Camera and Projector Arrays for Immersive 3D Video

Harlyn Baker
Hewlett-Packard Laboratories
Palo Alto, CA
650-857-2584
harlyn.baker@hp.com

Zeyu Li
EECS Department
University of California
Berkeley, CA
zeyuli@berkeley.edu

ABSTRACT
Applying recent advances in multi-imager capture and multi-projector display, we combine capabilities through the Nizza multimedia dataflow architecture to deliver low-cost wide-VGA-quality low-latency autostereoscopic 3D display of live video on a single PC. Supporting multiple users as they observe and interact against a life-sized display surface responsive to their positions, this facility will open new opportunities in mediated interaction.

Keywords
Multi-viewpoint capture, multi-viewpoint display, autostereo immersive display

1. INTRODUCTION
The future of human-computer interaction lies in eliminating the perceived barriers between people, and between them and their machines, and providing enhanced capabilities through an intelligent and responsive interface. The visual is the leading layer in this interface, and presents the first obstacle to delivering users an experience that is as immersive and compelling as reality—i.e., personal face-to-face interaction. Principal in bridging this chasm is delivering comfortable unencumbered 3D perceptions. Our focus here is on the capture and delivery of autostereoscopic (that is, without glasses) 3D video in a live and scalable life-sized demonstrator setting.

Establishing 3D video communication is a multi-viewpoint acquisition and display challenge. It involves capturing, transmitting, and reconstructing enough of the local lightfield—the set of rays emanating from a scene—to convince viewers that they are observing a reality. While needing rather sophisticated input and output elements, bridging these devices necessitates understanding the geometric relationships among them (cameras and displays) and configuring the signals to best present the captured reality. The challenge of our work lies in five major areas: multi-viewpoint capture, multi-viewpoint display, math and analysis for calibration across them, efficient compression to permit reasonable transmission, and smart processing of the signals to provide an interactive experience. This paper addresses the first three, leaving transmission and interaction for later study.

Immerscom, May 27-29, 2009, Berkeley, CA, USA.

2. SYSTEM ISSUES IN LIVE AUTOSTEREOSCOPIC DISPLAY
Our laboratory at HP is well situated for investigating immersive 3D communication in that we have major components of this challenge in house, established in use for related applications, and ready to be deployed for 3D. As part of earlier work in videoconferencing (Coliseum and Halo projects [2,3]), we developed a multi-imager camera system—the FanCamera—that can obtain, as of this moment, up to 72 wide-VGA video streams to a single PC at frame rate and without compression (a 24-imager linear configuration is shown in Figure 1). GPUs gave us the ability to reconfigure 24 of these streams in real time as a shapeable panorama, blending the parts into a unified view. At the same time Labs has been a world leader in combining projector outputs for scalable shaped display (Pluribus [5] and Panoply [7]). Our studies here have given us tools for calibrating systems of cameras and displays. Included in this is a novel approach to camera calibration that capitalizes on high quality homographies between pairs of imagers to develop a global optimal solution delivering

Figure 1. Multi-imager camera array.

Digital Object Identifier: 10.4108/ICST.IMMERSCOM2009.6238
http://dx.doi.org/10.4108/ICST.IMMERSCOM2009.6238

epipoles and fundamental matrices simultaneously for the entire system [8]. In addition, we employ the planar calibration method of Zhang [15] in comparative studies (to be published elsewhere) in obtaining the geometric relations among cameras.

Rectification – the reorienting of images so that their epipolar lines are located on corresponding scan lines – is an essential step in both recovering geometry from the scene and in structuring imagery for autostereo display. This operation minimizes the vertical disparity at corresponding points in the images, simplifying both the matching process of stereo analysis and the fusion of binocular vision. It is not possible to align cameras manually to the precision needed for either metric analysis or viewing – there are just too many imaging parameters involved to expect any cooperation from assembly. In addition, lenses usually must have their distortions removed before analysis or viewing and this necessary resampling can be integrated with epipolar alignment resampling. For free-viewpoint autostereo viewing, we must attain this rectification of the images.

