Calibrating Camera and Projector Arrays for Immersive 3D Display

Harlyn Baker[a], Zeyu Li[b], Constantin Papadas[c]
[a]Hewlett-Packard Laboratories, Palo Alto, CA (harlyn.baker@hp.com)
[b]EECS Department, University of California, Berkeley, CA
[c]Integrated Systems Development, S.A., Athens, Greece

ABSTRACT

Advances in building high-performance camera arrays [1, 12] have opened the opportunity—and challenge—of using these devices for autostereoscopic display of live 3D content. Appropriate autostereo display requires calibration of these camera elements and those of the display facility for accurate placement (and perhaps resampling) of the acquired video stream. We present progress in exploiting a new approach to this calibration that capitalizes on high quality homographies between pairs of imagers to develop a global optimal solution delivering epipoles and fundamental matrices simultaneously for the entire system [2]. Adjustment 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. Individual homographies [6] are obtained for a projector array that presents the video on a holographically-diffused retroreflective surface for participant autostereo viewing. The camera model adjustment means vertical epipolar disparities of the captured signal are minimized, and the projector calibration means the display will retain these alignments despite projector pose variations. The projector calibration also permits arbitrary alignment shifts to accommodate focus-of-attention vengeance, should that information be available.

.

Keywords: camera array, multi-viewpoint capture, multi-imager calibration, autostereo display, retroreflection

1. LIVE IMMERSIVE 3D VIDEO DISPLAY

While there has been much research in building devices to provide autostereoscopic viewing, this has been done almost exclusively in the context of presenting computer graphics (CG) models for off-line visualization. Little capability exists for capturing and presenting live multi-viewpoint 3D video. We present here a system and supporting methods for providing this immediate coupling of live 3D video capture and display to a participant on a single workstation. Our contribution covers multi-imager capture, calibration of these cameras for epipolar rectification, calibration of a projector array for displaying these streams for autostereo viewing, and a software infrastructure that runs all of this at frame rate using GPUs while imposing negligible CPU burden. We begin by discussing the display issues, move to capture and its calibration, then unify these ideas through description of a system that provides a realtime visualization loop for autostereo interaction.

2. MULTI-VIEWPOINT DISPLAY

2.1 Monocular versus binocular and autostereoscopic display

Most display technologies are monocular, where both eyes see the same thing. In contrast, a stereo display presents distinct viewpoints so an observer's left and right eyes see different perspectives, and this delivers a sense of 3D. Such binocular stereo display has been around for decades (since Wheatstone in the 1830's). Numerous mechanisms can deliver binocular stereo, including barrier filters over LCD displays (vertical bars act as a picket fence, channeling data in specific directions for the eyes) and complementary filters placed jointly over two overlapping projectors and over the two corresponding eyes (i.e., anaglyph, linear or circular polarization, or the narrow-pass filtering of Infitec). Shutter glasses provide a different sort of control by synchronizing production of left-right images with their alternate acquisition by the viewer's eyes. Two-view binocular display presents a limited experience—all observers receive the same fixed perspective; viewed statically, a scene takes on a cardboard-cutout appearance since proprioceptive

expectations are not met; without accomodation to changing position, a viewer's motion makes the static views deliver the appearance of a skewing/twisting scene. The goal of *immersion* suggests a natural, authentic, stable, and compelling experience. Wearing special devices and providing only two views works against this.

Autostereo implies that the perception of 3D is in some manner automatic, and not requiring devices such as glasses—either filtered or shuttered. Optics can provide this, for example with lenslet or lenticular elements arranged to make parts of an underlying composite image visible only from certain view directions. Typically, a lenticular display multiplexes separate images in cycling columns beneath its elements making them take on the color of selected pixels beneath when viewed from different directions. LCDs[1] or projection sources [3] can provide the pixels for such display. These techniques serve to recreate a *lightfield* that a user can sample by positioning himself at different locations within its presentation range. The notion of lightfield comes from the fact that the field of light that would arise from the true scene is recreated and directed as it would be were the scene actually present. Integral imaging presents a discrete version of a lightfield, while holography provides the continuous correlate.

