Three Dimensional Interaction with Autostereoscopic Displays

1. Research Team

Project Leader:	Prof. Alexander A. Sawchuk, *Electrical Engineering*
Other Faculty:	Prof. Isaac Cohen, *Computer Science* Prof. Chris Kyriakakis, *Electrical Engineering* Prof. Albert Rizzo, *IMSC and Gerontology*
Graduate Students:	Zahir Y. Alpaslan
Industrial Partner(s):	Dynamic Digital Depth (DDD)

2. Statement of Project Goals

The broad goals of this project are to develop new techniques of human interaction with stereo displays, particularly autostereoscopic (AS) displays. Autostereoscopic displays produce a 3D visual sensation to one or more observers without the use of glasses, goggles, helmets, or head-tracking. We are developing new techniques for interactive input and manipulation of three-dimensional data using a motion tracking system combined with an autostereoscopic display. Users interact with the system by means of video cameras that track a light source or a user's hand motions in space. Our principles are applicable to both single-user and multi-user AS displays ranging in size from laptops to large flat-panel systems [1].

3. Project Role in Support of IMSC Strategic Plan

The overall objective is to develop simple techniques for interaction with 3D display systems to enhance and extend the immersive experience for users in a variety of visualization, design, simulation, monitoring, security, entertainment, gaming, tele-conferencing, and command-control scenarios. The research may be extended to interactive real-time high-definition (HD) 3D stereo video, particularly within the context of interactive immersive environments with high quality audio. This project extends and supports other IMSC work on the Media Immersion Environment (MIE), Entertainment, Education and Communications visions as described in Volume One of this report.

4. Discussion of Methodology Used

Autostereoscopic (AS) displays make the 3D viewing experience more pleasant by removing the necessity of using glasses. Table 1 summarizes the two main types of current commercially available AS displays along with their image quality and applications. Lenticular based AS displays [2.3.4.5] have higher brightness and are better for multi-view (more than two displayed images at the same time) and multi-viewer (more than one person can see the effect) applications. However, because the slanted lenticular screen is used to display multiple views the image resolution of each view is reduced horizontally and vertically, making it difficult to read displayed small text. Barrier screen based systems [5.6] have lower brightness and are more appropriate for single viewer and single view situations. However, their resolution is reduced only in the horizontal dimension.

AS Display Type	Resolution	Text	Image quality	Depth Effect	Application	Cost
Lenticular (StereoGraphics, DDD etc.)	Reduced to 1/3 both in horizontal and vertical	Small text unreadable	Image edges are blurred	Movement parallax and image parallax	Multi view and multi viewer	\$4,000 - \$18,000
Barrier (Sharp)	Reduced to half only in horizontal	Readable	Sharp clean edges	Only image parallax	Single view, single viewer	\$3,300

Table 1. Differences between lenticular and barrier technologies in two commercially available AS displays.

The basic block diagram of our system applicable to general AS displays is pictured in Figure 1. Our current focus is on the unshaded blocks of Figure 1. The system consists of acquisition, display, interaction and tracking parts. Starting from the interaction volume on the lower right side, two cameras image a light source cursor held by the user. A tracking algorithm analyzes the images and extracts the position information. This information is combined with graphical models to create a scene with virtual objects. Virtual cameras produce synthesized views of the virtual scene while the graphics card interlaces up to nine views that are sent to the display. Combining these components forms a closed loop as seen in Figure 1. Figure 2 shows details of the two cameras tracking a light source (small flashlight) used as a cursor in the interaction volume.

The acquisition and display parts together are a 3D drawing and image manipulation program based on OpenGL and Visual C++. The acquisition part accepts drawing commands in the form of 3D (x, y, z) coordinates and draws 3D virtual objects in virtual space using these coordinates. The display part produces nine views of the virtual object scene created in the drawing part by shifting a virtual camera horizontally, and interlaces these nine views at the sub-pixel level to create the final interlaced image to be displayed on the autostereoscopic display.

