

# Using Planar Point Correspondence to Calibrate Camera Arrays for Light Field Acquisition

Ashley W. Stewart, Donald G. Dansereau

Queensland University of Technology, Australia; Stanford University  
a.stewart.au@gmail.com, donald.dansereau@gmail.com

## Abstract

Light field cameras are an emerging technology with unique post-processing capabilities. For certain light field applications such as synthetic aperture photography, existing calibration procedures are either metric calibrations that are difficult to execute with low-cost hardware, or non-metric procedures that are inflexible to arbitrary sub-camera pose or require cameras with tightly controlled poses. We present a novel and comparatively simple non-metric procedure for light field acquisition that estimates the geometric transforms between camera images with respect to a calibration plane. The procedure is suitable for mobile camera arrays, and flexible to unknown or varied sub-camera pose. It only requires one image from each camera of a calibration pattern positioned to span the camera array's full field of view. We also provide a quantitative measure of calibration quality, and use it to demonstrate the procedure's efficacy with our Raspberry Pi camera array. The results highlight the procedure's robustness to variable camera orientation in contrast to existing state of the art techniques. Finally, we present qualitative results by rendering light fields at varying levels of focus and occlusion, and demonstrate success in capturing and rendering light field video.

## 1 Introduction

*Light fields* describe the amount of light flowing through space in all directions. A light field *camera* captures an array of views of a scene using either an array of cameras [Yang *et al.*, 2002], a lenslet array [Ng *et al.*, 2005], or a single camera on a controlled gantry [Levoy and Hanrahan, 1996]. The resultant views are aligned using homographies so that light rays can be identified via a two-plane parametrisation consisting of a camera plane with cameras in  $s, t$  and an image plane with pixels in  $u, v$ .

Some applications of light fields include image-based rendering [Levoy and Hanrahan, 1996] and 3D geometry estimation [Wanner and Goldluecke, 2014]. Technologies such as 3D light field displays are also emerging [Chen *et al.*, 2014]. A popular application of light field cameras is synthetic aperture photography [Levoy *et al.*, 2004], which involves projecting images onto a focal plane and computing and rendering their average, enabling post-capture refocussing. With a sufficiently wide synthetic aperture, synthetic focussing can blur nearby occluders until they fade from view.

To achieve results in any such application, camera calibration is essential. Calibration processes have been described for monocular cameras since the early 1970s, first in photogrammetry [Brown, 1971], and later in computer vision [Ganapathy, 1984]. Monocular camera calibration involves estimating intrinsic and extrinsic camera parameters so that images can be later rectified for accurate analysis and rendering. Early light field capture was dominated by calibrated monocular cameras that travel along controlled gantries. Although this technique allows the transformations between viewpoints to be trivially calculated, it severely limits applications outside the laboratory, and restricts capture to static scenes. Camera arrays and lenslet arrays have greater potential, though finding the transformations between viewpoints becomes non-trivial, especially when orientations differ. As a result, calibrations for camera arrays and lenslet arrays have been developed that recover sub-camera poses.

Most camera array calibrations involve applying an extension to Zhang's monocular calibration [2000] across all viewpoints with additional optimisation steps [Ueshiba and Tomita, 2003; Xu *et al.*, 2014]. Calibrating camera arrays using such processes is challenging, especially outside the laboratory and for mobile arrays.

Vaish *et al.* [2004] point out that for many light field applications such as synthetic aperture photography, a metric calibration that recovers camera parameters is unnecessary. Instead, only the relative positions of viewpoints are needed to a scale to compute the projective transforms

necessary to render an image focussed at a given focal plane. Vaish achieves this by measuring the parallax of several scene points across all cameras and computing the nearest rank-1 factorization via Singular Value Decomposition (SVD). This non-metric procedure was shown to produce better qualitative results for synthetic aperture photography applications than those produced through metric calibrations.

Although Vaish’s procedure is conceptually simple, it assumes that all camera images are aligned at a reference plane. This allows scene point parallax in images to be considered a function of relative camera position. Non-uniform orientation among cameras voids this relationship as parallax becomes a function of camera orientation in addition to position. This makes camera positions irrecoverable by means of parallax measurement. This is a significant problem for low-cost consumer cameras prone to inconsistent construction. Our Raspberry Pi camera array, for example, was built to maintain front-facing orientations by mounting cameras onto an aluminium plate, yet variable orientation is still exhibited within their fixed camera boards.

