmented with computer models to guide, instruct, and inform the user.

2. BACKGROUND

Head-mounted displays have been used to interactively view and explore 3D data and scenes for a number of years, recently gaining more popularity [3;11]. The head is tracked, and imagery is generated appropriate for the viewing location and direction. Boom-mounted displays provide similar functionality, allowing for more-massive, high-resolution displays and greater ease of use in certain situations [6].

A different approach was taken by Fisher, who used a monitor fixed in place, allowing the eyes to move relative to the display. Videodisc technology was used to store and playback multiple images of a scene, from different viewpoints. The observer's head was tracked and the appropriate image for that viewpoint location was displayed on a CRT display, creating what Fisher termed *viewpoint dependent imaging* [2].

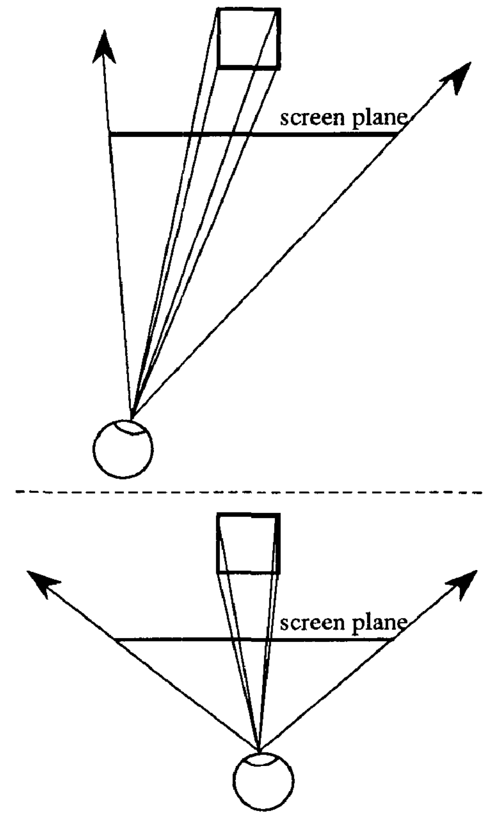
About the same time, a similar system was demonstrated by Diamond, *et al.*, using real-time image generation. Wire-frame rendering was used to generate the perspective projection appropriate for the observer's eyepoint, tracked by a light bulb on the head using a video camera. The authors described the effect of their monoscopic system as "dynamic parallax" [1].

The above technique was extended by Suetens, *et al.*, to provide a stereoscopic image, using electro-optical shutter glasses. A Polhemus sensor was used to track the head, and a stereoscopic wire-frame rendering was generated in real-time [10].

Venolia and Williams created a similar system, which provided for real-time shaded stereoscopic imagery. In order to provide more complex imagery than could be generated in real-time, they employed a "viewpoint array" similar to Fisher's approach. The precomputed images were stored in memory and were displayed based on the observer's horizontal location [12].

This paper provides more details than the above references on the transformations used to generate viewpoint dependent images. It also extends this technique to allow for a mobile display surface. By tracking both the head and monitor, greater flexibility is achieved in the exploration of 3D information, while retaining an isomorphic correspondence between the synthetic space and the real, laboratory space.

Figure 2: The perspectives and sizes of the 2D projections of 3D objects change as the viewpoint moves.



3. FIXED-DISPLAY MONOCULAR SYSTEM

Figure 2 shows an example of how the perspective projection of a 3D object is modified as the view changes. Points which lie at the same depth as the screen are the only ones which do not "move" relative to the screen as the viewpoint changes. Figure 3 depicts a stereoscopic, viewpoint dependent display. The display screen acts like a "window" into the three-dimensional space, cutting off the view of objects which lie outside the current viewing volume. Objects "behind" the screen are cut off just as we expect a real window to obscure objects. Objects in front of the screen and outside the viewing volume are also clipped. However, this is not a phenomenon we are familiar with from our everyday experiences. The "closer" objects are seemingly obscured by the screen, "further" back. This is often called a "window violation" and can significantly disrupt the depth perception of the scene, whether using a stereoscopic or monoscopic display.

To generate a viewpoint dependent image, a normal perspective rendering takes place, using a "window" onto the view-plane, which is off center from the vector which passes through the eyepoint and is normal to the display surface. Figure 4 shows an example viewing setup. The *window center* rendering parameter is used to shift the area to be rendered away from the view normal [4;9].