The interaction and tracking parts work like a 3D mouse. To interact with the system a user moves a small hand-held light source in the interaction volume defined by the field of view of two FireWire cameras. A tracking algorithm analyzes this live video frame-by-frame and finds the coordinate of the light source's center in the frames. When the cameras are perpendicular, simple combination of two axes from one camera and the third axis from the other camera gives the 3D coordinates of the light source in camera pixel coordinates. In general, the cameras do not have to be perpendicular.

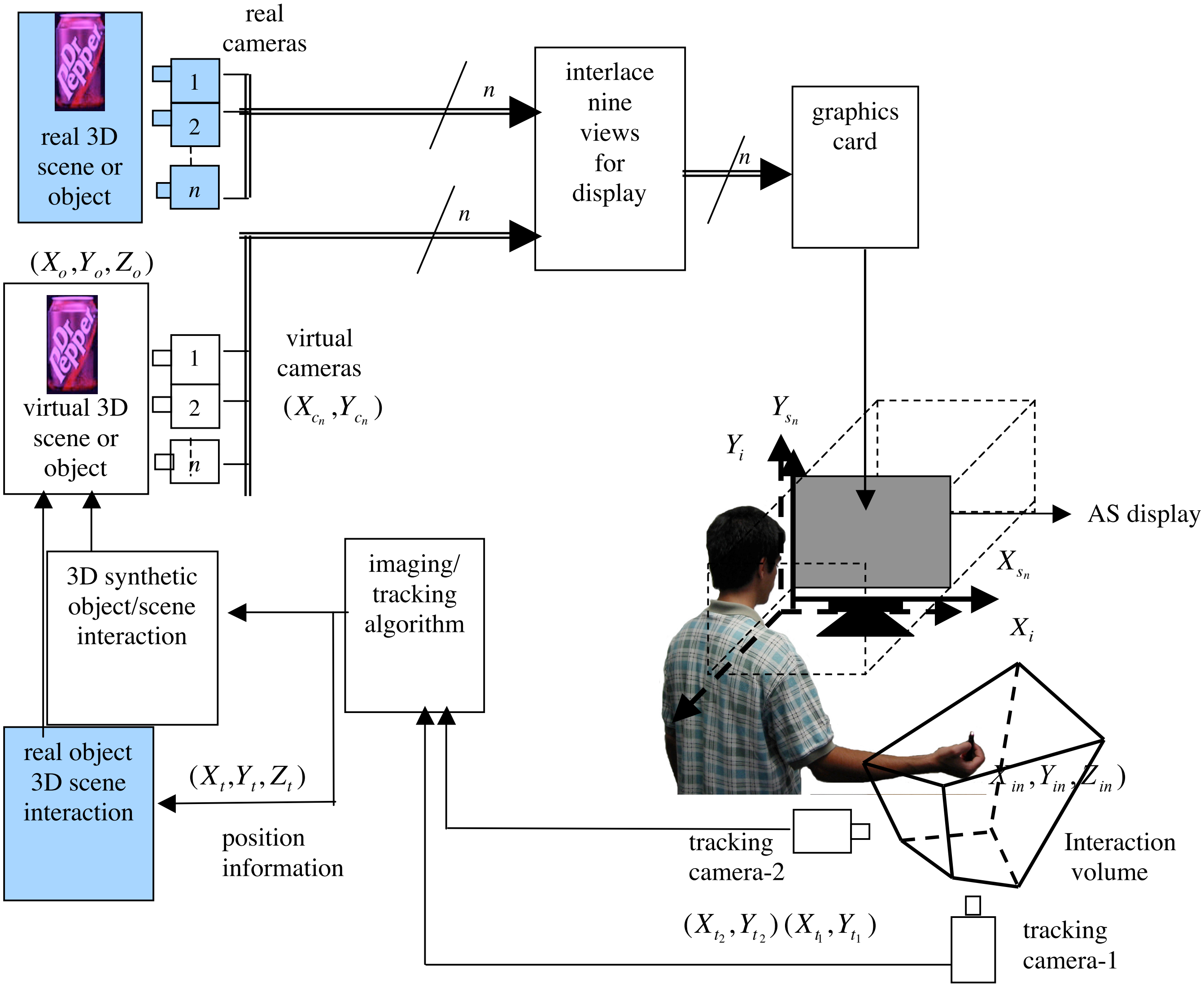


Figure 1. System Overview. Unshaded blocks represent the two components we are currently working on. Shaded blocks are the future system components. (X_o, Y_o, Z_o) are the coordinates of the virtual 3D scene. (X_{c_n}, Y_{c_n}) are the coordinates of the virtual 3D scene as measured by the detector array of camera n. Our system uses from two to nine virtual cameras in order to create the sensation of a 3D display. (X_{s_n}, Y_{s_n}) are the coordinates of view n at the AS display surface. (X_i, Y_i, Z_i) are the coordinates of the displayed points as perceived by a human observer. These coordinates depend on viewing distance, eye separation and many other parameters. (X_{in}, Y_{in}, Z_{in}) are the coordinates within the effective interaction volume. A cursor or user's hand must be in this volume for both cameras to track its position. (X_{t_1}, Y_{t_1}) and (X_{t_2}, Y_{t_2}) are the coordinates of the objects in the interaction volume as measured by the detector array of cameras 1 and 2 respectively. We process these coordinates to create (X_t, Y_t, Z_t), the 3D location of an object in the interaction volume as seen by both cameras.

In our current system we focus on implementing the interaction and display parts in separate physical volumes, although they could be overlapped and placed in register. Thus we do not have to track the user hand movements right in front of the display where the images appear; instead we are free to place the cameras anywhere