A Multi-imager Camera for Variable-Definition Video (XDTV)

H. Harlyn Baker and Donald Tanguay

Hewlett-Packard Laboratories, 1501 Page Mill Rd, Palo Alto, CA, USA
harlyn.baker/donald.tanguay@hp.com

Abstract. The enabling technologies of increasing PC bus bandwidth, multicore processors, and advanced graphics processors combined with a high-performance multi-image camera system are leading to new ways of considering video. We describe scalable varied-resolution video capture, presenting a novel method of generating multi-resolution dialable-shape panoramas, a line-based calibration method that achieves optimal multi-imager global registration across possibly disjoint views, and a technique for recasting mosaicking homographies for arbitrary planes. Results show synthesis of a 7.5 megapixel (MP) video stream from 22 synchronized uncompressed imagers operating at 30 Hz on a single PC.

1 Introduction

Several computing trends are creating new opportunities for video processing on commodity PCs: PC bus bandwidth is increasing from 4 Gbit/sec (PCI-X) to 40 Gbit/sec (PCI-e), with 80 Gbit/sec planned; Graphics cards have become powerful general-purpose parallel computing platforms supporting high-bandwidth display; Processors are available in multi-core multi-processor architectures for parallel execution. 16-lane PCI-express buses, for example, are standard equipment on many desktop, server, and workstation PCs; Nvidia's 3450 GPU is capable of multi teraflop of floating operations; HP and others sell dual-core dual-processor workstations for general computing uses, with Sun Microsystems offering 32 processing cores. The resulting central and peripheral processing power and data bandwidths make computers logical hosts for applications that once were either inconceivable, or required dedicated and inflexible hardware solutions.

While our investigations with these technologies are directed at development and use of camera arrays for multi-viewpoint and 3D capture, our initial achievements in combining the imagery from such a system—with high bandwidth multi-image capture, fast PC bus data transmission, and using GPUs for image manipulation—have been in building a novel ultra-high-resolution panoramic video camera that delivers varied levels of detail over its field of view. We term this XDTV for its flexible and user-selectable high-resolution format.

B. Gunsel et al. (Eds.): MRSC 2006, LNCS 4105, pp. 594 – 601, 2006.

2 Multi-imager Camera Array

The Herodion camera system (named after an amphitheater at the base of the Parthenon) is a high performance multi-imager CMOS capture system built around a direct memory access (DMA) PCI interface that streams synchronized uncompressed Bayer-format video to PC memory through a three-layer tree structure: an imager layer, a concentrator layer, and a frame grabber layer. Up to 24 imagers, grouped in 6's (see Figure 1), are attached to leaf concentrators. Two leaf concentrators connect to a middle concentrator, up to two of which can connect to the PCI-bus frame grabber. Different configurations are supported, down to a single imager. Details of the camera system can be found in an earlier publication [1]. In distinction to others [9], these data are uncompressed, synchronized at the pixel level, and running into a single PC. Redesign for PCI-X (for 8Gbps) is underway, which will give us 96 synchronized VGA streams. To support community developments in these areas, we have licensed the imaging system for commercial sale through our contractors [5].

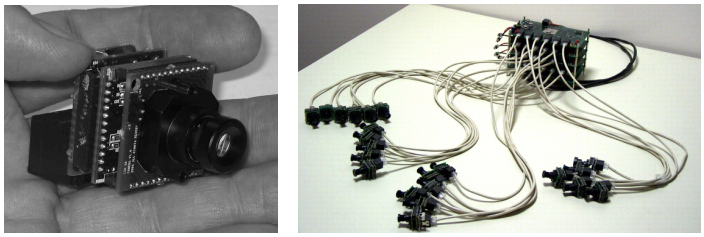


Fig. 1. Imager (left), and 24 attached to a bank of concentrators (right)

3 The Mosaic Camera

A mosaic camera is a composite of many other cameras—Figure 2 shows some of the mosaicking cameras we have built using the Herodion multi-imager system—we call them FanCameras. Calibration is the process of determining the parameters that map their data into a seamless image. We describe the calibration problem, formulate the solution as a global minimization, and describe how to calibrate in practice.

