367-06

Viewpoint Dependent Imaging: An Interactive Stereoscopic Display

Scott Fisher

Architecture Machine Group, MIT
105 Massachusetts Ave., Cambridge, Massachusetts 02139

Abstract

Design and implementation of a viewpoint dependent imaging system is described. The resultant display is an interactive, lifesize, stereoscopic image that becomes a window into a three dimensional visual environment. As the user physically changes his viewpoint of the represented data in relation to the display surface, the image is continuously updated. The changing viewpoints are retrieved from a comprehensive, stereoscopic image array stored on computer controlled optical videodisc and fluidly presented in coordination with the viewer's movements as detected by a body-tracking device. This imaging system is an attempt to more closely represent an observers interactive perceptual experience of the visual world by presenting sensory information cues not offered by traditional media technologies: binocular parallax, motion parallax, and motion perspective. Unlike holographic imaging, this display requires relatively low bandwidth.

Introduction

The design problem undertaken here is an attempt to more closely represent an observers interactive perceptual experience of the visual world by presenting additional sensory information in response to the visceral process of a user's interaction. The goal is to present a virtual display environment in which the user is immersed. In this display, necessary information relative to a user's point of observation is generated by several cues for visual depth perception and is presented to the viewer by means of state of the art display technology. Although similar to the familiar vocabulary of film and video camera movements, the emphasis of this display is on user control as opposed to vicarious, directed observation. These cues allow the viewer to virtually explore the image space as in a real environment - combining several disparate images into a coherent experience of that space. (Figures 1 and 2).

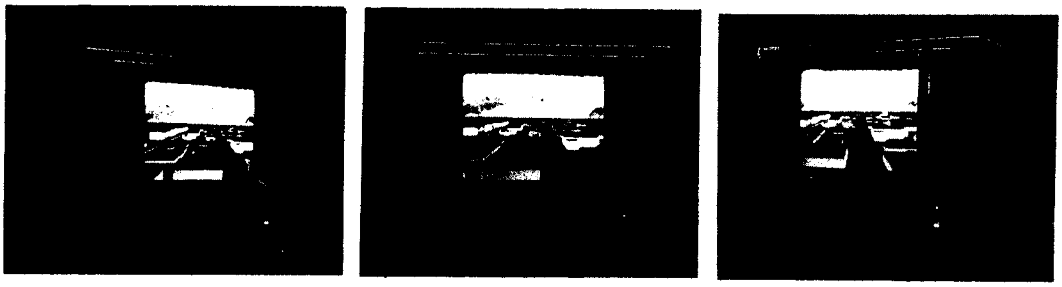

Figure 1. Correspondent viewpoints presented in response to user translation parallel to the display surface.

Figure 2. Typical range of user movement in X.

Fisher, S. "Viewpoint Dependent Imaging: An interactive stereoscopic display." Processing and Display of Three-Dimensional Data, J.J.Pearson, Editor, Proc. SPIE 367, 1982

A first priority is to define the viewer's position in coordinates relative to the plane of the display screen:

X = position along the horizontal axis parallel to the display surface.
Y = position along the vertical axis parallel to the display surface.
Z = position along the axis perpendicular to the display surface.

Change in position information for an active observer is described as translation along these axes relative to the screen and results in three important visual depth cues:

Translation parallel to the plane of the screen in the X and Y axes is called motion parallax. Equivalent to 'tracking' movements in cinematography (not to 'panning' moves), motion parallax is revealed as change in relative position of far and near objects in a scene. The amount and direction of change is described as horizontal or vertical parallax and is the strongest visual depth perception indicator over distances up to 10 meters and beyond.

A subset of translation in the X axis yields binocular parallax or binocular disparity. Because our eyes are horizontally displaced at an average of 65 mm, each sees a slightly disparate viewpoint. For any scene, there exist two cones of vision, with an apex to each eye corresponding to the center of projection or center of perspective for that particular viewpoint. Intersection by a plane perpendicular to this axis of projection will yield images which may be optically fused as a stereoscopic image. The availability of two projections of a given scene is particularly important with regard to resolving image ambiguities. Given only one cone of vision, as in traditional 2D displays, there exist an infinite number of possible object positions that the image displayed on the screen could be generated by. For example, a 5 foot high image of a face could be interpreted as either a closeup of a 'normal' sized head or as the head of a giant. With the additional viewpoint in binocular vision, the image is specifically located in the Z axis and the actual size confirmed. Depth information offered by this cue is effective to about 10 meters, beyond which, image disparaties are unresolvable.