If a 3D display method permits autostereo at only a *sweetspot*, we term it *constrained-viewpoint* display. More complex methods are needed in delivering *free-viewpoint* display—which permits viewing from a range of perspectives and perhaps for many viewers simultaneously. Our interest is in free-viewpoint autostereo of live video for multiple simultaneous participants.

2.2 Delivering a Lightfield

Over the past few years we have collaborated with researchers at UC Berkeley on a prototype front-projected multiview display using retroreflective and diffusing materials [9]. In this work, three projections were fully superimposed on the screen, with their sources positioned roughly behind and above each of three participants so each's view would be just of what came from that projector. The projector displays video collected from a symmetric remote site, taken from just above where each participant is projected on that remote site screen (see Fig. 1). This was, essentially, a coarse version of the Mitsubishi Research (MERL) 16-PC 3D TV display system [8]. Such an arrangement of projectors and cameras delivered more-natural eye contact and greater gesture awareness, with studies showing a measurable increase in trust of about 30% versus standard (i.e. non-multiview) video conferencing [10]. This prototype multi-view system, while monoscopic with fairly large view zones and a significant amount of visual crosstalk (see Fig. 1 right), motivated us in developing stereoscopic equivalents using more refined materials—including corner-cube retroreflectors and holographic diffusers—to give tighter control of the illuminant return to the point where it supported free viewpoint autostereo. Fig. 2 left shows a battery of pocket projectors, and Fig. 2 right gives a view of the retroreflecting holographic surface they aim at that we have assembled in our autostereo display studies.

Fig. 1. HPLabs-UCBerkeley MultiView Facility:
(left) projectors; (right) display and cameras.

Fig. 2. Autostereo display:
(left) above-viewer pocket projectors;
(right) diffusing retroreflection screen

Retroreflective display with diffusion works with front projection to permit viewers located within the diffusion zone to see the output of a projector, here placed above the viewer's head. All of the signal emitted by the projector would return to it alone were there no diffusion (i.e., no one would see anything), but with diffusion of, say, 1 degree horizontal

[1] http://www.wowvx.com/

by 30 degrees vertical at the display surface, anyone positioned within the cone defined by these values with respect to the screen will see the projected pixels (sketched in Fig. 4 left). Fig. 3 plots sampled horizontal diffusion values for a variety of materials, with our original lenticular in black and our chosen material in magenta (some of these materials were not available in roll/life-size form). Notice that the lenticular is still at 15 percent return at +/- 3 degrees, accounting for the noticeable ghosting we experienced. Magenta is at about 6 percent, so dropoff is still not ideal. Careful placement allows the projectors to somewhat-uniformly cover the participant view-zone area (sketched in Figure 4 right). Fig. 5 shows an overhead view of the projector-camera-displayscreen configuration, with sample projected rays indicated with their diffusion. Fig. 6 shows what the screen returns from a camera-projector pair in a single view zone.

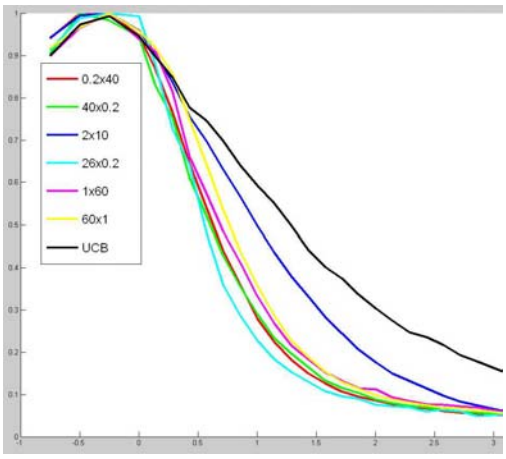

Fig. 3. Spreads of various diffusing materials (% vs degrees).

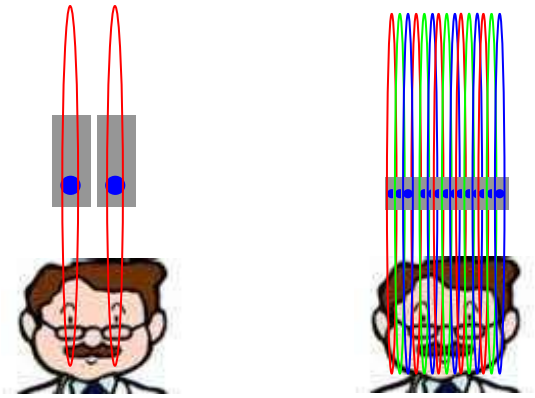

Fig. 4. (left) Two projectors with a view zone over each eye—binocular stereo; (right) the intended distribution—multi-viewzone autostereo.

