https://doi.org/10.2352/ISSN.2470-1173.2019.3.SDA-636

EPIModules on a Geodesic: Toward 360º Light-Field Imaging

H. Harlyn Baker[1], Gregorij Kurillo[1], Allan Miller[1], Alessandro Temil[1], Tom Defanti[2], Dan Sandin[3]
[1]EPIImaging, LLC, Los Altos CA, [2]University of California, San Diego, and [3]University of Illinois, Chicago

Abstract

We discuss the design and operation of an Epipolar-Plane Imaging light-field camera array and embedded-computation system designed for panoramic operation. Built from a planar configured EPI module, this is a hybrid capable of imaging in 360º x 3D. We discuss the geometry and constraints of EPI imaging, its connection with the design and structuring of the acquisition array, and the architecture of the EPIModule forming the basis for our panoramic capture. Finally, we show an arrangement of modules on a geodesic – the Truncated Icosahedron of football/soccer – which happens to suit the optical and mechanical properties of the first EPIModule we have built. Results are preliminary; more will be presented for the conference publication.

Panoramic Capture and Scene Modeling

A current theme in visualization developments is the presentation of in-situ imagery prepared for experiential reviewing – virtual reality and its relatives. Most approaches to acquiring and showing such data depend on cylindrical viewing to capture the 360º aspect for a participant's rotating head with some amount of three-dimensional content positioning to permit a little head displacement for synthesizing the 3D nature of the scene through viewpoint accommodation. Typically, such 3D is inaccurate and often inadequate for the task (and therefor inducing perceptual conflicts and nausea), since it is formed from binocular – or perhaps slightly more – stereoscopic processing of overlapping views from the acquiring camera's adjacent FOVs. Our perspective here is that such stereoscopic processing, being probabilistic, is inherently flawed and fragile, and an approach that can deliver more accurate and precise 3D content is preferable, regardless of its departure from the traditions of binocular capture. Because of this, we present an approach to passive 360ºx3D capture based on a methodology that has been shown to have these characteristics.

Camera Arrays and EPI Analysis

Epipolar-Plane Image analysis (EPI) is an approach to multi-view stereo that bears the cost of employing many imagers but has advantages of accuracy through redundancy, precision through a selectable baseline, results structured by spatial continuity rather than isolated sets of points, and computation linear rather than exponential in number of cameras. First developed in the 80's [1], EPI awaited it's moment until the advances brought by the smart phone revolution – high-quality tiny-format inexpensive image sensors, similarly refined optics, and processors scaling with Moore's Law. The result was a situation where camera count, size, and price were no longer impediments to their deployment in large numbers.

EPI Structuring

EPI analysis exploits the ganged use of the epipolar constraint in partitioning a viewed scene into planar slices, forming a pencil passing through a common axis – the axis upon which the cameras' centers of projection lie. These scene slices map to individual lines in the projective imaging surfaces (epipolar lines), and these lines may be stacked to form an epipolar-plane image (EPI) capturing all that lies in that scene slice (see *Figure 1*). Scene features appear as linear streaks (epilines) in these EPIs, with one observation from each contributing epipolar line. The slope of epilines reveals the distance to their associated feature. The set of epilines in any EPI depict all features that its epipolar plane intersects in the scene. Through this, range determination maps to linear filtering (of one sort or another). Our colleagues at HCI in Germany developed a Structure Tensor approach [2] that explicitly sought the best estimate of this slope by a scatter analysis of the gradients through pixels in the EPIs at a reference image (the middle of the set of observing cameras). Our original EPI approach (and as employed here) fits lines to the Laplacian zero crossings of the various EPIs (there is one EPI for each rectified reference image scanline). So in our work epiline continuity replaces correspondence search. Continuity in the spatial dimension is used to produce connected observations in 3D – using the structure of the Laplacian zero crossing of the spatial reference image – to produce what we term String Clouds in distinction to Point Clouds (see *Figures 2 and 3)*.