The monitor's and the observer's locations and dimensions are tracked and located in the rendering "world-space" with the 3D objects. The eye location is established as a constant translational offset within the head tracking coordinate frame. A coordinate frame is established for the monitor, which has its origin at the center of the display surface. Matters are simplified if the coordinate axes are aligned with the display normal and the "vertical"

Figure 3: An off-axis view onto a stereoscopic, viewpoint dependent display. The screen acts as a "window" into the space— clipping objects both in the foreground and background.

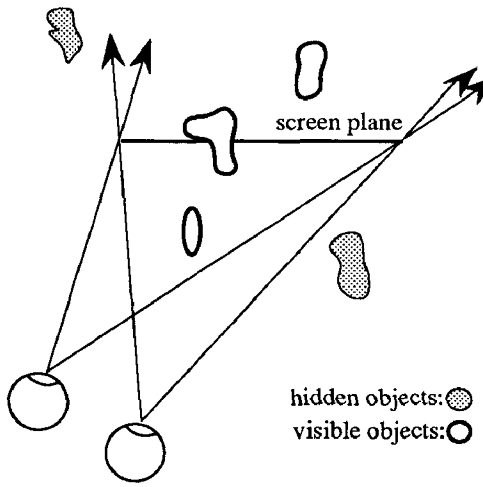
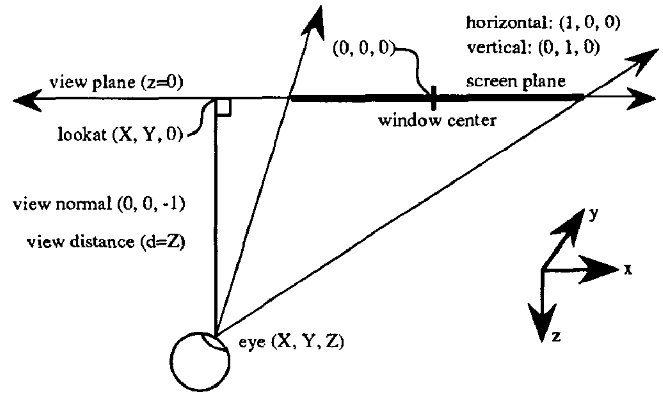


Figure 4: Shifting the "window center" based on the position of the eye generates the appropriate perspective for that viewpoint. The "window center" parameter is used in the rendering pipeline to control a shear transformation, which aligns the center-line of the viewing pyramid with the *z*-axis, in the coordinate system shown here.



and "horizontal" directions, such as the coordinate frame depicted in Figure 4.

The viewing parameters are set as follows: the *eyepoint* is set to the tracked location of the eye, in world space; the *view normal* is set to the "inwards" monitor normal, rotated (and not translated) into world space; the *view up* is set to the "vertical" monitor vector, rotated into world space; the *window half-size* is set to one half of the monitor's actual size; the *view distance* is set to the distance of the eye from the monitor plane, easily attainable by transforming the world-space location of the eye into the monitor's coordinate frame, and using the "height" of the eye, along the display normal ($-normal \bullet eye$); and the *window center* is set to offset the eye's position relative to the display surface's center: ($-horiz \bullet eye$, $-vertical \bullet eye$). These calculations assume that the display surface is planar.

This system was implemented using a Hewlett Packard Model 835 UNIX workstation, with a "Turbo-SRX" real-time polygonal rendering system (performance approximately 12 MIPS CPU, 38,000 shaded triangles per second). A Polhemus sensor was used to track the head. The display surface is fairly large (13" x 11"), with a resolution of 1280x1024.

This is the display system used in the experiment described in Section 6. The system has been used to view 3D objects and animations, qualitatively enhancing 3D perception significantly.

4. MOBILE DISPLAY MONOCULAR SYSTEM

By tracking the position and orientation of the display monitor, we can accommodate changes in its location in the rendering process, so that isomorphism is retained between the imagery and the real-world. The monitor can be moved to attain a better view of the data, or simply shifted to a more comfortable viewing position, without losing the correspondence to the real world coordinates.

The fixed-display method is extended simply by tracking the monitor, and adding the appropriate transformations. A monitor coordinate frame is established as above, only in this case, the monitor frame is a "child" of the display's tracking device coordinate frame, rotating to the normalized monitor space, and translating to the display center.