Refinement of the determined camera models to deliver minimal vertical misalignment in an epipolar sense is used to permit ganged rectification of the separate streams for transitive positioning in the visual field. This means that a single transform per camera will correct the data for viewing by any pair of eye viewpoints. The alternative would require *n*-choose-*2* transforms (all pairs) and be fraught with visual jitter as the different pair-wise transforms are applied. This structuring is key to arranging the video data both for 3D display and for the recovery of scene geometry—required if one wants to move on to a "*responsive*" interface—and represents an area where we have made advance. Figure 2 sketches this in the case of a 3-imager system. Top is the given configuration as determined by individual calibration, middle shows a directional alignment, and bottom indicates the synthesized least-squares calibration that permits epipolar or transitive alignment.

Figure 3 shows pre-rectification error plots for a twelve-imager camera array, using Zhang's method for calibration. Using the checkerboard corner vertices as reference points, we measured a mean error of 0.32 pixels, with a max error of 1.3 pixels. Maximum deviation from epipolar alignment was 0.8

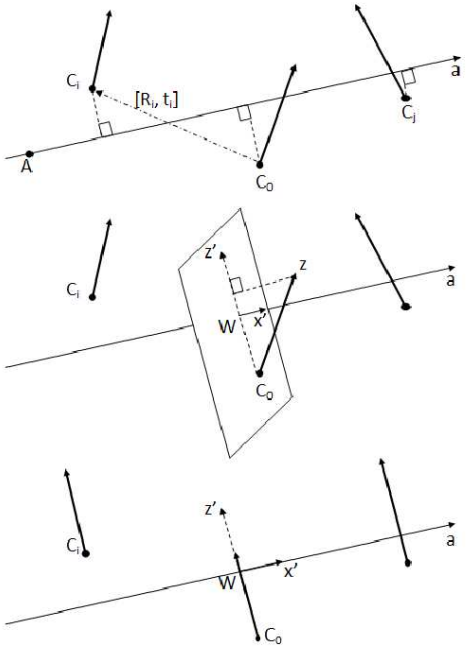

Figure 2. Calibrated then synthesized geometry for Epipolar alignment.

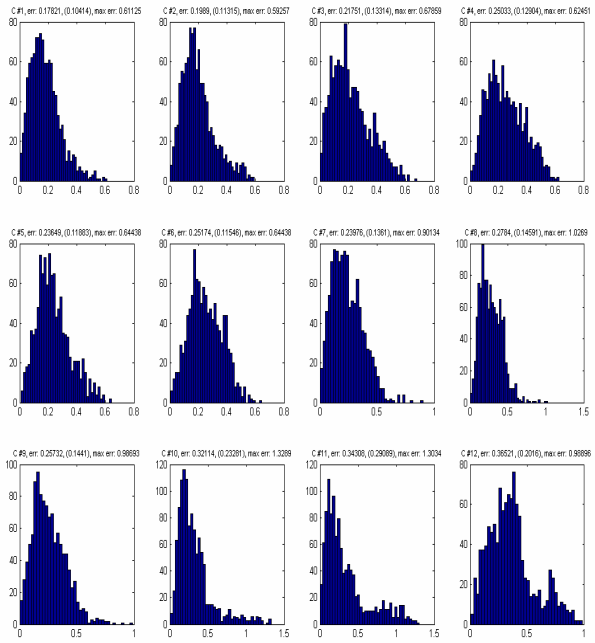

Figure 3. 12-imager calibration error distributions.

pixels. After the LSQ adjustments, max epipolar deviation was reduced to 0.6 pixels. Figure 4 (top) shows a set of selected epipolar lines across a band in 10 rectified images, with Figure 4 (bottom) indicating one of those full images.

Digital Object Identifier: 10.4108/ICST.IMMERSCOM2009.6238
http://dx.doi.org/10.4108/ICST.IMMERSCOM2009.6238

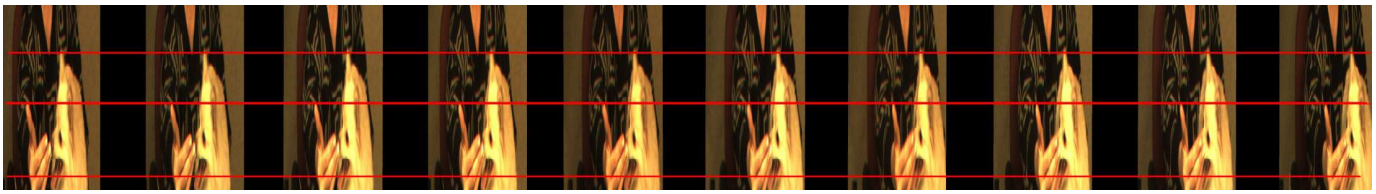

Figure 4. (top) Selected epipolar lines to reveal vertical alignment ; (bottom) one full image with selected lines.