3.1 Mosaicking Methods

Most mosaicking methods use point correspondences to constrain image alignment. In digital photography, panoramic mosaics [2, 7] are derived from the motion of a single hand-held camera. In photogrammetry, aircraft and satellites capture images which are stitched together to produce photographic maps. Having large areas of overlap (typically 20-50%), these solutions are generally not effective for a rigid camera arrangement because this overlap reduces total resolution. They also typically depend on a scene's visual complexity since they require distinguishable features from the content itself. Because our imagers do not move relative to each other, we can

Fig. 2. Experimental FanCameras: 2x3 2MP; 2x9 6MP; 3x6 6MP; 2x9 concave 6MP; 22-imager variable resolution 7.5MP

calibrate the system beforehand, and can select calibration patterns to suit our needs. Note that rectification is with reference to a plane in the scene, and objects not on that plane will appear blurred or doubled, depending on their distance from the plane and the separation of the imagers.

3.2 Our Mosaicking Solution

The goal of our application is to produce high resolution video from a large number of imagers. We have narrowed our focus with two requirements. First, to produce resolution that scales linearly with the number of imagers, we fully utilize native resolution by "tiling" their data. That is, we require that imagers overlap only to ensure spatial continuity in the final mosaic. Super-resolution techniques, on the other hand, require complete overlap and face limits on resolution recovery [6]. Second, we facilitate an initial solution by using a linear homographic model for mapping each imager into a common reference plane to produce the final mosaic. This homographic model supports two types of scenario: those where the imagers have a common centre of projection (*i.e.*, pure imager rotation), and those where the scene is basically planar (*e.g.*, imaging a white board). For the latter scenario, we can increase the range of valid depths by increasing the distance from the camera to the scene, or by reducing the separation between imager centres. In fact, continued reduction in camera sizes increases the validity of the homographic model.

Our calibration method is as follows: (1) Place imagers in an arrangement so that they cover the desired field of view, while minimizing overlap (we have an automated computer-assisted-design (CAD) method in design for constructing these camera frames). (2) Select (automatically) one imager C_0 as the reference (typically, a central one), with its image plane becoming the reference plane of the final mosaic. (3) Estimate a homographic mapping ${}^{0}_{i}\mathbf{H}$ for each imager C_i that maps it into imager C_0 to produce a single coherent and consistent mosaic.

3.3 Line-Based Processing

We calibrate the camera using line correspondences. Lines bring two major benefits. First, a line can be estimated more accurately than a point. It can be localized across the gradient to subpixel accuracy [3], and the large spatial extent allows estimation from many observations along the line. Second, two imagers can observe the same line even without any overlapping imagery. This significantly increases the number of constraints on a solution because non-neighbouring imagers can have common observations.

Line (*l*) and point (*x*) correspondence homography solutions (*H*) are related by:

$$\begin{aligned} x' &= Hx \\ l' &= H^{-T} l \end{aligned} \tag{1}$$

With lines defined in Hessian normal form ($ax + by - c = 0$), constraints for line equations *l* and *l'* in two images—*l* = *(a,b,c)* and *l'* = *(a',b',c')*—enter the optimization as:

$$\begin{bmatrix} 0 & a'c & -a'b & 0 & b'c & -b'b & 0 & c'c & -c'b \\ -a'c & 0 & a'a & -b'c & 0 & b'a & -c'c & 0 & c'a \end{bmatrix} \tilde{h} = 0 \tag{2}$$

where $\tilde{h}$ is a linearized version of the line homography H^{-T} . There is a pair of such constraints for each line observed between two images. We can chain solutions for a pairwise-adjacent set of solutions through the array of images. While forming a reasonable initial estimate of the solution, these homographies yield global inaccuracies because they do not incorporate all available cross-image relationships—for example, they allow straight lines to bend across the mosaic. We have developed a bundle adjustment formulation to minimize the geometric error over the whole mosaic by simultaneously estimating both the homographies of each imager and the line models behind the observations. The nonlinear least-squares formulation is:

$$\underset{{}^0_i\hat{\mathbf{H}},\, {}^0\hat{\mathbf{l}}_j,\, i\neq 0}{\arg\min} \sum_{i,j} d\left({}^i\hat{\mathbf{l}}_j, {}^i\tilde{\mathbf{l}}_j\right)^2 \tag{3}$$