### 1.1 Contributions

We propose a novel calibration solution that recovers the geometric transformations between images with respect to a calibration plane. Such transformations are useful because they are a function of relative camera *pose* rather than *position*. Our method can be achieved using only a single image from all cameras of a calibration pattern spanning the camera array’s full field of view. Transformations are estimated using feature detection and sample consensus algorithms. Another advantage of our solution is that it largely relies on prevalent algorithms that already exist as standard tools in most computer vision libraries, so implementing it requires few lines of code.

To support the use of our calibration, we also present a quantitative measure of calibration quality. Vaish et al. [2004] note that the usual indicators of calibration accuracy such as reprojection error cannot be measured for non-metric calibrations. Our measure of error is the total distance between SURF feature points in calibration images after projecting them onto the calibration plane using their estimated relative transformations. We have achieved a significant improvement in SURF feature consistency from our calibration compared to Vaish’s. This will be true for any camera array with variable camera orientation.

## 2 Related Work

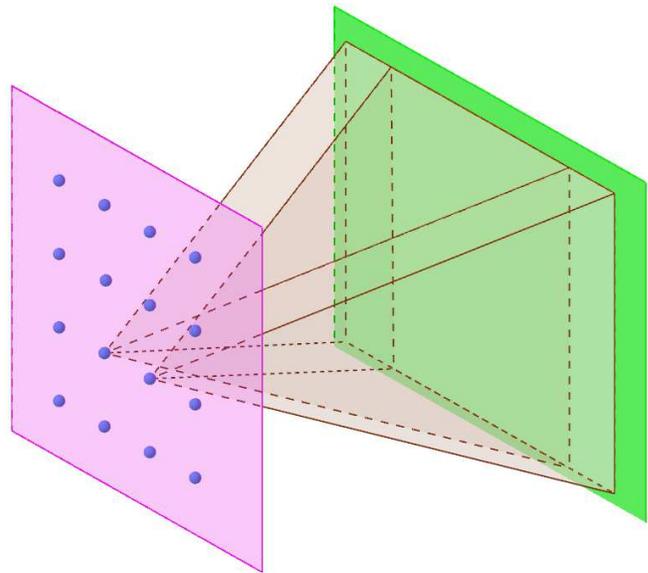
Dansereau et al. [2014] point out that image sets suffering from parallax can be used to construct light fields by estimating the geometric transformations between images and reprojecting them onto a common plane. This is essentially what we have recognised and are taking advantage of

in our method. However, theirs requires an initial estimate of camera pose to be known, parametrised by azimuths and elevations. This is possible when the pose of cameras is tightly controlled, such as in their Ocular Robotics *RobotEye* camera system. The poses of our cameras are not tightly controlled on their camera boards, so sub-camera poses cannot be recovered without performing a metric calibration that estimates camera parameters.

Vaish et al. [2004] also state that a possible area of future work is investigating extensions to their technique for cameras in general positions, as well as arbitrary reference planes using planar homologies. Following this suggestion, we have identified a technique to align camera images using planar point correspondence for cameras with variable orientation.

## 3 Calibration

Consider a camera array consisting of cameras with unknown poses, and a non-repeating planar calibration pattern positioned to span the array’s full field of view (see Figure 1). The image set captured in this scenario can be brought into focus at the calibration plane as a light field by applying geometric transformations to the image set. Our approach will estimate these transformations.



**Figure 1:** Example calibration setup with camera plane on the left, and calibration plane on the right. Between the planes is the projection of two adjacent camera views with significant overlap. Our procedure requires that it be possible to position a calibration pattern that spans the array’s full field of view with significant overlap between adjacent views.

### 3.1 Choosing a Calibration Pattern

The calibration pattern must be non-repeating so that erroneous feature matches are minimised. This means that common calibration patterns such as checkerboards are unsuitable. Additionally, the pattern must be sufficiently detailed so that feature identification is strongly encouraged. Li et al. [2013] have reverse engineered a pattern generator that yields high quantities of detectable features using random noise. The generator executes in multiple passes at varied scales so that plentiful features can be detected at a range of distances. Calibration patterns generated this way are suitable. We have also achieved good results using certain detailed paintings and posters. Though results may vary using paintings and posters, it is a convenient alternative to printing a potentially large pattern.