On the receiving display side, we determine individual homographies [6] for projectors in an array directed at a 3D display surface. The homographies adjust the projector outputs so that their content coincides and proper alignment is retained. The camera transforms mean that vertical epipolar disparities of the captured signal are minimized, and the projector transforms mean the display will retain these alignments despite projector pose variations. The projector calibration also permits arbitrary alignment shifts to accommodate focus-of-attention vergence, should that information be available for example through gaze tracking. Tying together our computational elements is HPL's Nizza dataflow

Figure 5. Multi-projectors; Retroreflective surface.

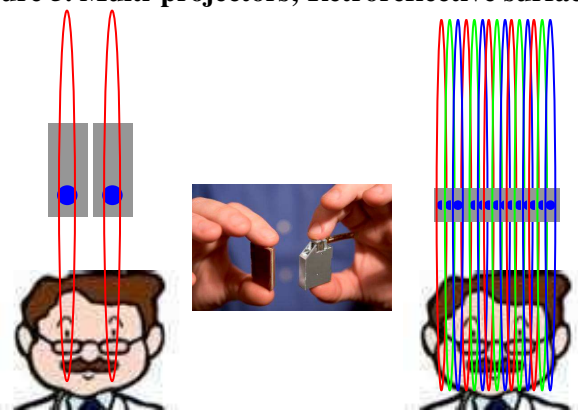

Figure 6. Two projectors with a view zone over each eye—binocular stereo (left); the intended distribution—multi-view zone autostereo (right), enabled by pico projectors (center).

architecture [13], significantly simplifying our developments and performance optimization.

A novel element of this solution is our 3D display surface. It uses retroreflective material with a diffusing layer that permits viewers located within the reflected diffusion zone to see the output of an overhead-mounted projector (Figure 5). All of the signal emitted by a projector and hitting a retroreflector would return to it alone if there were no diffusion (i.e., no one would see anything). But, with diffusion of, say, 1 degree horizontally by 30 degrees vertically at the display surface, anyone positioned

Digital Object Identifier: 10.4108/ICST.IMMERSCOM2009.6238
http://dx.doi.org/10.4108/ICST.IMMERSCOM2009.6238

within the ellipse defined by these values with respect to the projection source will see the projected image (sketched in Figure 6 left). Careful placement allows the projectors to uniformly cover the participant view-zone area (Figure 6 right), while calibration lets the images align properly for autostereo viewing. Figure 6 center shows a replacement for these projectors—once pico projectors establish themselves in the market—enabling the close packing of the right sketch.

The projectors of Figure 5 present 25 lumens each—completely unacceptable for general use but ideal here, given the enormous gain of the retroreflective surface. Figure 7 shows two dual early pico-projector systems of about 10 lumens each. Figure 8 shows the output of a single projector of Figure 5 on our retroreflective diffusing surface. Figure 9 shows a similar display surface where 3 view zones are super positioned for multi-view presentation [10].

Figure 7. A pair of Microvision (left) and 3M (right) pico projectors assembled in a binocular stereo mode.

A critical element at this stage is limiting the horizontal diffusion so that ghosting is eliminated. We are not there yet. Our diffusers provide visibility within abut a 1.6 degree lateral cone and drop off to about 4% signal at two interocular distances. These figures must be reduced to about a half degree and near zero to accommodate more densely spaced projection sources without cross talk. Figure 10 shows diffusion plots for our deployed material (Figure 8) in comparison with that of the earlier MultiView material (Figure 9) with ours being the tighter. Note that they both exhibit cross talk at one interocular (1.6 and 3.2 degrees at FWHM), and neither drops low enough to prevent white in one eye's view zone being seen as brighter than black in the other eye's view zone (floors of 6% and 21%). Removing this limitation of current diffusers is one of our next tasks.

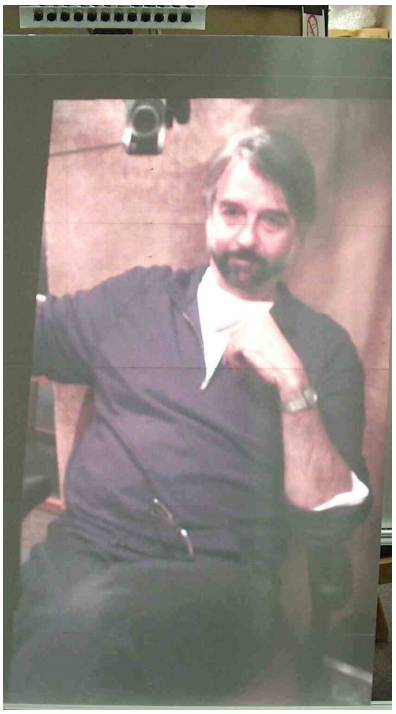

Figure 8. A single view zone projection.