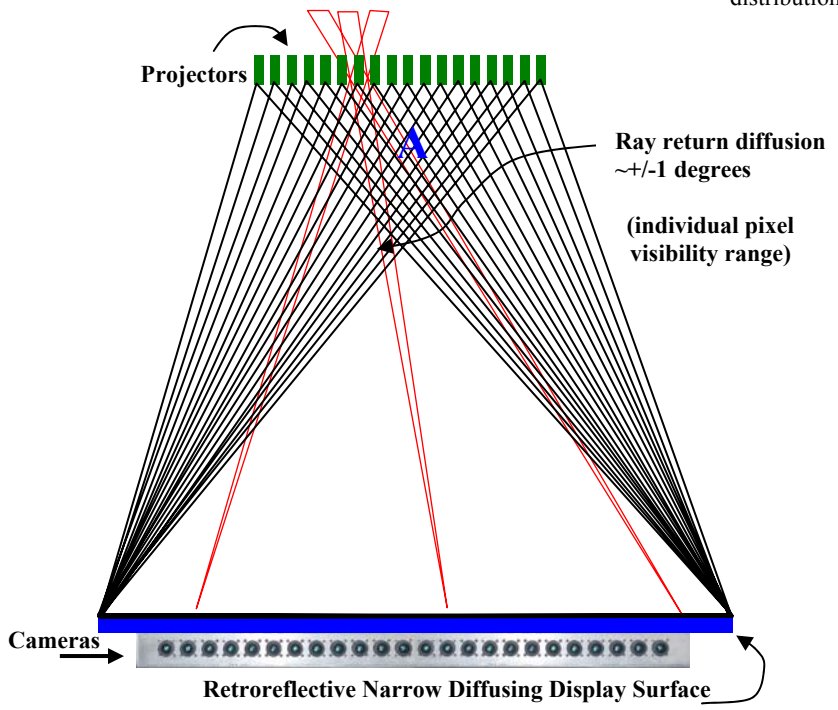

Fig. 5. Schematic of display system: Cameras and retroreflective diffusing screen at bottom (blue); projectors above (green); cast images as black-outlined triangles and demonstrating cast rays in red. Blue A marks the region of full autostereo; elsewhere exhibits windowing.

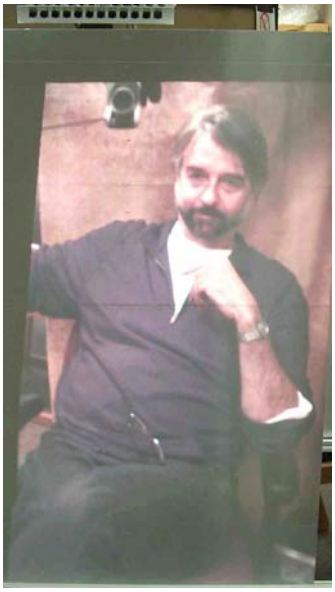

Fig. 6. A single viewzone projection; cameras above.

2.3 Autostereo multi-projector calibration

We presume that the imagery presented to the projectors is properly aligned for stereoscopic viewing (see Section 3 for what this constitutes and how to attain it). We wish to preserve this alignment despite the arbitrary poses of the various projectors. To achieve this, we must correct for placement, scaling, and keystoning of the imagery on the screen. A camera near the projectors views the screen and observes a calibration target as cast by each projector. Display involves a series of transforms, the last of which (the physical projection) in unavoidable and immutable—our adjustments must occur before this is applied. A homography H_i^P is computed for each projector Q_i with respect to the calibration target P. We select a reference projector Q_0 and seek a mapping H_0^i for all other projectors Q_i such that their presentation of the target aligns with that of Q_0; that is, we want $H_0^P = H_i^P * H_0^i$ or $H_0^i = \left(H_i^P\right)^{-1} * H_0^P$. There are a number of other image operations being performed, such as placing the video in a graphics display buffer, ensuring the display area is maximized for the given camera and projector overlaps, and bringing the reference projector into alignment with the viewers' interocular horizontal baseline, so there is somewhat more to the transform that this sketches. In addition, we want to ensure color is correct and consistent across the various projectors, and calibration target analysis is providing for this as well.