where the estimated parameters are the ideal point homographies ${}^0_i\hat{\mathbf{H}}$ from the imagers C_i to the reference imager C_0 and the ideal lines ${}^0\hat{\mathbf{l}}_j$, expressed in the coordinates of the reference imager. The function $d()$ measures the difference between ${}^i\tilde{\mathbf{l}}_j$, the observation of line $\mathbf{l}_j$ in imager C_i , and its corresponding estimate ${}^i\hat{\mathbf{l}}_j$. This measure can be any distance metric between two lines. We chose the perpendicular distance between the ideal line and the two endpoints of the measured line segments. This is a meaningful error metric in the mosaic space, and it is computationally simple. More specifically, we have selected $d()$ so that the bundle adjustment formulation of Equation 3 becomes

$$\underset{{}^0_i\hat{\mathbf{H}},\, {}^0\hat{\mathbf{l}}_j,\, i\neq 0}{\arg\min} \sum_{i,j} \left({}^0\hat{\mathbf{l}}_j^T \cdot {}^0_i\hat{\mathbf{H}} \cdot {}^i\mathbf{p}_j\right)^2 + \left({}^0\hat{\mathbf{l}}_j^T \cdot {}^0_i\hat{\mathbf{H}} \cdot {}^i\mathbf{q}_j\right)^2 \tag{4}$$

where ${}^{i}\mathbf{p}_j$ and ${}^{i}\mathbf{q}_j$ are the intersection points of the measured line with the boundary of the original source image.

The domain of the error function (Equation 4) has $D = 9(N_c - 1) + 3L$ dimensions, where N_c is the number of imagers and L is the number of lines. A typical setup has about 250 lines, or some 900 parameters. We use Levenberg-Marquardt minimization to find a solution to Equation 4, and a novel linear method to initialize the parameters—one which uses triples of constraints rather than the pairs of Equation 2, so that all abutting image relationships are considered.

3.4 Calibration Method

Digital projectors are employed as a calibration device. The projectors display a series of lines one at a time (see Figure 3). Imagers that see a line at the same instant j are actually viewing different parts of the same line $\mathbf{l}_j$ and therefore have a shared observation. Image analysis determines line equations ${}^{i}\tilde{\mathbf{l}}_j$. After all observations are collected, the solution of Equation 4 is found, paying careful attention to parameter normalization.

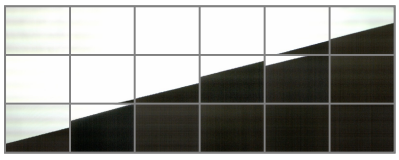


Fig. 3. A single calibration line pattern, as seen from 18 imagers.

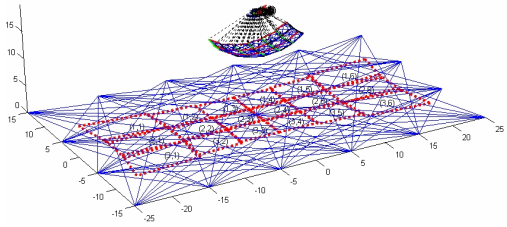


Fig. 4. Imaging geometry of the 3x6 mosaic camera. The 18 imagers are aligned using common observations of lines on a plane (shown in blue). The individual imager fields of view (outlined in red) illustrate minimal overlap.

A calibration process that presented coded lines simultaneously would be expeditious but, instead, we chose a method that provided reasonable throughput with sequential projection in our controlled-illumination laboratory space. Judicious use of observed constraints during the calibration allows us to reduce the lines presented to a minimal set that cover the space without bias. Figure 4 shows a simulation using a set of lines—each cast separately—for calibrating a 3x6 configuration. The results of the calibration are fed to the PC's GPU, and all video mappings occur there.

3.5 Redefinition of the Reference Plane

While we stated that features must lie on the reference (calibration) plane for blur-free mosaicking, we have developed a means to reposition this plane at will. Given two homographies, H_1 and H_2 relating two imagers I_a and I_b through two scene

planes Π_1 and Π_2, we can define a third homography, H_d, relating the imagers through an arbitrary third plane, Π_d. $H = H_2^{-1} \bullet H_1$ is an homography mapping I_a to I_b and then back to I_a. This transform has a fixed point that is the epipole e_a of the two imagers in I_a, and this fixed point is an eigenvector of H. Our desired arbitrary homography is $H_d = H_1 - e_b \bullet (x,y,z)$, where e_b is the image of e_a in I_b, and (x,y,z) is an expression for the desired plane Π_d = (a,b,c,d) with d normalized to be 1.0 [4]. From this, we derive H_d to suit our needs. Related, although less general, transformations have been computed using disparity [8]. Figure 6 shows a mosaicked image composed from 18 imagers (the 2x9 concave camera of Figure 2).