Figure 9. Multi view zone projection display.

3. NOVELTY OF APPROACH

The novel contributions of our efforts here include:

1. A single-PC solution to 3D capture and display of live multi-viewpoint video
2. Multi-camera calibration for optimal epipolar alignment and 3D resampling
3. Calibration of a multi-projector system for front-projection autostereocopic display
4. Design and configuration of the retroreflective diffusing screen
5. Demonstration of a real-time life-sized low-latency autostereo experience.

Digital Object Identifier: 10.4108/ICST.IMMERSCOM2009.6238
http://dx.doi.org/10.4108/ICST.IMMERSCOM2009.6238

4. CURRENT DEVELOPMENT STATUS

The system is in operation with 9 cameras and 9 projectors offering 7 discrete binocular view zones at the plane of the projectors (a near-infinity of valid 3D views is attainable through the sampling of viewer forward and backward motion in the area of frusta intersection). Video bandwidth in this configuration approaches a gigabit per second; our camera system can support 8 times this number (72 imagers). Display fan out is the current view-zone restrictor. We have shown the system locally and have a portable version for transporting to demonstrations. We have implemented rudimentary horopter selection and aim at having vergence accommodate to viewer position and to viewer gaze direction, to be determined through analysis of the acquired multi-imager video stream, Additionally, we are experimenting with pico projectors (Figure 7) for simple binocular stereo display.

5. COMPETING APPROACHES

Holografika [4] has a rear-projection life-size 3D display system utilizing many dozens of laser projectors. It costs about a half million dollars, requires over a dozen workstations, and is capable of displaying only computer graphics data—they, as others, have no source of multi-viewpoint video. The optics of their approach is similar to ours, only using transmission rather than reflection. SeeReal [12] have solid-state displays employing holographic means for compression and resampling, but again are limited to CG data sources. The only other comparable multi-imager camera system was developed at Stanford University [14]. It handles much less data, requires multiple PCs, uses MPEG compression, and has little capability of online capture and display. Mitsubishi Research [9] demonstrated a multi-PC 16-camera-projector system at SIGGRAPH 4 years ago. Clarity was low, scalability limited, cost high, and system integration lacking. Our solution provides a unified single-PC approach, where cameras, displays and computation could be delivered for under $10K. Further, the mechanisms of data rectification, transmission, and resampling enable our approach to be used for the recovery of scene 3D geometry [1].

6. NEXT DEVELOPMENTS

Our current plan is to work on issues of perceptual quality—including diffusion, projector density/fan-out, resolution, sharpness [11], color clarity, and related image processing means for experience improvement—and to build up the camera and projector counts to permit a much wider range of operation. The aim is to display the acquired live video on a life-sized surface on the area of 6 feet by 10 feet.