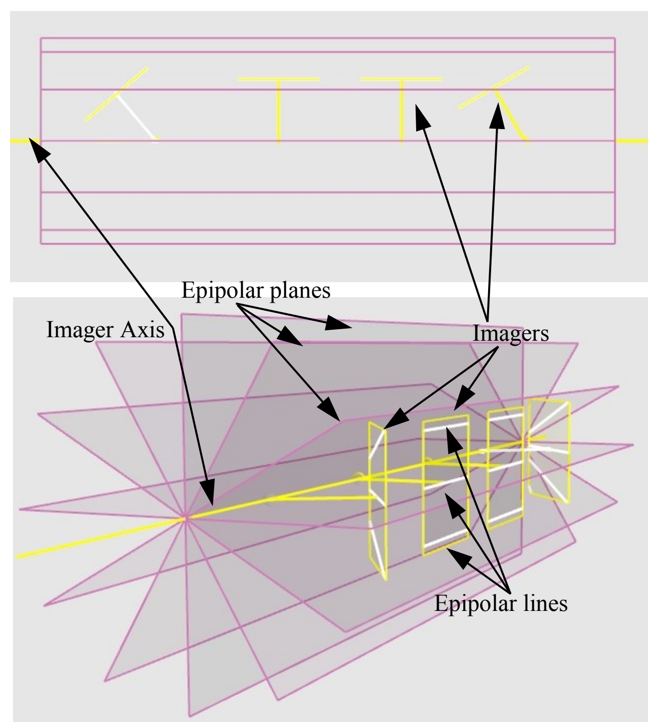


Figure 1. EPI Scene Partitioning: (top) overhead view; (bottom) side view.

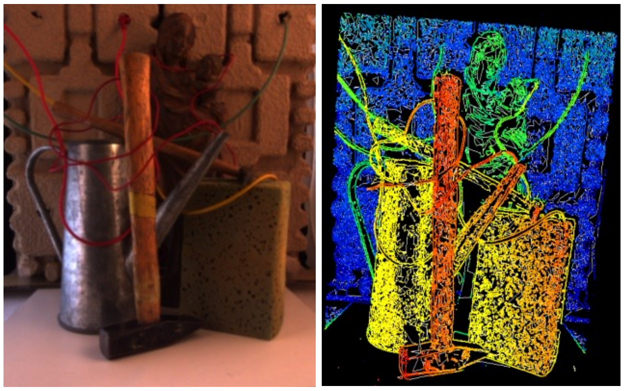


Figure 2. Typical range data from EPI on acquired imagery: String Clouds color coded by depth.

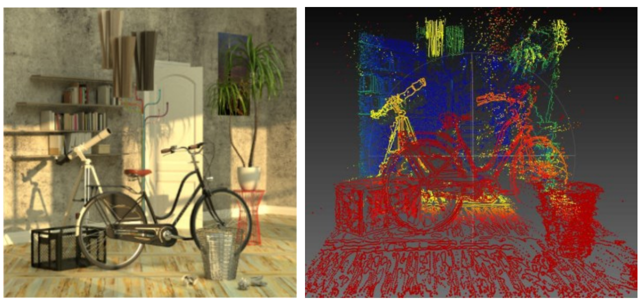


Figure 3. EPI String Clouds from a Light-Field Conference challenge dataset [4].

Designing for EPI Computation

Configuring discrete components for an EPI system leads to complex arrangements of many cables, connectors, components to adjust, and specialized hardware and software support for handling the multiple elements and their data streams. Further, discrete components may experience their own trajectories over time, making retention of calibration difficult. These factors impact manufacturability and maintainability of an EPI system. Having experienced these issues in the past, we chose to build an EPI camera that had no cables, minimal connectors, an architecture that provided all data as a unit, and that could benefit from fabrication advances such as wafer-level integration (WLI).

Implementation on an MPSoC

In addition to these usage concerns, we wish to configure imagers for specific task requirements and to combine modules for ganged use without major redesign effort. Seeking an imaging solution that could be placed in a variety of locations at a variety of scales ranging from centimeter-sized robot fingertips through meter-width light-field frames while keeping communication and processing bandwidth manageable, we looked for a processing system that could unify the processing and packaging. For this we selected the multi-processor-system-on-chip (MPSoC) Zynq UltraScale+ solution from Xilinx[1].

The Platform

This Zynq MPSoC has a quad-core 64-bit Cortex-A53 ARM, a dual-core 32-bit Cortex-R5 ARM, a Mali-400 GPU, and ~600K-logic-cell FPGA in a 900 pin package. We add to this 8 GB of RAM (but designed for 16), micro-SD flash (up to 128 GB), and gigabit Ethernet and Thunderbolt3 for IO. In our initial build – and to support placement of different imager sets on the same computational platform – we configured the system as two PCBs; a camera head with individual data/control lines plus synchronization clocking, and a processor PCB with the Zynq, whose programmable logic controls the camera and receives its outputs. The two boards snap together. FPGA code organizes the streams into stacked frames – gamma mapped for quality signal retention after byte-level compression – and puts it on the bus to memory and the processors.

Figure 4. An EPIModule: camera head PCB at left, processor PCB at right.