### 3.2 Transform Estimation and Image Rectification

Once a calibration image has been captured by each camera, feature matching can begin. We identify features across all views using Speeded-Up Robust Features (SURF) [Bay et al., 2008]. This generates a point matrix and a feature descriptor matrix for each image. Unique, matching points between adjacent image pairs are then collected if the error between feature vectors is within a threshold. Not all image pairs need to be compared; we compare only successive and adjacent image pairs (i.e. with an image set  $I$ , we match features in  $I_0 \leftrightarrow I_1, I_1 \leftrightarrow I_2, \dots$ ). The sets of points matched between each successive image pair become the main inputs into the Maximum Likelihood Estimation Sample Consensus algorithm (MLE SAC) [Torr and Zisserman, 2000], which provides an initial set of geometric transformations. MLE SAC uses the same strategy as the more common RANSAC [Fischler and Bolles, 1981], but chooses solutions according to maximum likelihood rather than the number of inliers, overcoming non-linear constraints between parameters.

The transformations returned by MLE SAC will be relative to the first camera’s image, causing a progressive distortion of the later images. This can be resolved by choosing an alternative image as the anchor and applying its inverse transform to all the others, so that the anchor image becomes the least distorted. We measure the  $u, v$  limits of each image after projecting them onto the calibration plane via their estimated transforms to identify the central image and use it as the anchor. Choosing the central image as the anchor will reduce overall distortion.

The final set of transformations can be used to rectify any image set captured by the array into light field alignment at the calibration plane.

### 3.3 Synthetic Aperture Focussing

If relative camera positions are known, then images can be translated into alignment at any plane parallel to the calibration plane for synthetic focussing. With a desired fo-

cal depth  $d$  relative to the calibration plane, and cameras with relative positions in  $\Delta P$ , each camera image  $C_i$  should each be translated by  $-d\Delta P_i$  to bring them into alignment [Vaish et al., 2004]. Results are rendered by taking the average of all translated images.

## 4 Implementation and Results

Our Raspberry Pi camera array uses Raspberry Pi V1 camera modules arranged in a 4x4 grid (see Figure 2). Their orientations are unknown, though they are known to be non-uniform. Our cameras are approximately planar and evenly spaced.



Figure 2: Our Raspberry Pi Camera Array

Although we suggest using a calibration pattern generated via noise functions [Li et al., 2013], we demonstrate the flexibility of our solution by using an image of a painting displayed on a TV positioned 300mm in front of the camera array. The painting used is Leonid Afremov’s *Farewell to Anger*. Our horizontal and vertical sub-camera fields of view are  $\phi = (53.5^\circ, 41.41^\circ)$ , and distances between cameras are each roughly 34mm. This makes our horizontal and vertical  $u, v$  overlaps on the calibration plane approximately 89% and 85%.

Our qualitative results include focussed light field stills, including animations of stills in which the level of focus varies between planes of interest (see Figure 3). Notice that the objects in the focal plane are clear, while the objects elsewhere appear blurred, and occluders fade from view when the focus is on the background. Clear focus levels are indicative of a high quality calibration.

We have also captured light field video of indoor and outdoor scenes with panning focus and moving objects (see Figure 4). Our synthetic focussing results and robustness to occluders compares well with the results others have achieved using similar procedures.



(a) Unrectified image from one of the Raspberry Pi V1 camera modules



(b) Light field focussed on the office chair in the background



(c) Light field focussed on the occluding hand in the foreground

**Figure 3:** Rendered light field still demonstration displaying an office scene at two levels of focus with an occluding hand (focus animations available online) [Stewart and Dansereau, 2017]



