Using Plane + Parallax for Calibrating Dense Camera Arrays

Vaibhav Vaish* Bennett Wilburn† Neel Joshi* Marc Levoy*

*Department of Computer Science †Department of Electrical Engineering
Stanford University, Stanford, CA 94305

Abstract

A light field consists of images of a scene taken from different viewpoints. Light fields are used in computer graphics for image-based rendering and synthetic aperture photography, and in vision for recovering shape. In this paper, we describe a simple procedure to calibrate camera arrays used to capture light fields using a plane + parallax framework. Specifically, for the case when the cameras lie on a plane, we show (i) how to estimate camera positions up to an affine ambiguity, and (ii) how to reproject light field images onto a family of planes using only knowledge of planar parallax for one point in the scene. While planar parallax does not completely describe the geometry of the light field, it is adequate for the first two applications which, it turns out, do not depend on having a metric calibration of the light field. Experiments on acquired light fields indicate that our method yields better results than full metric calibration.

1. Introduction

The 4D light field is commonly defined as the radiance along all rays that intersect a scene of interest. In practice, light fields are acquired by obtaining images of a scene with a camera from several different viewpoints (which often lie on a plane). These images constitute a dense sampling of rays in the light field. The number of images is typically large, ranging from hundreds to tens of thousands. Lights fields are useful for several applications in computer graphics and vision:

(i) Image Based Rendering: If we have all the rays that intersect a scene, we can easily compute an image of the scene from any viewpoint outside its convex hull by selecting those rays which pass through the viewpoint. For discretely sampled light fields, we reconstruct rays passing through any chosen viewpoint by resampling the acquired rays. Using a two-plane parametrization, this can be done efficiently so new views can be computed at interactive rates [4, 8].

(ii) Synthetic Aperture Photography: Light fields can be used to simulate the defocus blur of a conventional lens, by reprojecting some or all of the images onto a (real or virtual) "focal" plane in the scene, and computing their average. Objects on this plane will appear sharp (in focus), while those not on this plane will appear blurred (out of focus) in the resulting image [7]. This synthetic focus can be thought of as resulting from a large-aperture lens, the viewpoints of light field images being point samples on the lens surface. We call this *synthetic aperture photography*. When the aperture is wide enough, occluding objects in front of the focal plane are so blurred as to effectively disappear (see Fig. 4, and supplementary videos). This has obvious uses for surveillance and military applications.

(iii) Estimating Scene Geometry: Recovery of 3D geometry from multiple images has been an active area of research in computer vision. Moreover, light field rendering and compression can be improved if even approximate geometry of the scene can be recovered. An attractive technique for estimating shape from light fields is plane sweep stereo [1, 11]. The light field images are projected on each of a family of planes that sweep through space. By observing how well the images align, we can recover the different layers of geometry.

Current implementations of these applications require accurate calibration of the camera setup used to acquire light field. Most light fields are acquired by a calibrated camera moving on a controlled gantry [7, 8]. Recently, several researchers including ourselves have built large arrays of video cameras to capture light fields of dynamic scenes [14, 15]. Calibrating these arrays is challenging, and especially difficult for acquisitions outside the laboratory. Interestingly, not all applications of light fields require a full metric calibration. For example, resampling the light field to render a novel view requires the locations of the viewpoints used

to capture the images, but synthetic aperture photography (and plane sweep stereo) needs just homographies to project each image onto different planes in the scene. Stated more precisely, we address the question: is it possible to implement the applications enumerated earlier without a complete metric calibration of the array ? If so, what is the minimal calibration required, and what are good ways to compute this calibration ?

In answering these questions, we have found it useful to characterize light fields using a plane + parallax representation. Specifically, we assume that the images of the light field have been aligned on some reference plane in the world, and we are able to measure the parallax of some points in the scene that do not lie on this reference plane. We can demonstrate that when the cameras lie on a plane parallel to the reference plane (and we have an affine basis on the reference frame) we can recover the camera positions, up to an affine ambiguity. To do so, we need to measure the parallax of just one scene point not on this reference plane (in all views). These camera positions are adequate for conventional light field rendering [8] and for synthetic aperture photography on focal planes that are parallel to the camera plane. We also derive a rank-1 constraint on planar parallax. This facilitates robust computation of the camera positions if we have parallax measurements for more points. No knowledge of camera internal parameters or orientations is needed.

Although having planar cameras and a parallel reference might seem restrictive, it is essentially the two plane parametrization used to represent light fields (see Fig. 1). All light field camera arrays we are aware of have their cameras arranged in this way. A planar arrangement of cameras is one of the critical motions [10] for which metric self-calibration is not possible and projective or affine calibration is the best we can achieve.