we want. Another advantage of this implementation is the increased interaction volume. Removing the restriction of interaction with the objects appearing in front of the display plane allows us to interact with objects appearing behind the display plane. In this case interaction volume at least doubles in size. Implementation of not-in-register interaction requires the user to interact with the virtual objects with the help of a cursor that represents the light source in virtual space.

A cursor also avoids the accommodation vergence conflict of the human observer. To perceive a 3D sensation, a user's eyes have to focus on the display plane. However, if the user wants to do in-register interaction without a cursor such as [7.8], he needs to focus on his hand and the display plane at the same time. Since the hand cannot touch the display plane, this forces the user to focus on two different planes, causing eyestrain. One solution to this problem uses active display systems with moving lenticular sheets or lenses in front of the display plane [7.8.9.10.11.12]. While we cannot move the lenticular sheet in our StereoGraphics 202 AS display, we may add it later. Another future application of our system involves the shaded blocks in Figure 1. By adding multiple camera real-time video capture, it is possible to interact with 3D video of live scenes.

5. Short Description of Achievements in Previous Years

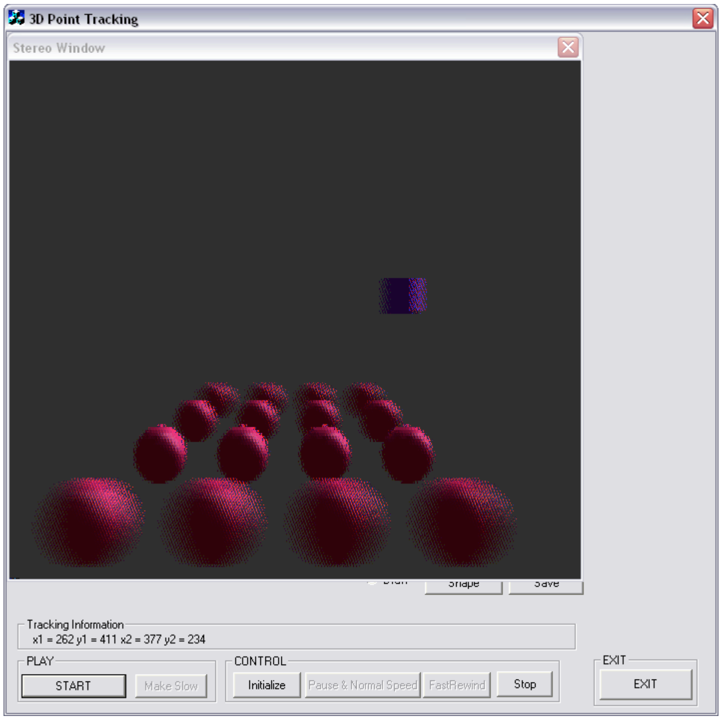
This is a very recent direction for IMSC 3D visualization work. We began with initial experiments in acquiring, displaying and interacting with stereo images by assembling a camera system to acquire still frame stereo images for experimentation, and assembled a simple two-camera stereo video acquisition system. We used two computer projectors fitted with polarizers for experiments with display of the video using polarizing glasses and anaglyph (red-blue) glasses, and acquired an autostereoscopic display system and software needed to drive it. We have been doing software development in preparation for the integration work done over the last year.