(a) Video 1, background in focus



(b) Video 1, foreground in focus



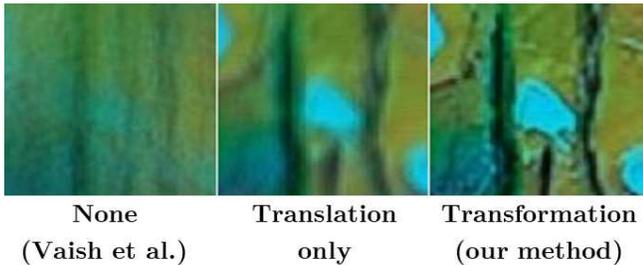
(c) Video 2, background in focus



(d) Video 2, foreground in focus

**Figure 4:** Screenshots from some of our light field videos (full videos available online) [Stewart and Dansereau, 2017]

To qualitatively demonstrate the effectiveness of applying relative view transformations to handle non-uniform orientation, we compare the visual clarity of synthetic focussing results rendered with no alignment, alignment by translation only, and alignment by transformation (see Figure 5). No alignment is akin to applying Vaish et al.’s calibration, ignoring varied camera orientation. We compare our results with Vaish even though their method is not well suited to our implementation, because it is the only other non-metric calibration that we are aware of. Alignment by translation is a more primitive version of our technique that applies only translations to align images at a point. The results highlight the benefits of applying the more complete transformations recovered by our procedure.



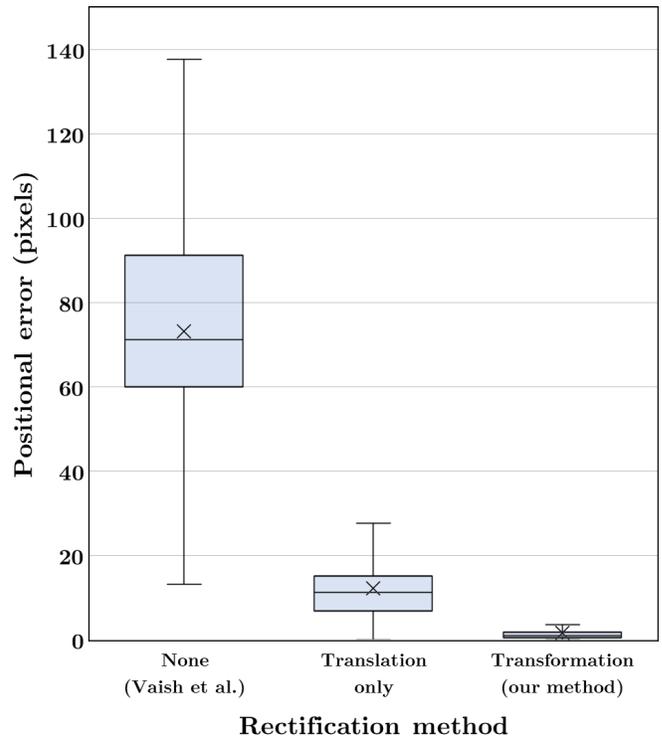
**Figure 5:** Close-up of synthetic focussing results using three rectification techniques. Notice the progressive increase in visual clarity, the best of which was achieved by our geometric transformation method.

We obtain quantitative results by measuring the pixel distance between matching SURF features in calibration images projected onto the calibration plane. This provides a reliable measure of focal error. We compare the results of the same three rectification techniques as in our qualitative comparison (see Figure 6).

We found that running multiple passes of our procedure yields more precise view transformations. Multiple passes can be executed by running the procedure successive times against the rectified calibration set. For our setup, two passes yield a significant improvement, with diminishing returns achieved thereafter (see Table 1).

**Table 1:** Feature error across our aligned calibration images after several calibration passes.

Pass #	Average <i>u</i> -feature error (pixels)	Average <i>v</i> -feature error (pixels)
1	5.092	5.155
2	1.178	0.884
3	1.184	0.917
4	1.160	0.925



**Figure 6:** Positional error of SURF features measured in an image set for three rectification techniques. Notice the progressive decline in error, the lowest of which was achieved by our geometric transformation method.

## 5 Conclusions and Future Work

As research in light field technology has progressed and applications have expanded, so too has the need for robust calibrations. Our calibration procedure can be carried out with camera arrays constructed with low-cost hardware. It is also exceptionally robust; the only restriction on the camera array itself is that it must be possible to position a calibration pattern that spans its full field of view, and that sufficient features can be detected and matched between images. Our calibration is also remarkably easy to implement, with most of the processes needed (SURF, MLESAC, and noise generators) already built into current standard computer vision libraries and tools such as OpenCV and MATLAB.