3. MULTI-VIEWPOINT CAPTURE

3.1 Multi-camera video

To support immersive 3D capture at video rates, a camera system must manage considerable bandwidth. A dozen VGA cameras streaming Bayer format imagery at 30 fps is nearly a gigabit per second—or nearly double to triple the sustained capacity of the currently best USB and 1394 busses. While bus performance is increasing, it is not tracking the exponential rise in imager bandwidth. Additionally, synchronization presents challenges for firewire and USB, usually requiring external signaling.

Systems with large numbers of cameras typically use multiple PCs, low frame rate, and/or high compression. Wilburn [12] designed and built a custom PCI card solution that captured 25 VGA cameras onto a single PC with hardware compression to disk. Four such systems were combined to store MPEG'd video from 100 cameras. Yang [13] built a camera that rendered live light fields from an 8x8 arrangement of unsynchronized webcams, treating the cameras as individual sources for desired viewpoints—they did not provide imagery in concert. Zhang and Chen [14] built a system to capture and render unsynchronized JPEG'd video from 48 commodity QVGA Ethernet cameras.

Camera Link offers higher bandwidth but, with its large connectors, 28-strand cabling, lack of power or synchronization signaling, and current provision for only single or dual camera configurations, does not suit dense massive synchronized capture. Gigabit Ethernet similarly fails because of its use of variable packets for communication, reducing throughput and precluding synchronization signaling. A recent special-purpose addition to the Point Grey Research camera line from an affiliate company in Japan provides 25 fixed imagers using a single lane of the PCI-express bus[2]. Considering the growth in imager size and the requirements of large numbers of cameras for multi-viewpoint capture, we have to look to other technologies for high-bandwidth immersive 3D capture. In our work, we have selected to use a low-voltage differential-signalling (LVDS) protocol and to develop a hierarchic multiplexing architecture to support its multi-stream video capture.

3.2 Herodion multi-imager video capture

Herodion is HPLabs' 72-imager synchronized uncompressed capture system—designed and developed in collaboration with ISD in Athens[3]—operating at about 6Gbps over a 133-MHz PCI-X bus on a single PC [1]. Immersive video

[2] http://www.ptgrey.com/products/profusion25/index.asp
[3] http://www.isd.gr Products/Herodion

requires both lots of video and densely spaced capture. Most cameras cannot be positioned in such close proximity since lenses themselves are much larger. While optimal multi-viewpoint spacing would be about 3 mm[4]; our imagers permit about 20 mm—quite sufficient for initial studies. We know of no other camera system providing anywhere near this density, and none provides such bandwidth, synchronization, or flexibility. Considering the device: we use commodity cell-phone-type CMOS imagers[5] operating over 8-strand LVDS cable with Ethernet connectors and a communication protocol tied to an efficient DMA engine. Imagers are synchronized at the pixel level, and multiplexed in a line-coupled manner for maximum bus utilization and data throughput on a single PC.

Our imagers can be configured as desired, with optics and positioning to meet acquisition needs. We have configured these cameras for ultra-wide panoramic capture (achieving up to 24 Mpixel resolution at 30 fps), with varying foveas and arbitrary shapes,[6] and combined them for linear and 2D light field acquisition. Fig. 7 shows a raw camera configuration, and Fig. 8 presents one of our panoramic cameras (above) and a 24-imager linear 3D capture camera (below). Fig. 9 shows a 6-megapixel panoramic composite snapshot from the 30 Hz video produced by this camera. Market forces are continuously increasing imager resolution while reducing cost to give us ever-better devices at lower price.