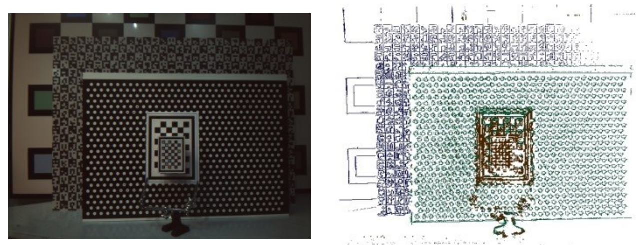


Figure 5. Color-coded range String Clouds from an EPIModule; estimates are feature-based, located at intensity discontinuities (zero crossings of a Gaussian Laplacian).

We run a Xilinx Linux variant (Petalinux) that contains our code supported with OpenCV and OPENGL functionality, and drivers to handle the various data streams (imagery in, storage, range results, etc., out).

The camera is comprised of seventeen Aptina AR0135 image sensors at 1.3 megapixels, configured in a plus shape sharing a central reference view (*Figure 4*). The shared intersection view lets us combine estimates of more vertical features (as detected with a horizontal band of imagers) and more horizontal features (as detected with a vertical band of imagers), with estimates formed with respect to the central imager. This orthogonal image-set capture enables more accurate and precise estimation of features oriented along either of the baselines. With our estimation approach, the shared imager can be anywhere since we can choose at will which is to be the reference; in Structure Tensor the center tap of the gradient kernels defines the frame of reference. We could, for example, have L's or T's as well. Figure 5 shows the reference image from a frame whose range is shown color-coded to its right. Notice that range is defined at edges – zero crossings – typically filling about 10% of the image. By definition these are the most localizable elements of a scene – others are, in a sense, hallucinations since they lack contrast.

---

[1] XCZU9EG-1FFVC900E

Geodesic EPIModule

Our EPIModules are designed to deliver a point cloud of 3D feature estimates within a roughly 64ºx48º FOV. In collaboration with the Electronic Visualization Laboratory at the University of Illinois Chicago and the University of California San Diego's Qualcomm Institute [3] we designed and built three of these to be configured on a partial geodesic aimed toward capturing 3D in full 360º. For this, we combine modules in a designed framework to form the desired shape. The modules run independently, or in a master/slave arrangement where synch control for all comes from one.

The geodesic form seemed most appropriate as it tiles the sphere in a uniform manner and, if the geometrical and optical properties are aligned, can deliver the desired coverage in a compact form. While we have been constrained by the FOV built into our modules which, given the 4π steradians required, could present a challenge, experimentation and simulation have brought us to selecting a truncated icosahedron. This is comprised of 20 hexagons and 12 pentagons – a standard soccer ball.

Designing within this, we find we can cover the full 4π steradians at about three meters and beyond through judicious selection of imager orientations on the various facets. For our three-facet contribution to the project we chose an adjacent triple as shown in *Figure 6*, and as constructed in *Figure 7*.

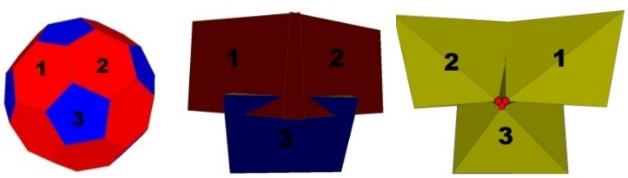


Figure 6. *Truncated Icosahedron, pentagons blue, hexagons red. Three selected frusta from front (center), rear (right).*

Figure 7. Three populated faces of geodesic.

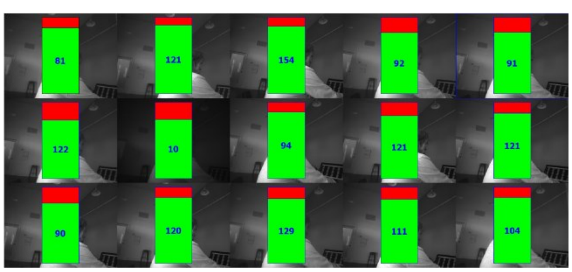


Figure 8. Focus tool: green indicates current focus at imager, red is best so far, value is contrast score.

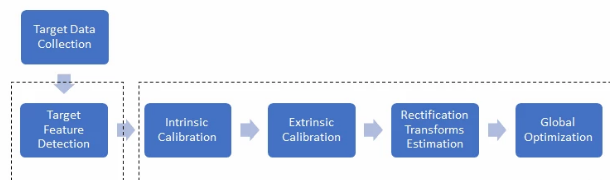
Capture Experiments

We focus the array using a graphical feedback process that shows us the relative contrast at each imager. Simple focus adjustments let us rapidly do as well as or better than the human eye (*Figure 8*) – we turn the lens until the green fills the superpositioned red rectangle. This interface connects the camera to its visualization package, including calibration facility, individual and collective register settings, point cloud visualize-ations, etc.