Our qualitative aperture focussing results compare well with those achieved using similar, non-metric calibrations. Vaish et al. [2004] have also shown that non-metric calibrations that minimise parallax on a reference plane produce better qualitative results than metric calibrations. Our quantitative results highlight our procedure’s robustness to non-uniform orientation.

Finally, we also demonstrate the capture and rendering of light field video using low-cost hardware. Currently, little work has been done on light field technology that exploits

or identifies the unique properties of the temporal axis. Our calibration enables even the most entry-level technology to be used in this exciting area. Work in synthetic aperture photography, for instance, shows that objects can be tracked in light field video through dense occluders [Joshi *et al.*, 2007].

## Acknowledgements

We would like to express our gratitude to the Australian Research Council Centre of Excellence for Robotic Vision for supporting this project (project number CE140100016). We would also like to extend a special thanks to Rafe Denham for his work on the camera array, and finally to Steven Martin for his assistance in capturing outdoor light field video, and for his alterations to the camera array.

## References

- [Bay *et al.*, 2008] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.
- [Brown, 1971] Duane C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37:855–866, 1971.
- [Chen *et al.*, 2014] Renjie Chen, Andrew Maimone, Henry Fuchs, Ramesh Raskar, and Gordon Wetzstein. Wide field of view compressive light field display using a multilayer architecture and tracked viewers. *Journal of the Society for Information Display*, 22(10):525–534, 2014.
- [Dansereau *et al.*, 2014] Donald G. Dansereau, David Wood, Sebastian Montabone, and Stefan B. Williams. Exploiting parallax in panoramic capture to construct light fields. In *Proceedings of Australasian Conference on Robotics and Automation*, pages 1–9. Australian Robotics & Automation Association ARAA, 2014.
- [Fischler and Bolles, 1981] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [Ganapathy, 1984] Sundaram Ganapathy. Decomposition of transformation matrices for robot vision. *Pattern Recognition Letters*, 2(6):401–412, December 1984.
- [Joshi *et al.*, 2007] Neel Joshi, Shai Avidan, Wojciech Matusik, and David J Kriegman. Synthetic aperture tracking: tracking through occlusions. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.
- [Levoy and Hanrahan, 1996] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.
- [Levoy *et al.*, 2004] Marc Levoy, Billy Chen, Vaibhav Vaish, Mark Horowitz, Ian McDowall, and Mark Bolas. Synthetic aperture confocal imaging. In *ACM Transactions on Graphics (ToG)*, volume 23, pages 825–834. ACM, 2004.
- [Li *et al.*, 2013] Bo Li, Lionel Heng, Kevin Koser, and Marc Pollefeys. A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 1301–1307. IEEE, 2013.
- [Ng *et al.*, 2005] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11, 2005.
- [Stewart and Dansereau, 2017] Ashley W. Stewart and Donald G. Dansereau. Using Planar Point Correspondence to Calibrate Camera Arrays for Light Field Acquisition, 2017. <https://astewartau.github.io/Planar-LF-Calibration>, [Online; accessed 19-11-2017].
- [Torr and Zisserman, 2000] Philip H. S. Torr and Andrew Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.
- [Ueshiba and Tomita, 2003] Toshio Ueshiba and Fumiaki Tomita. Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 966–973 vol.2, 2003.
- [Vaish *et al.*, 2004] Vaibhav Vaish, Bennett Wilburn, Neel Joshi, and Marc Levoy. Using Plane + Parallax for Calibrating Dense Camera Arrays. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–2–I–9 Vol.1, June 2004.
- [Wanner and Goldluecke, 2014] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE transactions on pattern analysis and machine intelligence*, 36(3):606–619, 2014.
- [Xu *et al.*, 2014] Yichao Xu, Kazuki Maeno, Hajime Nagahara, and Rin-ichiro Taniguchi. Mobile camera array calibration for light field acquisition. *CoRR*, abs/1407.4206, 2014.
- [Yang *et al.*, 2002] Jason C. Yang, Matthew Everett, Chris Buehler, and Leonard McMillan. A real-time distributed light field camera. *Rendering Techniques*, 2002:77–86, 2002.
- [Zhang, 2000] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, November 2000.