5a. Detail of Accomplishments During the Past Year

Our current system tracks the motion of a flashlight cursor running at 15 fps and draws an image with 640x480 resolution in real-time. The current main functions of the system are move, draw and save. The move function translates an object's location in stereo space coordinates(X_i, Y_i, Z_i). Figure 2 shows the picture of a cube moving in AS display space.

The draw function creates a solid object. In order to save processing time we implemented it as a wire-frame drawing tool. The user first creates a wire-frame object made of dots. When the user finishes drawing the wire-frame object he can stop the tracking and make the drawing a solid object by using a basic shape such as a sphere as shown in Figure 3.

We save the figures as text files. Instead of saving the whole image we save the 3D coordinates the program uses for creating and transforming an object. The file size for an image such as the one in Figure 3 is just a few KBs.



x1 = 262 y1 = 411 x2 = 377 y2 = 234

Figure 2. Demonstration of move function. In this screen shot of a monoscopic display, a 4x4 array of spheres serve as depth level indicators in the stereo window. The single cube moves in stereo space according to the user commands. The fuzziness of the displayed spheres comes from the interlacing algorithm. When viewed with a lenticular display the row with the largest spheres appears in front of the display surface, the second row of spheres appears at the display surface and the rows with the smallest spheres appear behind the display surface.