3.3 Optimal multi-imager calibration

A requirement of using camera arrays for metric analysis such as this is that their intrinsics and relative poses be known. Little has been done in this area for large numbers of cameras, since few researchers have access to or need of such data. We have developed a special approach to this calibration that capitalizes on high quality planar homographies between pairs of imagers to develop a global optimal solution delivering epipoles and fundamental matrices simultaneously for an entire multi-imager camera system. The method exploits what we identify as the Rank-One-Perturbation-of-Identity (ROPI) structure of homologies in posing a unified estimator for the parameters. Our solution permits us—for the first time—to consider all calibration observation planes and multiple cameras simultaneously in framing a solution, which we solve robustly by SVD. The method works as follows. Homographies are formed in the usual way—although we use both line and point solutions [6]. For corresponding features l_A and l_B we have $l_B \sim \tilde{H} \cdot l_A$, with each providing 3 constraints (of which 2 are independent). We form a system of such constraints and solve for $\tilde{H}$.

$$\begin{bmatrix} a_B \\ b_B \\ c_B \end{bmatrix} \sim \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \cdot \begin{bmatrix} a_A \\ b_A \\ c_A \end{bmatrix} \Bigg| \begin{bmatrix} 0 & 0 & 0 & -c_B a_A & -c_B b_A & -c_B c_A & b_B a_A & b_B b_A & b_B c_A \\ c_B a_A & c_B b_A & c_B c_A & 0 & 0 & 0 & -a_B a_A & -a_B b_A & -a_B c_A \\ -b_B a_A & -b_B b_A & -b_B c_A & a_B a_A & a_B b_A & a_B c_A & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_9 \end{bmatrix} = 0$$

Considering initially the two-camera two-plane case, we form a special matrix H composed from two such $\tilde{H}$ matrices

Fig. 7. A 24-imager Herodion configuration.

Fig.8. A 7MP 24-imager FanCamera and 24-imager multi-view capture camera

[4] 0.1 degrees of separation for a viewing distance of about 2 meters—typical pupil size [7].
[5] Micron Imaging, http://www.aptina.com/products/image_sensors/mt9v032c12stc/#overview
[6] http://www.hpl.hp.com/personal/Harlyn_Baker/

that is a homology rather than a homography (see [2] for details). This matrix can be decomposed as $H = R + A = R + ev^T$ where R is a rotation matrix and ev^T is rank 1. In this situation, R is actually the identity, I, and we refer to H as a rank-one perturbation of an identity. e is an epipole, and v is the image of the line of intersection of the two planes used in computing the contributing $\tilde{H}$ matrices. We then form a simultaneous solution:

$$\begin{cases} {}^{i,j}H_1 = I_{3x3} + e_1\,{}^{i,j}v^T \\ \vdots \\ {}^{i,j}H_k = I_{3x3} + e_k\,{}^{i,j}v^T \\ \vdots \\ {}^{i,j}H_K = I_{3x3} + e_K\,{}^{i,j}v^T \end{cases}$$

which we can restructure as a least squares system $J = \sum_{k=1}^{K}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left\| {}^{i,j}H_k - I - e_k\,{}^{i,j}v^T \right\|_F^2$

and solve as:

$$\begin{pmatrix} \cdots {}^{i,j}H_1 - I \cdots \\ \cdots {}^{i,j}H_2 - I \cdots \\ \cdots \;\vdots\; \cdots \\ \cdots {}^{i,j}H_K - I \cdots \end{pmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_K \end{bmatrix} \begin{bmatrix} {}^{1,2}v^T & {}^{1,3}v^T & \cdots & {}^{n-1,n}v^T \end{bmatrix}$$

Since ev^T is rank 1, by Eckhart-Young[7] we can determine the minimum-error rank-one matrix of the right hand side. Optimal vectors e and v arise from its first left and right singular vectors. Multiple planes (and corresponding homographies $\tilde{H}$) are introduced to provide redundancy in determining the single epipole estimate e (we don't require the varying image intersection v's). From e and $\tilde{H}$ we can directly obtain a fundamental matrix relating the pairwise-contributing cameras: $F \cong [e']_\times \tilde{H}$ [2]. Projection matrices can be computed from the resulting fundamental matrix and epipole [6]. A metric camera matrix can be formed with the introduction of the intrinsics, *K*.

This discussion was with respect two cameras: we generalize to multiple cameras and multiple planes in determining robust global estimates for all epipoles (see [2]).

Fig. 9. Video-rate panoramic composite from the Herodion FanCamera.

4. TRANSFORMS FOR HORIZONTAL-PARALLAX AUTOSTEREOSCOPIC DISPLAY

Autostereo binocular display involves the simultaneous presentation to each eye of the imagery appropriate for its perspective. Numerous methods are available for delivering these data to the viewer, but a critical concern is that the images obey what is termed the *epipolar constraint*—their horizontal lines are positioned in correspondence so that the planes defined by the two centers of projection (the centers of the two eyes' optical systems) and any point in the viewed scene intersect the two (left and right eye) image forming surfaces in collinear lines that are parallel. The collinearity ensures that there is no vertical disparity across views; the parallelism is required so that—for *any* pair of views—the collinearity of any pair satisfies all pairs. Sufficient condition for this is that imagers share a common image plane and focal length and that their centers of projection lie along a line in space. It is difficult—perhaps infeasible—to build a multi-imager cameras system with this property. Optic variations, attachment of the imager to its mount, and numerous other physical opportunities for deviation from perfection make it most unlikely that we will have such configurations.

[7] http://voteview.com/ideal_point_Eckart_Young_Theorem.htm

Our challenge then is to determine how we can work with what we can build—to measure the variance from the desired configuration, to assess the error this brings, and to perhaps design corrective measures for situations where the implied error is beyond our operational limits.

4.1 Transitive rectifications

Consider a simple situation in horizontal-only autostereo display. A viewer's eyes are sampling the images from, say, camera 1 and camera 2 to achieve some 3D viewing perspective. Sampling of a different perspective comes from moving the head laterally so the eyes sample different camera images. Say the left eye now sees the data from camera 2 and the right eye sees the data from camera 3 (the head has moved to the right by one view zone). The alignment transforms (epipolar rectifications) that applied between camera 1 and camera 2 for collinear structuring induced some mapping $\Phi^2_{1,2}$ on camera 2. The transforms needed to rectify the camera pair 2-3 induces a mapping $\Phi^2_{2,3}$ on camera 2. It should be clear that if $\Phi^2_{2,3}$ differs from $\Phi^2_{1,2}$ then the scene will appear to change in the transition as the viewer moves. Fig. 10 shows a sampling of corresponding epipolar lines for a 9-by-2 multi-camera configuration, with colors coding the direction of adjacent constraints—the pairwise transforms differ significantly. Consider further the viewer moving forward from the original position so that his eyes intersect the images presented from camera 0 and camera 3, rather than being restricted to just the data from adjacent cameras 1 and 2. The viewer is sampling the lightfield generally. In this set of observed points (rays), the eyes will be seeing corresponding features from non-adjacent cameras—that is, the rectification employed originally in an adjacent pair-wise manner now must be consistent across non-adjacent views. Clearly, a single set of transforms to a common base is needed so each image is mapped once in serving all others. This is simple when generating views from CG models—just position the synthetic cameras along a straight line and aim them orthogonal to that line in the same direction with the same focal length. It is somewhat more challenging when working from acquired multi-viewpoint video imagery. This is one of the contributions of this paper: making do with the imperfections of a constructed multi-viewpoint camera. We describe a system for acquiring such multi-viewpoint video, structuring it for multi-viewer autostereo display, and present a facility that delivers this experience on a single workstation.