Two mobile display systems were implemented. The first used the high-resolution HP display, allowing it to tilt and swivel. The display could be translated as well, but it is quite bulky. The Polhemus sensor was mounted on a "boom," away from the EM field of the CRT. It is an important issue to mount the sensor as close as possible to the monitor's center, however, since error and noise in the orientation sensing will be amplified by distance. Movement of the monitor proved useful for adjusting the view, and for exploring the data without losing the correspondence between object space and real space. The display was quite "jittery," unfortunately, due to tracking noise. However, a mode can be employed to deactivate monitor tracking when it is not being moved, to reduce the overall noise. Ideally, a low-noise tracking system would be employed, such as measuring the joint angles in the monitor base.

The second mobile display used a small (2.5"x1.8"), hand-held LCD screen, tracked by a Polhemus, which could be freely moved in space. This system was interesting due to its high mobility— the user could quickly explore 3D data, from many different positions and orientations. The small screen is certainly limiting, but the results indicate that larger screens are worth exploring in this context.

5. STEREOSCOPIC SYSTEM

The extension of the above systems to include stereo is very simple. The second eyepoint is located in world-space in the same manner as the first eye, with a different translational shift from the tracked point (e.g. the polhemus sensor). A second rendering is generated from the second viewpoint, and the left and right eye images are displayed in the appropriate manner for the type of stereoscopic display used.

A tracking device should be used which detects orientation, as well as position, so that the two eyes are accurately located in space. In addition, the "roll" of the head can be detected, as it tilts towards the sides, and the stereo imagery is automatically offset in the appropriate direction. This can be especially important when the display is mobile, since it may take on unusual viewing configurations. The stereoscopic display must be able to support these types of rotations— for example, some polarized systems use linear polarization, which will not allow "rolls."

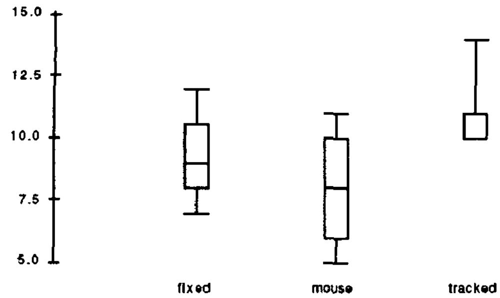
Due to a lack of equipment, we have not, as yet, experimented with a non-HMD stereoscopic display.

6. EXPERIMENT

An informal experiment was conducted to test the effect of viewpoint dependent control on the speed required to manually locate a three dimensional target location. The fixed-monitor, moving viewpoint system was used, as described in Section 3. A second Polhemus sensor was used to track the hand location.

The experiment progresses as follows: A red cube, 2 cm per side (in modeling space and "real" space), appears on the display to act as the target. A blue cube, also 2 cm per side, is displayed, and acts as a cursor, tracking the motions of the hand. The cursor, in the depicted 3D space, moves with the same magnitude and directions as the tracked hand, simply offset by a translation. The task is to align the cursor cube to the target cube (translation only,

Figure 5: The experiment results from the "expert" subjects. The different phases of the experiments are shown across the plot on the *x* axis, and the number of successful target matches is shown on the *y* axis. The mean score is indicated by the central horizontal bar. The line boxes, partially overlapping the grey boxes, indicate the median 25%-75% range of the scores.



no orientation) within a given distance tolerance (1 cm). Once aligned, the target moves to a new random location within the workspace. The subject is instructed to reach the target as many times as he or she can, within the given, fixed time limit.

There are three phases of the experiment: one in which the view is fixed and unchanging, one in which the viewpoint can be moved using a mouse, and one in which the viewpoint is directly controlled by head movements.

Eleven subjects were run through the experiment, four novices and seven experts (subjects familiar with real-time rendering and tracking systems). Figure 5 shows the data from the expert subjects. The novice subjects had the lowest scores, and their results were more widely varying than the experts. In general, performance did increase under viewpoint dependent control, although not dramatically. Use of the mouse generally decreased the score.

Qualitatively, the subjects preferred the viewpoint dependent control, especially as compared to the mouse control, which most found confusing. Some subjects considered the "jitter" in the view, due to the noise from the polhemus tracker, to be distracting; others thought it helped give a better sense of the depth, due to the small amount of resulting motion parallax. This effect could be tested experimentally.

7. DISCUSSION

Providing renderings based on the true viewing parameters of the observer and display has proven to enhance the 3D perception of real-time graphics, in our applications and experiments. Qualitatively, these displays significantly enhanced depth perception via motion parallax, and the ability to "look around" objects and explore the 3D scene, using intuitive motions. These displays generated significant interest and excitement in the lab.