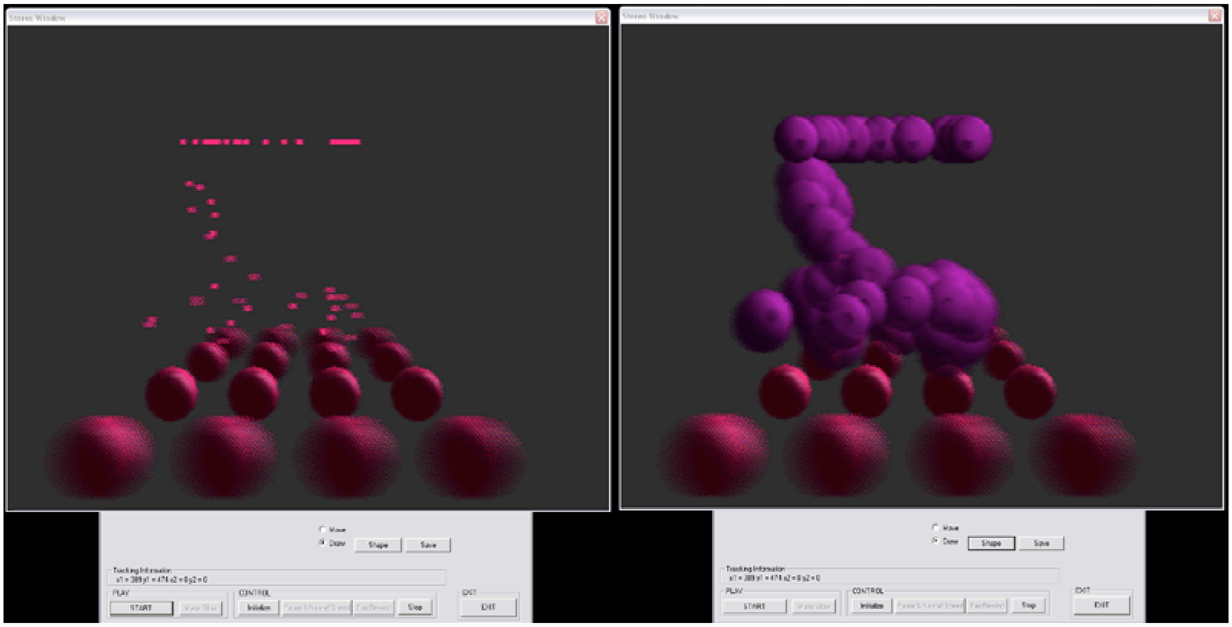


Figure 3. Demonstration of the draw function. The figure on the left shows the wire-frame version of the object drawn by user. When the user finishes the wire-frame version he can connect the dots by using a basic three-dimensional figure such as a sphere.

We also implemented a rotation tool that can rotate the object by clicking buttons on the desktop using a mouse. In the future the rotation and other manipulation functions will be moved to stereo space.

6. Other Relevant Work Being Conducted and How this Project is Different

The earliest reference to interaction with an autostereoscopic display is by DeWitt [6]. He refers to the possibility of interaction but describes no details to our knowledge. In 2000, the MUltimo3D group at the Heinrich-Hertz-Institute in Germany built an autostereoscopic display requiring head-tracking [7.8]. The system consists of a gaze tracker, a head tracker, a hand tracker and an autostereoscopic display for viewing and manipulating objects in 3D. The head tracker makes the images appear in a fixed position so the user has a look-around capability (being able to see around the sides of an image). Gaze tracking activates different applications in the desktop, and the hand tracker navigates and manipulates objects in space rather than a mouse. The Multimo3D hand tracking system is accurate to approximately 2-3 cm. More recently, Berkel at the Philips Labs [14] described the tracking of user's hand and fingers with magnetic fields using sensing electrodes around the edges of a display. They use this information to interact with an autostereoscopic display.

While our approach is related to that of the Multimo3D group, we have different advantages and challenges associated with a multi-zone and multi-view display. Using a two-camera configuration, it is possible to reduce dramatically the computation and processing time needed for tracking light sources used as a 3D cursor. We are not worried about tracking the head for stabilizing the image. Our AS display has nine images inside a viewing zone, therefore it already has a built-in look-around capability. Our interaction algorithm allows a user to interact with images both appearing inside the display and in front of the display. The Multimo3D system allows only interaction with images appearing in front of the display. Since we are using a multi-user AS display, more than one user can see the interaction at the same time and more than one person can interact with the display (by tracking two or more light sources). Connecting two systems via Internet, it is possible to have 3D interaction. For example, one user's drawing can appear and float in space in front of another user at a remote location. The Multimo3D system does not need any handheld objects for help in tracking, therefore gives user more freedom. Our ultimate goal is to have multiple users interact with our AS desktop system at the same time.

7. Plan for the Next Year

In the future we are planning extensive user testing. We will investigate how our system changes the human implementation of tasks using the display, the effectiveness of the user interface and the effects of accommodation and vergence conflict for in-register implementation.

We also will explore system improvements with the addition of 3D binaural, 5.1-channel or 10.2 channel immersive sound. Research shows that 3D audio can enhance the stereoscopic experience [15.16]. We are building our software in a modular structure so that adding new parts is easy to do. With additional programming we can modify this software for use in virtual sculpting and remote virtual interaction. Later our system can be extended to using hand gesturing instead of a cursor. Because of the increase in the speed of USB 2.0 it is also possible to use USB-based cheaper web-cams instead of FireWire cameras to track the user input as in [17]. We view this work as a foundation toward studying the larger problem of 3D operating system and 3D computers.