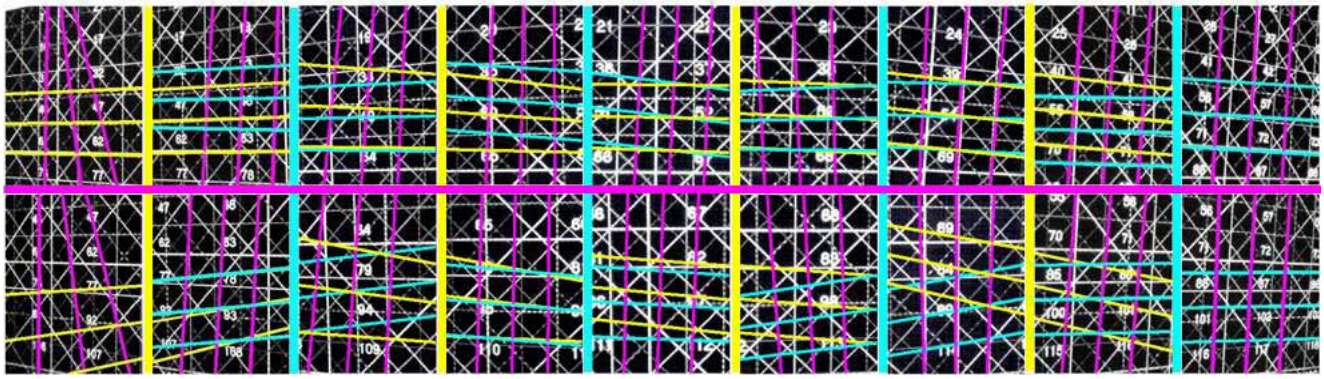

Fig. 10. Samples of corresponding epipolar lines from all adjacent pairwise mappings; yellow is odd-even pairs (1:2, 3:4, …), cyan is even-odd pairs (2:3,4:5, ...) and magenta is vertical pairs (1:1, 2:2, …).

4.2 Approximating the Ideal

We now deal with the issue of "correcting" the model to accommodate reality—the fact that we will not ever have perfect coplanarity and optics in our multi-imager system. We begin with initial estimates of the individual camera poses (Fig. 11). For evaluation purposes, we do this with both the ROPI based method above and using a grouped variant of the standard Zhang camera solver [15], augmented by a robust method for parameter estimation [11] that exploits the general version of the ROPI analysis we discussed above (details of this ROPI-Zhang comparative analysis will be presented elsewhere). We fit a line to the estimated camera centers. This defines the *x*-axis of the camera center-of-projection line. We chose a *z* direction as the mean of the estimated pose *z*-vectors, and define *y* as orthogonal to *x* and *z*, then readjust *z* to be orthogonal and minimizing the projective distortion in the acquired images. All image planes

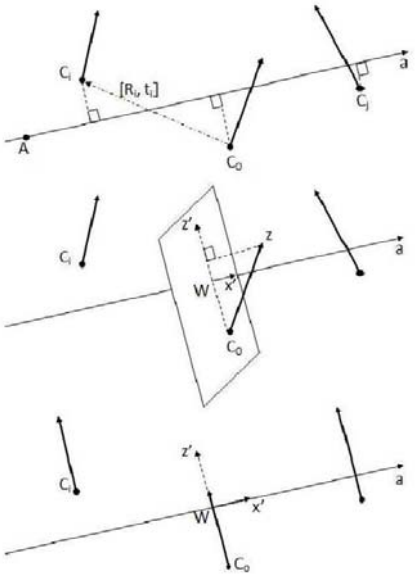

Fig. 11. Camera alignments: (top) computed poses and best-fit center line; (middle) re-alignments; (bottom) move CP path