Calibration

Calibrating the focused individual 2D camera arrays and their groupings is a critical part of the system. We build on prior developments [5] in determining the intrinsics and extrinsics of the module imagers, determining the optimal rectification and resampling parameters to minimize epipolar alignment deviations. This consists of several steps:

1. Determining intrinsic parameters of individual cameras (e.g. lens distortion, focal length, optical center).
2. Determining extrinsic parameters defining relative position and orientation of each camera in the array.
3. Calculating and refining rectifying homographies. To bring the set into epipolar alignment.



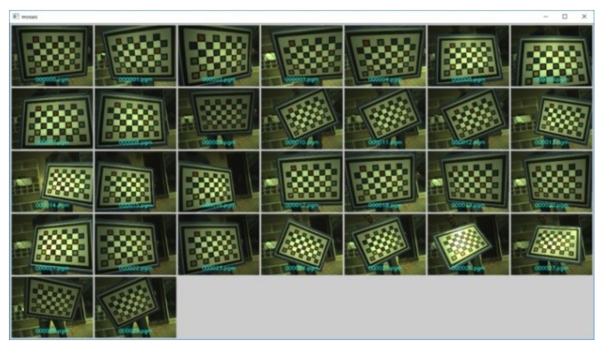


Figure 9. Sample views of the calibration target.

We do this jointly for the horizontal and vertical sets of imagers. These analyses follow this structure:

Each calibration requires a few dozen or so frames from the full set, which appear in *Figure 9* seen from the reference-imager's perspective. A cross-validation with target images captured over greater depth and orientation ranges tells us that we get better intrinsics when the imagery fill the fields of view, while we get better extrinsics and global depth measurements when we position the target at a larger range of depths.

With the individual EPIModules calibrated, we then perform a cross-module calibration/registration that manages to estimate EPIModule relative poses through extrapolation and cross validation, despite lacking views of the full target [5]. Figure10 shows reference views from modules 2, 1, and 3 of Figure 7, and Figure11 shows their individual string clouds, and integrated as a combined string cloud.

Expected Performance and Resolutions

The 1280x1024 imagers give about 20 pixels per degree spatial resolution for their range maps, with Laplacian edge localization of about 5 bits subpixel. This is expected to give us adequate detail localization for viewer comfort. Depth precision with this baseline (8 cm) is about 0.1% of depth (see *Figures 12 and 13*).

Other Considerations

While the effort here addressed panoramic capture over the sphere, we are also configuring EPIModules for planar capture – tiled as large integrating apertures – for traditional light-field capture. This work has been supported by NSF SBIR Phase I award 1648388 to EPIImaging, LLC and a contract from UIC under NSF award 1456638 **(**Sensor Environment Imaging Instrument – SENSEI**).**

Figure 10. Registered imagery for combining three EPIModule data (#2 left, #1 right, and #3 lower of Figures 6 and 7).