2. Related Work

Light fields were originally introduced in computer graphics for image-based rendering [4, 8]. The notion of averaging images taken from adjacent cameras in the light field to create a synthetic aperture was described in [8]. The use of a synthetic aperture to simulate defocus was described by Isaksen et al [7]. Using synthetic imagery, they also showed how a sufficiently wide aperture could be used to see around occluders. Favaro showed that the finite aperture of a single camera lens can be used to see beyond partial occluders [3]. In our laboratory, we have explored the use of the synthetic aperture of a 100-camera array to see objects behind occluders like dense foliage, and we present some of these results here.

The plane + parallax formulation for multiple views has been studied by several researchers [2, 5, 6, 12]. Triggs uses a rank constraint similar to ours for projective factorization, but requires knowledge of projective depths [12]. Seitz's computation of scene structure from four point correspondences assuming affine cameras falls out as a special case of our work [9] . Our rank constraint requires no prior knowledge of depths, and works equally well for perspective and affine cameras.

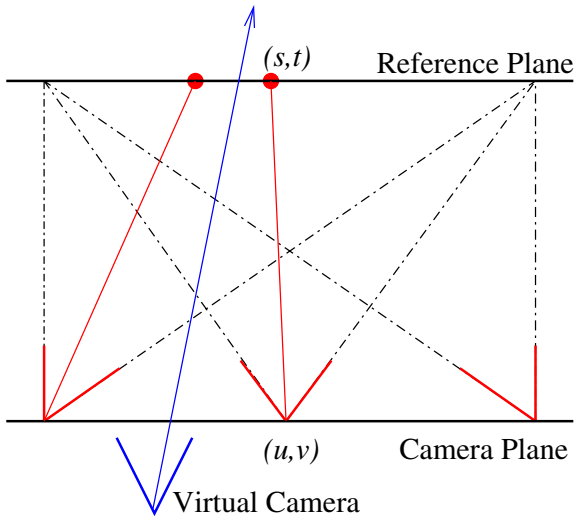


Figure 1: Light field rendering using the two-plane parametrization. The image viewpoints lie on the camera plane, and images are aligned on a parallel reference plane. Rays are parametrized by their intersection (u, v) with the camera plane and (s, t) with the reference plane. Rays desired for a novel view are computed by applying a 4-D reconstruction filter to the nearest samples in (u, v, s, t) space.

Rank constraints on homgraphies required to align images from different views onto arbitrary scene planes were studied by Zelnik-Manor and Irani [16, 17]. They derive a rank-3 constraint on relative homographies (homologies) for infinitesimal camera motion, and rank-4 constraint for general camera motion. Our rank-1 constraint is a special case of the latter.

3. Parallax for Planar Camera Arrays

We shall now study the relationship between planar parallax and camera layout for a planar camera array. We assume, for now, that the images of these cameras have been projected onto a reference plane parallel to the plane of cameras, and we have established an affine coordinate system on this reference plane. This can be

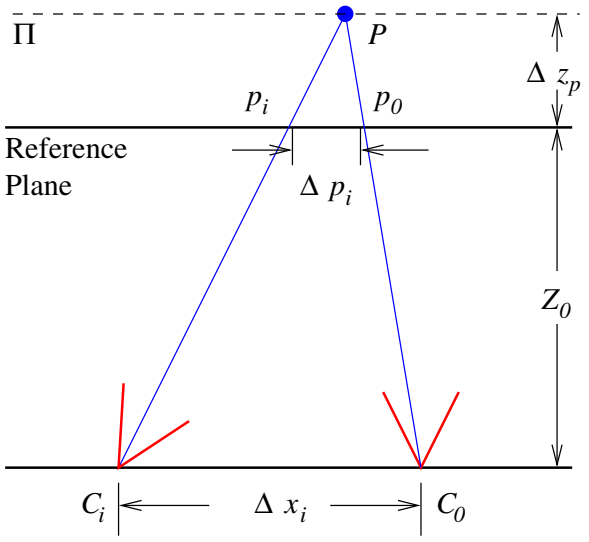


Figure 2: Planar parallax for light fields. A point P not on the reference plane has distinct images p_0, p_i in cameras C_0, C_i. The parallax between these two depends on the relative camera displacement Δx_i and relative depth $\frac{\Delta z_p}{\Delta z_p + Z_0}$.

done by identifying features on a real plane in the scene which is parallel to the camera plane and on which affine structure is known; we use a planar calibration grid for this. In terms of the two plane parametrization, we have established an s, t coordinate system on the reference plane. This is similar to image rectification employed in stereo, where the reference plane is at infinity. We begin by deriving a rank-1 constraint on planar parallax and show how to use it for image-based rendering and synthetic aperture photography.