8. Expected Milestones and Deliverables

We expect to have a real-time 30 fps version of the tracking software fully integrated with a lenticular display system. We will optimize it using OpenGL and other graphics packages such as DirectX. We expect performance improvements as we execute more of the rendering commands on the PC display card. We will complete a mathematical description of the 3D image acquisition and display process for multiple cameras in various geometries and utilize these models in a revised software system. We will explore applications of the system in areas such as network security, command and control, and begin preliminary work on integration with spatial audio.

9. Member Company Benefits

Companies whose business is in visualization, design, simulation, monitoring, security, entertainment, gaming, tele-conferencing, and command-control systems will benefit from the results of this research. The addition of desktop autostereoscopic displays having an interaction capability that does not require clumsy goggles or head-mounted displays to existing and future immersive environments will be of great benefit.

10. References

[1] Z.Y. Alpaslan and A.A. Sawchuk, "Three-Dimensional Interaction with Autostereoscopic Displays," *Proc. Stereoscopic Displays and Virtual Reality Systems XI Symposium*, Proc. SPIE, vol. 5291, San Jose, CA, 2004.

[2] StereoGraphics Corporation, www.stereographics.com

[3] StereoGraphics, “The SynthaGram Handbook”, http//www.stereographics.com/products/synthagram/The%20SynthaGram%20Handbook%20v71.pdf

[4] L. Lipton, M. Feldman, “A new autostereoscopic display technology The SynthaGram™”, *Proceedings of SPIE Vol. 4660, Stereoscopic Displays and Virtual Reality Systems IX*, 2002.

[5] DDD Group Plc., www.ddd.com

[6] Sharp Systems of America, http://www.sharp3d.com/

[7] J. Liu, S. Pastoor, K. Seifert, J. Hurtienne, “Three dimensional PC toward novel forms of human-computer interaction”, *Three-Dimensional Video and Display Devices and Systems SPIE CR76*, 5-8 Nov. 2000 Boston, MA USA.

[8] Heinrich-Hertz-Institute, mUltimo3D group website http://imwww.hhi.de/blick/

[9] H. Kakeya, “Autostereoscopic 3-D Workbench," www2.crl.go.jp/jt/jt321/kake/auto3d.html, Multimodal Communication Section, Communications Research Laboratory, MPHPT, Japan.

[10] H. Kakeya, Y. Arakawa, “Autostereoscopic display with real-image virtual screen and light filters”, *Proceedings of SPIE Vol. 4660, Stereoscopic Displays and Virtual Reality Systems IX*, 2002.
[11] S. Yoshida, S. Miyazaki, T. Hoshino, T. Ozeki, J. Hasegawa, T. Yasuda, S. Yokoi, ”A technique for precise depth representation in stereoscopic display” *The Visual Computer*, Vol. 17, No. 1, 2001.
[12] C. F. Neveu, L. W. Stark, “The Virtual Lens”, *Presence*, Vol. 7, No. 4, pp 370-381 August 1998.
[13] T. DeWitt, “Visual Music Searching for an Aesthetic”, *Leonardo*, Vol. 20, No. 2, pp. 15-122, 1987.
[14] C. van Berkel, “Touchless Display Interaction”, *SID 02 DIGEST,* 2002.
[15] J. Freeman, J. Lessiter, “Hear There & Everywhere The Effects of Multi-channel Audio on Presence”, *ICAD 2001 - The Seventh International Conference on Audio Display*, Helsinki University of Technology, July 29 - August 1, 2001.
[16] J. A. Rupkalvis, “Human considerations in stereoscopic displays”, *Proceedings of SPIE Vol. 4297, Stereoscopic Displays and Virtual Reality Systems VIII*, 2001.
[17] M. Andiel, S. Hentschke, T. Elle, E. Fuchs, “Eye-Tracking for Autostereoscopic Displays using Web Cams”, *Proceedings of SPIE Vol. 4660, Stereoscopic Displays and Virtual Reality Systems IX*, 2002.