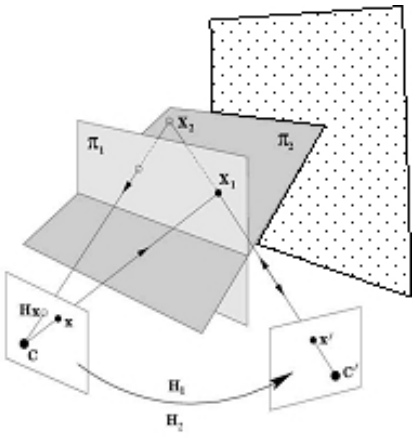


Fig. 5. Redefining the homography between two imagers based on two rectifying homographies and a third plane.

Fig. 6. The 2x9 concave camera of Figure 2 produces a wide-field-of-view 6 MP video stream, of which this is a sample image.

3.6 Variable Resolution Imaging

The camera we present here is one designed to observe a large flat surface—for example a work surface in a videoconferencing room—providing overall context imaging at a somewhat low resolution and detailed imaging of specific high-resolution (or "hotspot") locations where users may wish to place documents or other artifacts to be

shared. The work surface is shown in Figure 7 (top), and the camera designed to image it was shown in Figure 2, lower right. This camera has 22 imagers, grouped into three 2-by-3 hotspots with four wider-viewing imagers between them. The effect is to have the overall work surface imaged at one resolution, and the hotspots imaged in much higher quality. The camera is calibrated using projected lines, as described, with the resulting homographies being evaluated with respect to the test pattern in Figure 7 (bottom). Figure 8 (top) shows the viewed images of this test pattern, which are then positioned and blended as shown below. Figure 9 shows a frame from the live mosaic, with the inset detailing the transition visible between high and low resolution image fields.

This overhead-viewing variable-resolution camera runs at 30 Hz, delivering 60 pixels per inch of resolution at its three hotspots. These can be selected and windowed under program control, or through a user interface (gaming joystick – near subject in Figure 9). More interestingly, we are developing user interface gestures to command attention at these locales.

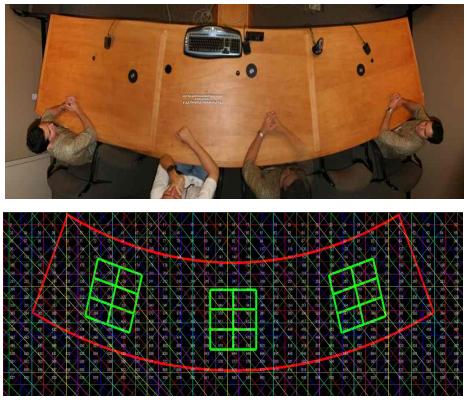


Fig. 7. The overhead view of scene to be captured at variable resolution (top); the calibration evaluation test pattern for this camera (bottom).

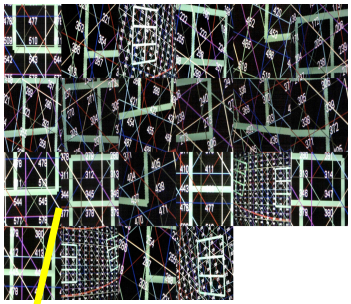
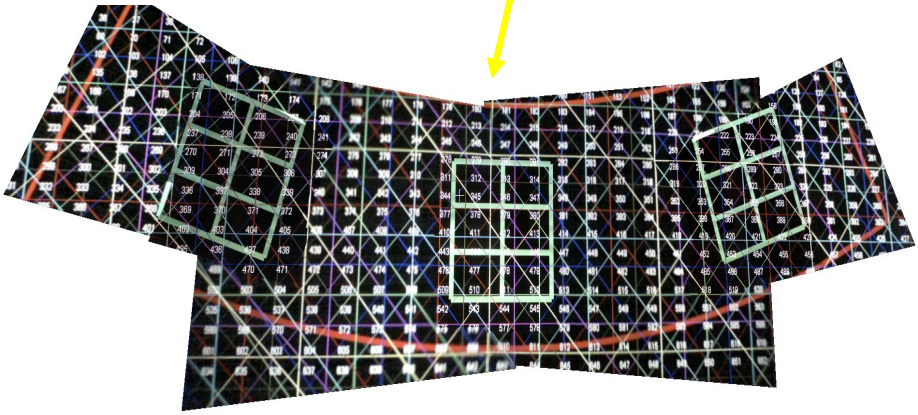


Fig. 8. 22-imager camera images as acquired (above); as mosaicked (below).