3.1 A Rank-1 Constraint on Parallax

Consider a light field acquired by a planar array of cameras $C_0, \ldots, C_m$, whose images have been aligned on a reference plane parallel to the plane of cameras, as shown in Fig 2. Consider the images $p_0 = (s_0, t_0)^T, p_i = (s_i, t_i)^T$ of a point P in the two cameras C_0, C_i. Since P is not on the reference plane, it has parallax $\Delta p_i = p_i - p_0$. From similar triangles,

$$\Delta p_i = \Delta x_i \frac{\Delta z_p}{\Delta z_p + Z_0} \tag{1}$$

The parallax is the product of the relative camera displacement Δx_i and the relative depth $d_p = \frac{\Delta z_p}{\Delta z_p + Z_0}$ of P. This is easy to extend to multiple cameras and multiple scene points. Choose camera C_0 as a reference view, with respect to which we will measure parallax. Suppose we observe images of n points $P_1, \ldots, P_n$ with relative depths $d_1, \ldots, d_n$ in cameras $C_1, \ldots, C_m$. If the parallax of P_j in C_i is Δp_{ij}, we can factorize the matrix of parallax vectors:

$$\begin{bmatrix} \Delta p_{11} & \ldots & \Delta p_{1n} \\ \vdots & \ddots & \vdots \\ \Delta p_{m1} & \ldots & \Delta p_{mn} \end{bmatrix} = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_m \end{bmatrix} \begin{bmatrix} d_1 & \ldots & d_n \end{bmatrix}$$

$$\mathbf{\Delta P} = \mathbf{\Delta X}\ \mathbf{D} \tag{2}$$

Here $\mathbf{\Delta P}, \mathbf{\Delta X}, \mathbf{D}$ denote the matrix of parallax measurements, (column) vector of camera displacements with respect to the reference camera and (row) vector of relative depths respectively. Thus, parallax measurements form a rank-1 matrix. It is easy to show that this also holds when the cameras are affine (camera centers lie on the plane at infinity).

3.2 Recovering Camera Positions

Consider the parallax vector for the point P_1, $\Delta P_1 = \left[\Delta p_{11}^T \ldots \Delta p_{m1}^T\right]^T$. From (2), this is a scalar multiple of the camera displacements relative to C_0, i.e. $\Delta P_1 = d_1 \mathbf{\Delta X}$. Thus, parallax measurement of a single point gives us the camera positions up to a scale[1]. In practice, we measure parallax for several scene points and compute the nearest rank-1 factorization via SVD. The rank-1 constraint lets us recover $\mathbf{\Delta X}$ from multiple parallax measurements in a robust way.

3.3 Image-Based Rendering

Once we have the relative camera positions, the light field can be represented in the two plane parametrization shown in Fig. 1. Each ray can represented as a sample (u, v, s, t) in 4D space, where (u, v) is its interesction with the camera plane (i.e. the position of the camera center) and (s, t) the intersection with the reference plane (i.e. pixel coordinate on the reference plane) [8, 4]. In conventional light field rendering, the rays needed to render a novel view are computed by applying a 4D reconstruction filter to the sampled rays. Unlike conventional light field rendering, in our case the $(u, v)-$ and $(s, t)-$ coordinate systems are affine, and not Euclidean. However, an affine change of basis does not affect the output of a linear reconstruction filter (such as quadrilinear interpolation) used to compute new rays. Hence, the recovered camera positions are adequate for rendering a novel view.

[1] More precisely, the camera positions are known up to a scale in terms of the coordinate system on the reference plane. Since we have an affine coordinate system on the reference plane, the camera positions are also recovered up to an affine ambiguity.

3.4 Synthetic Aperture Photography

To simulate the defocus of an ordinary camera lens, we can average some or all of the images in a light field [7]. Objects on the reference plane will appear sharp (in good focus) in the resulting image, while objects at other depths will be blurred due to parallax (i.e. out of focus). The reference plane is analogous to the focal plane of a camera lens. If we wish to change the focal plane, we need to reparametrize the light field so that the images are aligned on the desired plane.

Suppose we wish to focus on the plane Π in Fig. 2. To align the images on Π, we need to translate the image of camera C_i by $-\Delta p_i = -d_i \Delta x_i$. Let $T = -\left[\Delta p_1^T \dots \Delta p_m^T\right]^T$ be the vector containing the required translations of each image. The rank-1 constraint tells us that this is a scalar multiple of the relative camera positions, i.e. $T = \mu \left[\Delta x_1^T \dots \Delta x_m^T\right]^T$. Using different values of μ will yield the required image translations for reference planes at different depths. μ is analogous to the focal length of a simulated lens. The rank-1 constraint is useful because it lets us change the focal length (depth of the reference plane) without explicit knowledge of its depth, or having to measure parallax for each desired focal plane. In our experiments, we let the user specify a range of values of μ to generate images focussed at different depths.