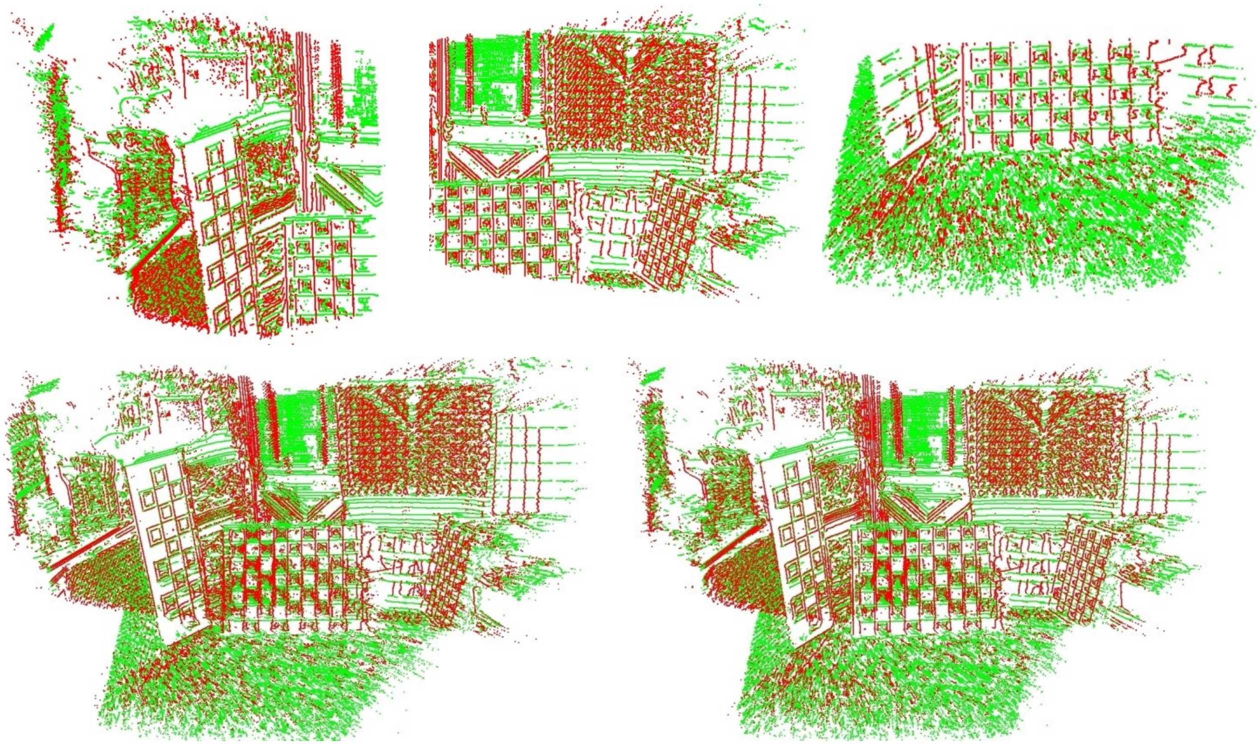


Figure 11. Individual string clouds (above) and aligned point cloud (below) from the EPImodules images of Figure 10.
Green features are from the vertical EPIModule imagers (S axis), red are from the horizontal EPIModule imagers (T axis).
The bottom figure is for crossed-eye stereo viewing.

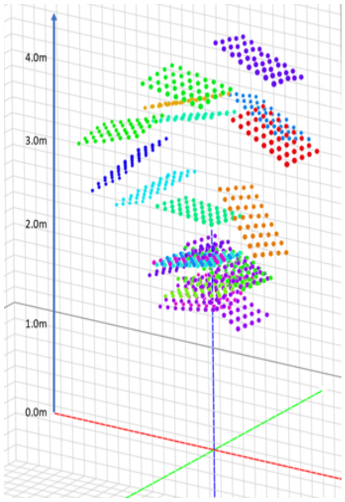


Figure 12. Location, orientation, and planarity of calibration target

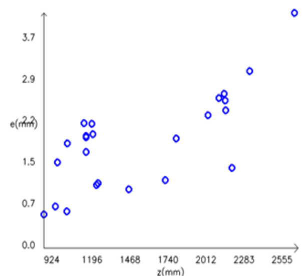


Figure 13. Mean absolute planarity

References

[1] R.C. Bolles, H. H. Baker., D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," International Journal of Computer Vision, 1:1, 1987.

[2] S. Wanner, J. Fehr, B. Jähne. "Generating EPI representations of 4D light fields with a single lens focused plenoptic camera," *Advances in Visual Computing*, Springer Berlin Heidelberg, 2011.

[3] NSF SENSEI Award to Electronic Visualization Laboratory, University of Illinois, Chicago, 2014. https://www.nsf.gov/awardsearch/showAward?AWD_ID=1456638

[4] CVPR LCV4LF dataset, 2017. http://hci-lightfield.iwr.uni-heidelberg.de/

[5] G. Kurillo, H. Baker, Z. Li and R. Bajcsy, "Geometric and Color Calibration of Multiview Panoramic Cameras for Life-Size 3D Immersive Video," *2013 International Conference on 3D Vision - 3DV*, Seattle, WA, 2013.

Author Biographies

Baker has degrees from the University of Western Ontario, Edinburgh University, University of Illinois at Urbana-Champaign, and has been on staff at Edinburgh, Stanford, SRI International, Interval Research, HP Labs, Heidelberg University and University of Illinois, Chicago. He, Kurillo (research engineer at UC Berkeley and UC Davis), Miller (serial entrepreneur), and Temil (hardware/software design engineer), have been running EPIImaging, LLC, a light-field camera startup since 2013. Defanti is a research scientist at Qualcomm Institute, UCSD, and a distinguished professor emeritus at UIC. Sandin is a senior research scientist and professor emeritus at the School of Art and Design UIC.