Finally, viewer translation in the Z axis yields motion perspective and is equivalent to 'dolly' shots in cinematography (not 'zoom'). The plane of the screen is perpendicular to this axis and the projection size of the imaged scene is a function of the viewer's distance from that plane. As the viewer moves along this axis, object size relationships are formed. The displayed image appears lifesize when viewed from a position that is equal to the product of image magnification and the focal length of the original taking lens. At this position the projected image will subtend the same angle of vision in the viewer's eyes as for the camera taking the original scene. When combined with binocular parallax cues (i.e. a stereoscopic image) the result is 'orthostereoscopic'.

The main task of this display is to record and store sufficient visual information about a given environment so that these three cues are readily accessible. This is done by imaging every possible viewpoint within a given viewing area and storing them as viewpoint arrays.

Display Implementation

Recording viewpoint arrays

Test images are of two basic types: Computer generated imagery and imagery of real objects and environments.

For recording real objects, a mechanical camera track was positioned so that a 16mm movie camera is shuttled along the track in the X,Y, and Z axes. One frame of film is exposed at predetermined intervals until the complete matrix of viewing positions is recorded. (Figure 3.) Over a total distance of 90cm translating in X, (i.e. horizontally parallel to the screen) intervals of 0.1 cm, 1cm, and 2 cm were used depending on the object/camera separation. Closeup shots require smaller intervals. Larger intervals were used for translation in Y and Z. Various shooting algorithms were tried in an attempt to determine the most useful. The density of the array is changed by varying the camera trigger intervals; Greater intervals for a less dense array also result in an increase of apparent translated speed in the final display.

Arrays of computer generated imagery were formed by defining the viewing matrix coordinates within existing 3D data bases and generating one viewpoint frame at a time. Each frame was then recorded on a 35mm animation camera.

When the viewpoint arrays are recorded, the 16mm and 35mm film is edited and transferred to 2inch video tape, from which an optical videodisc master is made. Several identical copies of the disc are then pressed. Use of the videodisc medium offers two important advantages:

1. Storage density is approximately 50,000 still frames per disc side at relatively low cost.
   and
2. All frames can be randomly accessed at a worst case rate of 2 to 4 seconds. Under computer control, multiple discs may be used to reduce formatting and search-time problems.

A disadvantage in this application is that arrays must be stored linearly in contrast to their original 3D matrix positions. Because some information can be accessed quicker than other, a priority of access must be determined. In this display, it is first assumed that viewer motion in the X axis is most important for motion parallax and binocular parallax cues. A short search time, for example, is necessary when moving from a standing to a sitting position.

367-06

Figure 3. Camera track for recording viewpoint arrays.

Playback configuration

This display is essentially an interactive movie that is accessed spatially rather than temporally: 5,000 frames of a normal film run about 3 minutes. Here an image of 5,000 frames offers as many viewpoints. The crux of this display is to match up these viewpoints stored on optical videodisc with the viewer's position of regard. The technologies used to achieve this are as follows:

The observer's position is tracked by a low-frequency magnetic body-tracking device manufactured by Polhemus Navigational Sciences, Inc.. The tracking range from the magnetic field source to the sensor worn on the user's head is approximately a 1 meter radius hemisphere. (Figure 4.) It operates at 40 points per second and reads three position coordinates as well as three degrees of attitude. Values are then stored in the mainframe computer.

A database is created for the information on videodisc to match up the image arrays with these position coordinates by defining start and end points for image array sequences along a specific axis. As the viewer moves, the videodiscs are computer controlled to match the viewer's direction and speed (to 30 frames/sec).

The actual display medium can take several forms. Much of the imagery is shot to be viewed on a standard video monitor because of the small tracking range ability. Ideally the display is used with a large screen video projector in a rear projection format. In both configurations, the ability to present lifesize images is demonstrable.

As described so far, this system yields a '2D' display rich in motion parallax and motion perspective. These cues may provide sufficient information about depth relationships in the scene. But since each eye sees a slightly different viewpoint, image arrays from the X axis are also used to provide important binocular parallax cues. During the videodisc mastering process stereoscopic image arrays are formed by interlacing identical frame sequences taken along the X axis and offset by about 6 frames or 6.5cm. The frame corresponding to the left eye view is placed in the even interlace of the video frame and the offset frame for the right eye is placed in the odd line interlace. (Figure 5.) Viewing of the resulting stereoscopic image is by means of piezoceramic viewing glasses, 'PLZTs', worn by the user along with the body-tracking sensor. (Figure 4.). Each lens of the viewing glasses acts as a light valve and opens and closes in synch with the field rate of the TV frame. As a result, each eye sees only its correct field and a stereoscopic image is perceived.