The ability to focus at different depths is very useful - if the cameras are spread over a sufficiently large baseline, objects not on the focal plane become blurred enough to effectively disappear. In section 4, we demonstrate how this can be used to see objects occluded by dense foliage.

4. Experiments and Results

We implemented the method described above for calibrating relative camera positions and synthetic aperture photography, and tested it on a light field captured outdoors. Our experimental setup is shown in Fig 3. We used an array of 45 cameras, mounted on a (nearly) planar frame 2m wide. Our lenses had a very narrow field of view, approximately 4.5^o. The image resolution was 640x480 pixels. The goal was to be able to see students standing behind dense shrubbery about 33m from the array, by choosing a focal plane behind it. To obtain image correspondences, we used a 85cm x 85cm planar grid pattern consisting of 36 black squares on a white background, mounted on a stiff, flat panel. We built an automatic feature detector to find the corners of the black squares in the camera image and match them to the grid corners. The central camera was selected as the reference view.

First, we placed the calibration grid about 28m from the array and approximately parallel to it. This supplied the position of the reference plane. A homography was applied to each camera to align its image onto this reference plane. We used our knowledge of the grid geometry to establish an affine coordinate system on the reference plane.

Next, we moved the calibration grid to six other positions, ranging from 28m to 33m from the target. The motion of the grid was uncalibrated. These gave us plenty of correspondences for measuring parallax. We measured the parallax of the image of each corner in every camera with respect to the corresponding corner in the reference camera. We observed a total of 6 x 36 x 4 = 864 points in the world. The rank-1 factorization of the matrix of parallax vectors gave the camera locations relative to the reference view. The RMS error per parallax observation between the output of the feature detector and computed rank-1 factorization was 0.30 pixels. This suggests that our assumption of a planar camera array and a parallel reference plane was fairly accurate.

We then computed synthetic aperture images on focal planes ranging from approximately 28m to 45m using the method described in section 3.4. The synthetic aperture images for different positions of the focal plane, along with some of the original light images are shown in Fig. 4. The reader is encouraged to see the electronic version of this paper for color images. The supplementary videos show sequences of synthetic aperture images as the focal plane sweeps through a family of planes parallel to the reference plane that spans the depths of our scene. The sharpness of objects on the focal plane indicates that the images are well aligned. The fact that we can focus well on focal planes beyond our calibration volume indicates the accuracy of our technique.

4.1 Comparison with Metric Calibration

We also performed a full metric calibration of the camera array, using the same set of images of the calibration grid. We used a multi-camera version of Zhang's algorithm [18]. Zhang's method computes intrinsic parameters and pose with respect to each position of the calibration grid for each camera independently. The calibration parameters returned by Zhang's method were used as an initial estimate for a bundle adjustment, which solved simultaneously for all the camera parameters and motion of the calibration grid that minimize reprojection error. Details of the implementation are available on our website [13]. The bundle adjustment returned calibration parameters with an RMS er-

ror of 0.38 pixels. The orthographic projections of the computed camera centers onto the reference plane were used to compute synthetic aperture images using the method of section 3.4, with the same light fields aligned on the same reference plane as above. We have obtained good results with this method in the laboratory at ranges of 1.5m-8m.

A comparison of the synthetic aperture images generated using the parallax based calibration and using full metric calibration is shown in Fig. 4. Errors in relative camera positions manifest themselves as misfocus in the synthetic aperture sequences. The synthetic aperture images from both the sequences are comparable in quality when the focal plane is within the calibration volume. However, we are not able to get sharp focus using metric calibration when the focal plane moves beyond the calibration volume. In particular, we are never able to see the students behind the bushes or the building facade as well as we can with the parallax-based method. We suspect that metric calibration did not perform as well as parallax based calibration for the following reasons:

1. Metric calibration solves for a much larger number number of parameters, including camera intrinsics and rotation. Our method needs to solve only for relative camera positions and not the intrinsics and rotations, which are factored into the homography that projects images onto the reference plane.

2. The metric calibration does not exploit the fact that the camera locations are almost coplanar.

3. At large distances, the calibration grid covers only a small part of the field of view. This could result in unreliable pose estimation, leading to inaccurate initialization for bundle adjustment, which may then get trapped into a local minima. Our method does not require pose estimation or nonlinear bundle adjustment.

We emphasize that metric calibration and our parallax-based method are not computing the same parameters, nor making the same assumptions on camera configurations. Thus, it is not correct to compare their relative accuracy. For applications like light field rendering and synthetic aperture photography, where planar cameras are commonly used, experiments indicate the parallax based method is more robust. For applications like metric reconstruction, or calibrating more general camera configurations, we would have to use metric calibration.

5. