Stereoscopic Video Acquisition, Display, Transmission and Interaction

1. Research Team

Project Leader: Prof. Alexander A. Sawchuk, *Electrical Engineering*

Other Faculty: Prof. Isaac Cohen, *Computer Science*
Prof. Chris Kyriakakis, *Electrical Engineering*
Prof. Skip Rizzo, *IMSC and Gerontology*

Graduate Students: Zahir Alpaslan

Industrial Partner(s): Dynamic Digital Depth (DDD)

2. Statement of Project Goals

The broad goals of this project are the development and application of high-resolution 3D stereo video and display technology and its integration into various IMSC vision and application research projects. We investigate stereo video acquisition, display and transmission capabilities and evaluate its effectiveness in immersive applications. A second goal is to investigate human interaction with stereo displays, particularly desktop autostereoscopic (no-glasses) displays. The overall objective is to enhance and expand the immersive experience for users in a variety of variety of entertainment, gaming, simulation, tele-conferencing, social gathering and performance scenarios using stereo display techniques.

3. Project Role in Support of IMSC Strategic Plan

Video images have provided a virtual view of distant times and locations for over seventy years. While video technology has matured from gray-scale to big-screen color and digitally processed imagery, there are many challenges remaining to attain the full potential of very high resolution (better than broadcast) 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), Remote Media Immersion (RMI) and new projects such as Distributed Immersive Performance (DIP) described in Vol. 1 of this report.

4. Discussion of Methodology Used

Acquisition - We investigate several increasingly sophisticated techniques for the acquisition of stereo video information [1, 2]. Here *camera calibration* is critical – the spatial registration, pan, focus, and zoom of the two stereo cameras must remain identical or the right-left images to each eye will not match [3]. The two cameras are attached to a rigid optical mount ensuring stable spatial registration and panning for live stereo video capture. A more difficult problem is to match the camera zoom and focus tracking. An uncalibrated zoom could produce a type of motion sickness to users. Similarly, if the focus changes do not track, the image in one eye will appear blurred. We will measure these factors for each camera and determine appropriate correction techniques. For minor mismatches, a lookup table correction to the zoom and focus control signals may be adequate. For extreme mismatch, we will develop image registration or machine vision algorithms as needed. A second important factor is *vergence*, the angle between lines drawn from the two eyes of an observer to a single point viewed in a scene (effectively the toe-in angle between cameras). For distant objects, the angle is essentially zero and fixing the cameras rigidly mimics conjugate eye movements. For close scenes, a calibrated vergence is needed to mimic the imaging done by human eyes. We are considering a camera mount to approximate dynamically the effects of vergence by image shifts calibrated to the lens focus distance.

Display - We utilize both traditional two-view (right-left) stereo image display systems and the latest technology multiple-view stereo displays. Several techniques are being utilized and tested in this project, including systems with and without supplementary glasses, and both projection or direct view systems using flat-panel or plasma displays [1, 2]. Traditional virtual-reality (VR) or augmented reality (AR) single-user head-mounted displays are not acceptable for this project because they are heavy, bulky and have a limited field of view. Single projector displays use synchronized active shutter glasses to provide right and left eye images, while two projector displays use passive polarizing glasses for this purpose. Both of these options are being compared and evaluated.

One relevant issue for any type of display is ***projector calibration***. Ideally, we want a perfect single registered image from the two projectors. We will initially place them either side-by-side or stacked vertically and apply auto-calibration techniques to minimize the relative image distortion due to the offset. Given the controls available, we expect that a good static calibration will be achieved for stacked projectors. We will explore more advanced optical beam-combiner designs that keep the optical axes of the two projectors aligned. ***Intensity and color calibration*** of the whole acquisition and display process is critical to maintain stereo image quality. While we expect that intensity tracking between the two cameras and displays should be very good, our system will make calibration measurements and devise a dynamic gray-scale lookup correction scheme to minimize any variations. Future cameras will have on-chip signal processing for adaptively shifting the output range and gain variations. Colorimetry difficulties arise from the nonlinearities of the analog image sensor and electronics. For a first order correction, we will calibrate the black and white endpoints of the RGB response curves and. We are also developing methods to calibrate multiple points of each camera and display system response using color charts to minimize colorimetry differences by incremental tuning.

In addition, we are acquiring very new multiple-view ***autostereoscopic*** display systems that produce stereo video without the use of glasses. This glasses-free feature has tremendous potential advantage in this project in improving the immersive experience. These displays use liquid crystal (LC) or plasma flat-panel technology, and are currently available in sizes exceeding 50". They utilize lenticular cylindrical lenslet arrays or barrier-screen (“picket-fence”) display overlays to produce a large number of independent views (typical nine or more) to observers [1, 2]. These independent views are simultaneously displayed on the screen and one or more or observers see smoothly blended different left and right eye images as a function of their position [4, 5]. One recently announced display utilizes an LC panel with native 3820x2400 pixel resolution [2]. These displays require considerable processing of transmitted stereo information and very high performance graphics cards to produce the multiple views from either two-view or multiple-view stereo source data. We are developing the software necessary to implement these techniques.

Transmission - This project requires very high fidelity multi-node audio-visual communication over local area and wide-area shared networks. The single greatest limiting factor for human interaction in this immersive project is the ***effective latency***. Compression is used to overcome bandwidth limitations of network transmission, at the expense of greatly increased delay. The nature of future interactive IMSC projects makes the delay due to compression to be intolerable, requiring the use of high bandwidth networks to transmit uncompressed (or minimally compressed) content [6, 7]. Initial experiments have shown that required maximum latencies range from 10-100 ms depending on the experimental conditions and content.

Table 1 is a condensed summary of several possible stereo 3D video transmission formats that we will explore. Not all formats we have considered are listed, only the most promising ones for immersive applications. Here the second column shows the additional bandwidth needed for general *N*-view stereo in addition to the usual 2D compressed bandwidth, while the third column specifies the 3D flexibility: the adjustment of stereo gain, focal points and adaptability for different displays. The next two columns describe the backward compatibility with 2D displays and the spatial resolution relative to the original 2D resolution. The last two columns list the ability to see around objects and the computational complexity of generating 3D images from the transmission form.

The interdigitated format is *N* (≥ 2) pre-rendered views, interleaved and ready for display on a specific 3D display device. Time multiplexed views are full or slightly reduced resolution views transmitted in time sequence: e.g. frame1-view1, frame1-view2 … frame1-view*N*, frame2-view1, etc. Multiple streams has all *N* views encoded at full or close to full resolution as *N* separate streams (*e.g.*, MPEG-2) without any interstream coding. Spatial tile encodes *N* views in the same spatial resolution as a single frame (e.g. nine views in a 3x3 grid). Source and depth sends acquired or synthetic stereo depth or disparity information at lower resolution in addition to the 2D frame at standard resolution. Source and NN transmits neural net structures that are used to build depth maps during decoding.

Format	Additional Compressed Bandwidth	3D Controll-ability	2D Back Compatible	Spatial Resolution	Motion Parallax/ Look-around	Decode CPU
1. Interdigitated	Very High	No	No	High	Yes	Low
2. Time-Multiplexed Views	Very High	No	Yes	High	Yes	Low
3. Multiple Streams	Very High	No	Yes	High	Yes	Low
4. Spatial Tile	Medium/ Low	No	Yes	Low	Yes	Low
5. Source and Depth	Medium	Yes	Yes	High	Limited	Medium
6. Source and NN	Very Low	Yes	Yes	High	Limited	Very High
7. CG models	Low?	Yes	Yes	High	Yes	High

Table 1. Some Stereoscopic 3D Transmission Formats [8].

CG models send full video binary X3D/VRML models, lighting and animation of synthetic or real-world scenes. A Yes entry in the 3D controllability column denotes techniques containing detailed stereo information that can drive a wide variety of two-view or multiple-view displays.

As part of this project we will evaluate these formats on the basis of perceived image quality and latency due to compression. We will develop several different stereo video compression techniques. The techniques are highly dependent on the exact acquisition and display formats used. Left and right stereo pairs acquired from cameras are highly correlated, thus significant compression is possible by extracting this information from interdigitated or time-multiplexed data and transmitting it with reduced bandwidth. For autostereoscopic displays, the multiple views can either be transmitted in parallel (each at reduced resolution) or generated at the destination site from two-view information. Thus various levels of processing are needed to convert transmitted stereo information for display, depending on its format (two-view or multiple-view). We note that other tradeoffs between the characteristics of an individual format are possible and that CG models are not always available.

Interaction - We are investigating techniques for human observers to interact with an autostereoscopic (no glasses) display using off-the-shelf video cameras and user hand gestures. The objective is for a user to manipulate the position and orientation of displayed real or virtual 3D objects and to outline, draw or interact with the object. In our initial approach the user draws figures in front of an autostereoscopic display using a tiny point source flashlight. Two cameras at 90-degree angles track the flashlight and send its position in 3D space to the computer. Using OpenGL, the computer redraws these points in virtual space. The next step is to create multiple images from a set of nine or more virtual cameras to drive the typical autostereoscopic display.

5. Short Description of Achievements in Previous Years

This is a new IMSC effort begun recently. Many previous IMSC projects have made provisions for stereo image and video display, but only recently has new technology greatly simplified the acquisition and display of stereo information in a convenient manner.

5a. Detail of Accomplishments During the Past Year

We have begun initial experiments in acquiring, displaying and interacting with stereo images. We assembled 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. We are acquiring an autostereoscopic display system and software needed to drive it. We finished writing the software that can use the output of two cameras for tracking a flashlight and generate multiple images for driving the display. Currently we are waiting for the display and its SDK to arrive so we can combine our program and test them.

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

There are very few reported results on the effectiveness of stereo video and high quality audio in creating an immersive experience [8]. One paper reports that, combined stereo video and 5.1 channel audio, are mutually beneficial in helping to maintain visual left-right eye convergence [9]. In the area of interaction with autostereoscopic displays, DeWitt described the possibility of interaction but hasn't done research on it to our knowledge [10]. The Multimo3D group at Heinrich-Hertz-Institute in Germany implemented a multi-modal interface with an autostereoscopic display [11]. Their system tracked a user's hand, head and eyes with cameras and fed the information to a one-user autostereoscopic display for interaction. More recently, Berkel at Phillips Labs has shown that it is possible to track a user's hand and fingers with magnetic fields using sensing electrodes in active matrix form or around the edges of a display 12]. They are using this information to interact with an autostereoscopic display.

Our work differs from Multimo3D in following ways: we can achieve higher precision than theirs because of the size of the light source being tracked; because the autostereoscopic display is a multi-user device, several people can see the interaction at the same time and interact with the display by tracking multiple light sources. Compared to other work, our initial approach uses more expensive frame-synchronized cameras to simplify the processing for tracking. In the future it may be possible to use cheaper cameras at the cost of additional processing for synchronization.

7. Plan for the Next Year

We initially use two relatively inexpensive and compact calibrated frame-synchronized IEEE-1394 cameras viewing the same scene with an effective optical axis displacement adjustable down to the typical human inter-ocular distance (around 10 cm). Another approach is to investigate new stereo video camera designs that multiplex two images onto a single sensor array making the system more compact and potentially eliminating some calibration problems. Examples include a frame multiplexing system made by NuView, and an optical spatial multiplexing system from Canon. Finally we will explore the use of HD-SDI cameras for higher effective resolution. Several new types of compact HD cameras with small physical size have recently become available and may be mounted directly side-by-side at the correct inter-ocular distance. In this way we expect to avoid the use of bulky beam-steering optics that are needed to bring the optical axes of two larger cameras to the close inter-ocular spacing. We are acquiring an autostereoscopic display and setting up two-projector large screen displays for experimentation. When these facilities are operational, we will explore interaction over the Internet and with hand gesture recognition. We are planning a set of experiments that include: psychophysical tests to determine the minimum necessary resolution and image quality; maximum tolerable latency; and the effectiveness of various stereo video compression and display modes in immersive applications.

8. Expected Milestones and Deliverables

Over the next year we will we demonstrate NTSC or DV quality stereo video acquisition; transmission over local-area networks; and projection or autostereoscopic display between two sites. We will begin work on the evaluation of stereo video compression and low-latency transmission; and conduct psychophysical tests on the effectiveness of stereo in various IMSC immersive applications. We will investigate the integration of stereo video with multichannel audio acquisition, rendering and synchronization in conjunction with other integration projects. In subsequent years we will upgrade to HD quality stereo video acquisition, transmission and display, and extend our experiments to off-campus partners connected by a wide-area shared network such as Internet2.

9. Member Company Benefits

Companies whose business is in high performance networks, the creation of immersive environments for teleconferencing, education, gaming, and entertainment will benefit from the results of this research. The addition of stereo video capability and display technology that does not require clumsy goggles or head-mounted displays to existing and future immersive environments is of great benefit.

10. References

[1] Dynamic Digital Depth, Inc. web site: http://www.ddd.com.

[2] Stereographics, Inc. web site: http://www.stereographics.com.

[3] A. Woods, T. Docherty, R. Koch, "Image Distortions in Stereoscopic Video Systems", *Stereoscopic Displays and Applications IV, Proceedings of the SPIE*, Vol. 1915, San Jose, CA, (1993).

[4] N.A. Dodgson, "Analysis of the Viewing Zone of Multi-View Autostereoscopic Displays", *Stereoscopic Displays and Applications XIII Conference, Proceedings of SPIE*, Vol. 4660 (2002).

[5] C. van Berkel, "Image Preparation for 3D-LCD", *Stereoscopic Displays and Virtual Reality Systems VI, Proceedings of the SPIE,* Vol. 3639, (1999).

[6] D. McLeod, U. Neumann, C.L. Nikias and A.A. Sawchuk, "Integrated Media Systems," *IEEE Signal Processing Magazine*, vol. 16, no. 1, pp. 33-76, January 1999.

[7] C.L. Nikias, A.A. Sawchuk, U. Neumann, D. McLeod, R. Zimmermann and C.C Kuo, "Total Immersion," *OE Magazine,* vol. 1, no. 7, pp. 20-23, July 2001.

[8] 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, Proceedings of the SPIE,* Vol. CR76, November 2000, Boston, MA (2000).

[9] J.A. Rupkalvis, “Human Considerations in Stereoscopic Displays”, *Stereoscopic Displays and Virtual Reality Systems VIII, Proceedings of SPIE*, Vol. 4297 (2001).

[10] T. DeWitt, “Visual Music: Searching for an Aesthetic”, *Leonardo*, Vol. 20, No. 2, pp. 15-122, (1987).

[11] Heinrich-Hertz-Institute, mUltimo3D group web site: http://imwww.hhi.de/blick/.

[12] C. van Berkel, “Touchless Display Interaction”, *SID 2002 Digest,* Society for Information Display, (2002).