Image Content

Particular attention has been paid to the image content of scenes shot for the viewpoint dependent imaging display. Content ranges from scenes that are especially rich in response to viewer movement, to environments that are 'binocular specific' and cannot be adequately re-presented by 2D images. Examples of real objects depicted are: Machine parts, circuit boards, complicated vegetation, lifesize faces (for teleconferencing applications), and anatomical sections (for medical applications). Sample computer generated images used are molecular structures, aerial and street views of Aspen, Colorado and graphic representations of body motion over time. Simulated movement in a real environment is also represented by cutting to cycling real-time footage stored on the videodisc whenever the user pauses along an axis of the viewpoint array.

367-06

Figure 4. PLZT viewing glasses and source and sensor (on glasses) for body-tracking.

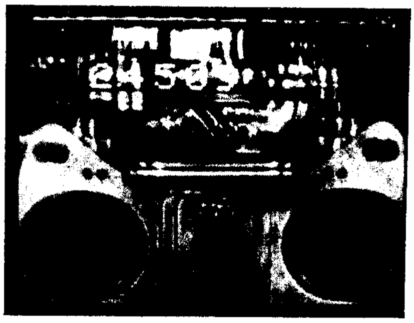

Figure 5. Interlaced 3D image with videodisc frame number.

Display Features

An important feature of this display is the relatively low bandwidth required for transmission. A major obstacle to widespread use of 3D TV previously has been the high bandwidth requirements. For a simple,good resolution, color 3D display configuration, at least twice normal bandwidth is required for use of two channels. In more sophisticated 3D displays such as holograms, every possible viewpoint would be transmitted at a cost of enormously high bandwidth. In this display, by only transmitting the correct viewpoint for a user's position relative to the screen at any given time, the bandwidth remains equal to that of one normal broadcast channel. Tracking the viewer and continuously updating the image eliminates redundant and irrelevant information.

These means for continuously updating a 3D display for an active observer also provide a solution to an insistent problem in other 3D presentations - both still and dynamic. As a viewer moves in front of a typical 3D image, it appears to have reverse motion parallax. The scene appears to pivot around the plane of the screen - making far objects move opposite to the viewer's direction of movement and near objects follow. (Figure 6). Because the viewpoint dependent imaging system continually corrects for viewer translation, correct motion parallax is always perceived. Along similar lines, the presentation of lifesize imagery is usually limited to only one correct viewing position. In this display the ability to quickly update the image maintains the effect as the user moves about. This is a key factor in creating a virtual window through which a user can explore a three dimensional image space.

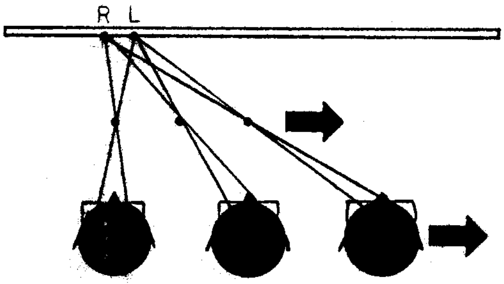

Figure 6. Reverse motion parallax in traditional 3D images. Perceived image foreground follows viewer motion.

Current developments

Current research in the development of this display includes use of a write-once optical videodisc prototype developed by Matsushita E.I.Co.,Ltd. for on-site image array storage as might apply to medical imaging applications. Additionally, effective image interpolation algorithms are under development to improve viewpoint array access, density and storage.

With several tracking and input stations, more than one user can access personalized or viewpoint specific images on a common display surface. In this configuration, the piezoceramic glasses are modified to present different viewpoints to each of several users rather than to each eye of a single user.

Further development of this display includes substitution of a remotely controlled, stereo camera system in place of pre-computed or stored images, enabling real time exploration of inaccessible environments while under direct viewer control.

Acknowledgements

Thanks to Honeywell, Megatek, and Matsushita for the loan of prototype 3D viewing glasses and to Ken Carson of the Architecture Machine Group who provided brilliant programming assistance for implementation of the viewpoint dependent imaging system.

The work reported herein was supported by the Advanced Research Projects Agency of the Department of Defense, under contract number MDA 903-81-C-0097 and was produced at the Architecture Machine Group within the emerging program for Arts and Media Technology at MIT.