4 Conclusions

While demonstrating a specific use of this variable resolution imaging capability, we believe it is clear that many application areas may benefit from such flexibility in placing pixels. Surveillance and related monitoring tasks, where areas at varying levels of interest are under observation, are candidates for such imaging. Beyond planar capture, the multi-imaging capability presents benefit for large-scale scene observation, with the imagers aimed in varied directions for more thorough scene observation. And, of course, we intend to use this system for ongoing lab work in multi-viewpoint capture coupled to multi-viewpoint display.

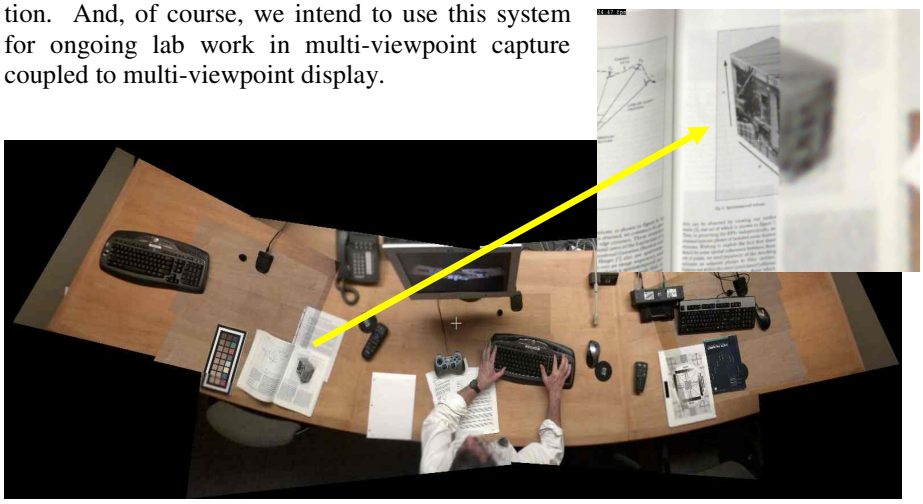


Fig. 9. 22-imager variable resolution camera images mosaicked, with detail of the high- versus low-resolution imaging at a boundary (PTZ'd and rotated by GPU).

References

1 Baker, H. Harlyn, D. Tanguay, C. Papadas. Multi-viewpoint uncompressed capture and mosaicking with a high-bandwidth PC camera array. In *Proc. IEEE Workshop on Omnidirectional Vision* (2005).

2 Burt, P.J., E.H. Adelson, "A multi-resolution spline with application to image mosaics," *ACM Trans. on Graphics*, 2:2 (1983).

3 Canny, J. A computational approach to edge detection. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8 (1986), 679-698.

4 Hartley, R., A. Zisserman, *Multiple View Geometry in Computer Vision,* Cambridge University Press, (2000).

5 Integrated Systems Development, S.A., Athens, Greece. http://www.isd.gr

6 Robinson, D., P. Milanfar. Statistical performance and analysis of super-resolution image reconstruction. In *Proceedings of Intl. Conf. on Image Processing* (2003).

7 Sawhney, H.S., S. Hsu, R. Kumar. Robust video mosaicing through topology inference and local to global alignment. In *Proc. 5th European Conference on Computer Vision*, vol. II, (1998) 103-119

8 Vaish, V., B. Wilburn, N. Joshi, M. Levoy, "Using Plane + Parallax for Calibrating Dense Camera Arrays," *IEEE Conf. Computer Vision and Pattern Recognition (*2004).

9 Wilburn, B., N. Joshi, V. Vaish, M. Levoy, M. Horowitz. High speed video using a dense camera array. In *Proc. Computer Vision Pattern Recognition* (2004).