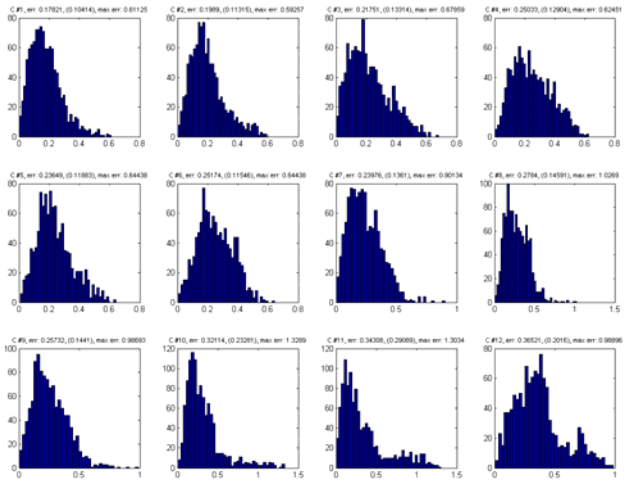

Fig. 12. Multi-imager epipolar error distributions

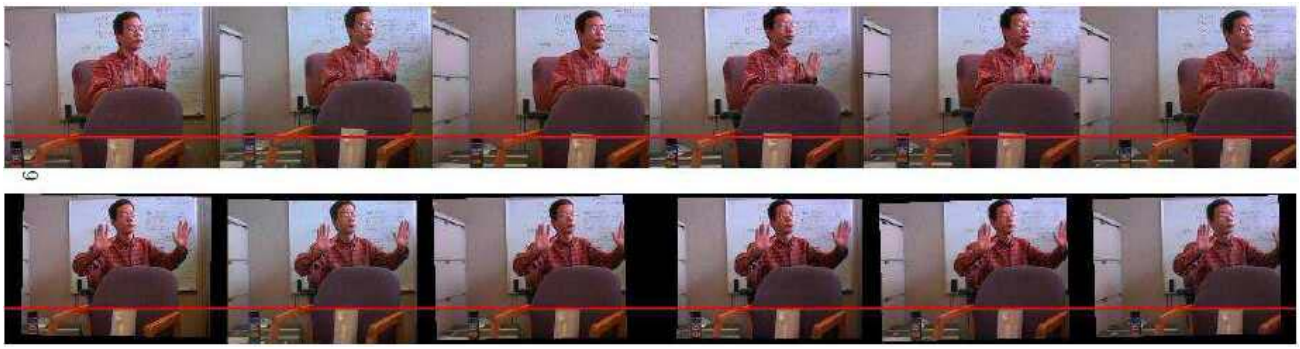

Fig. 13. Original and realigned images with sampled line.

are then reoriented about their centers [5] with remapped and resampled scanlines parallel. A final adjustment involves use of a non-linear optimization method to minimize the epipolar-line variance of calibration points as mapped through the various camera models. This adjusts the estimated rectifying homographies through application of another homography that works to bring corresponding points on the calibration checkerboard target to a common epipolar line (while avoiding degenerate mappings). Fig. 12 shows the distribution of resulting epipolar-alignment errors, and Fig. 13 shows originating and resampled images from six cameras of a 12-camera configuration to demonstrate the corrected mappings. The final stage of optimization decreased the mean scanline deviation by about 35% to a sixth of a pixel. Note that this analysis is also necessary for geometric reconstruction from such multi-imager data [4].

4.3 An autostereoscopic immersive pipeline

Our architecture involves full-framerate capture of multiple synchronized wide-VGA (752x480) imagers, writing of these images into GPU frame buffers with epipolar-based geometric reshaping and individual projector-camera color resampling, and splitting these to their destination projectors. Input multiplexing is handled by our Herodion hardware;

output multiplexing is done with Triplehead2Go[8] display adapters from Matrox. We can only drive twelve 800x600 projectors currently, so sample just 12 of our cameras. The software infrastructure is Nizza, a dataflow middleware application developed inside HP Labs that facilitates modular development of high performance multi-threaded and multi-core media processing and computational elements. Our current plan is to build up the camera and projector counts to permit a much wider field of operation, and to display the acquired live video on a life-sized surface with considerably increased area.

REFERENCES

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

[2] H. Baker, Z. Li, C. Papadas, "Exploiting Homologies in Global Calibration of a Multi-imager Array," 3DTV-Con (2007).

[3] T. Balogh, "Method and apparatus for producing 3D picture," U.S. Patent 5 801 761 (1998). http://www.holografika.com/

[4] R. Bolles, H. Baker, D. Marimont, "Epipolar-Plane Image Analysis: An Approach To Determining Structure From Motion," International Journal Computer Vision (1987).

[5] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," Machine Vision and Applications, 12-1 (2000).

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

[7] W. Ijsselsteijn, B. Harper, "Virtually There? A Vision of Presence Research," Presence–IST, EC Public Deliverable (ftp://ftp.cordis.lu/pub/ist/docs/fet/fetpr-4.pdf) (2001)

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

[9] D. Nguyen, J. Canny, "MultiView: Spatially Faithful Group Video Conferencing," Proc. CHI, ACM Press (2004).

[10] D. Nguyen, J. Canny, "MultiView: Improving Trust in Group Video Conferencing through Spatial Faithfulness," Proc. CHI (2007).

[11] R. Schreiber, Z. Li, H. Baker, "Robust Software for Computing Camera Motion Parameters," Journal Mathematical Imaging and Vision, Springer (2008).

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

[13] J.C Yang, M. Everett, C. Buehler, L. McMillan, "A real-time distributed light field camera," Proceedings of 13th Eurographics Workshop on Rendering, Eurographics Association (2002).

[14] C. Zhang, T. Chen, "A self-reconfigurable camera array," ACM SIGGRAPH Sketches (2004).

[15] Z. Zhang, "Flexible Camera Calibration by Viewing a plane from Unknown Orientations," ICCV (1999).

[8] http://www.matrox.com/graphics/en/products/gxm/th2go/