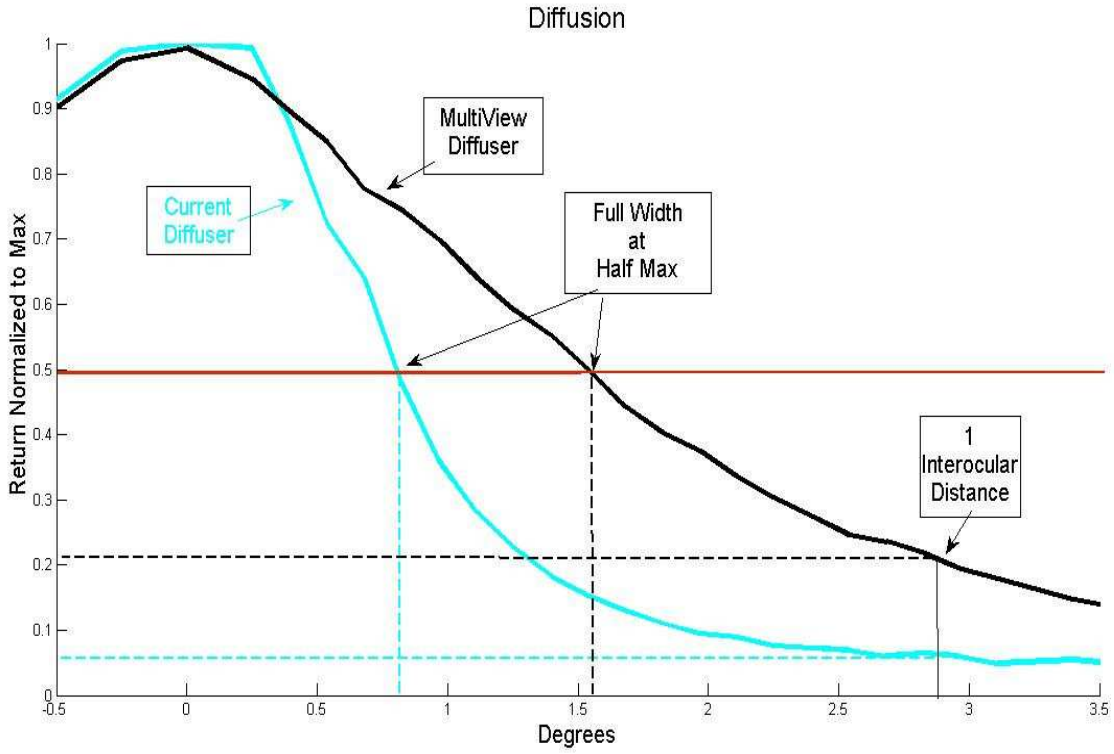

Figure 10. Diffusion angles with respect to max.

Digital Object Identifier: 10.4108/ICST.IMMERSCOM2009.6238
http://dx.doi.org/10.4108/ICST.IMMERSCOM2009.6238

7. REFERENCES

[1] H. Baker, R. Bolles, "Generalizing Epipolar-Plane Image Analysis on the Spatiotemporal Surface,'' *Intl Jour Computer Vision,* 1989.

[2] H. Baker, N. Bhatti, D. Tanguay et al, "Understanding Performance in Coliseum, an Immersive Videoconferencing System, '' *ACM TOMCCAP,* 2005.

[3] H. Baker, D. Tanguay, "A Multi-Imager Camera for Variable-Definition Video (XDTV),'' Springer-Verlag, MRCS, 2006.

[4] T. Balogh, "Method and apparatus for producing 3D picture," U.S. Patent 5 801 761, Sep. 1, 1998.

[5] N. Damera-Venkata and N. L. Chang, "Realizing Super-resolution with Superimposed Projection," IEEE *ProCams*, 2007.

[6] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Press, 2000.

[7] M. Harville, B. Culbertson, et al, "Practical Methods for Geometric and Photometric Correction of Tiled Projectors on Curved Surfaces," IEEE *ProCams*, 2006.

[8] Z. Li, H.Baker, R. Schreiber, "Calibrating a Multi-Camera Array: Estimation of Epipoles and Fundamental Matrices in Multi-Camera Multi-Plane Systems,'' HP Labs Tech Report HPL-2007-31 (currently internal access only).

[9] W. Matusik, H. Pfister, "3D TV: A Scalable System For Real-Time Acquisition, Transmission, And Autostereoscopic Display Of Dynamic Scenes," *SIGGRAPH'04*, 2004.

[10] D. Nguyen, J. Canny, "MultiView: Spatially Faithful Group Video Conferencing, '' *Proc. CHI 2004*, ACM Press, 2004.

[11] F. Russo, **"**An image enhancement technique combining sharpening and noise reduction**''** *IEEE Trans. Instrumentation and Measurement*, 51:4, 2002.

[12] http://seereal.com/en/autostereoscopy/index.php, SeeReal Technologies Inc.

[13] D. Tanguay, D. Gelb, H. Baker, "Nizza: A Framework for Developing Real-time Streaming Multimedia Applications,'' HP Labs Tech Report HPL-2004-132 (currently internal access only).

[14] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, M. Horowitz, "High speed videography using a dense camera array,'' CVPR, 2004.

[15] Z. Zhang, "Flexible Camera Calibration by Viewing a plane from unknown orientations, '' *ICCV*, IEEE Press, 1999.

Digital Object Identifier: 10.4108/ICST.IMMERSCOM2009.6238
http://dx.doi.org/10.4108/ICST.IMMERSCOM2009.6238