The mobile LCD prototype display is too small to be of use for many applications, but it demonstrates very intriguing viewing qualities. The objects displayed on it are convincingly 3D, not so much in that they "look" 3D, but rather, in that the 3D nature of the data is so easy to explore.

There are interesting differences between these displays and HMDs. These displays are particularly non-intrusive and non-disorienting, since most of the eyes' FOV remains within the real world, and visual jitter does not, therefore, strongly conflict with the vestibular system. Higher effective resolutions are achieved, since the pixels occupy smaller visual angles.

Tracking noise is currently a problem in these prototypes, especially in the mobile-monitor systems. Tracking systems are available which generate significantly lower noise than Polhemus trackers. In particular, articulated arms could be used to measure monitor positions with high accuracy and low noise.

The experiment helped confirm the utility of viewpoint dependent imaging in 3D picking operations. Further experiments should be designed in which a more complete understanding of the 3D scene is required, perhaps adding orientation criteria and more complex environments. In this experiment, the task seemed too simple and quick to execute, in that the subjects would not take the extra time to obtain multiple views unless it was required. An experiment which "rewards" visual exploration would be more appropriate to investigate the perceptual benefits derived from interactive display techniques.

8. ACKNOWLEDGMENTS

This research was initiated through the class *Telerobotics and Human Supervisory Control*, taught by Thomas Sheridan at MIT. The rendering software was written by David Chen. The Polhemus interface code is by David Sturman. I also wish to thank my advisor, David Zeltzer. Thanks to everyone who took the experiment, and to everyone in the lab who helped out. This work was supported in part by grants from NHK (Japan Broadcasting, Co.), and equipment grants from Hewlett-Packard and Apple Computer.

9. BIBLIOGRAPHY

1. Diamond, R., A. Wynn, K. Thomsen and J. Turner. Three-Dimensional Perception for One-Eyed Guys. *Computational Crystallography*. Oxford, Clarendon Press (1982).
2. Fisher, S. S. *Viewpoint Dependent Imaging: An Interactive Stereoscopic Display*. Master's Thesis, Massachusetts Institute of Technology. (1981). Also see: *Proc. SPIE- Display of Three-Dimensional Data* (Bellingham, WA, 1982). Vol. 367.
3. Fisher, S. S., M. McGreevy, J. Humphries and W. Robinett. Virtual Environment Display System. *Proc. 1986 ACM Workshop on Interactive Graphics* (Chapel Hill, NC, October, 1986), 77-87.
4. Foley, J. D., A. van Dam, S. K. Feiner and J. F. Hughes. *Computer Graphics: Principles and Practice*. Reading, MA, Addison-Wesley (1990).
5. Goldstein, E. B. *Sensation and Perception* (3rd edition). Belmont, CA, Wadsworth Publishing (1989).
6. McDowall, I. E., M. Bolas, S. Pieper, S. S. Fisher and J. Humphries. Implementation and Integration of a Counterbalanced CRT-Based Stereoscopic Display for Interactive Viewpoint Control in Virtual Environment Applications. *Proc. SPIE Stereoscopic Displays and Applications* (San Jose, 1990) (1990).
7. Okoshi, T. *Three Dimensional Imaging Systems*. Academic Press (1976).
8. Sheridan, T.B. *Telerobotics and Human Supervisory Control: Cooperative Action by People and Computers*. MIT Press. (In press).
9. Smith, A. R. The Viewing Transformation. Technical Memo No. 84. Computer Division, Lucasfilm, Ltd. (May 4, 1984).
10. Suetens, P., D. Vandermeulen, A. Oosterlinck, J. Gybels and G. Marchal. A 3-D Display System with Stereoscopic, Movement Parallax and Real-time Rotation Capabilities. *Proc. SPIE-Medical Imaging II: Image Data Management and Display* (Part B) (Newport Beach, CA, January, 1988). Vol. 914, 855–861.
11. Sutherland, I. E. A Head-Mounted Three-Dimensional Display. *Proc. the Fall Joint Computer Conference* (1968), 765-776.
12. Venolia, D. and L. Williams. Virtual Integral Holography. *Proc. SPIE- Extracting Meaning from Complex Data: Processing, Display, Interaction* (Santa Clara, CA, February, 1990), 99–105.
13. Ware, C. and S. Osborne. Exploration and Virtual Camera Control in Virtual Three Dimensional Environments. *Proc. 1990 Symposium on Interactive Graphics* (Snowbird, Utah, 1990). In *Computer Graphics* 24, 2 (1990), 175-183.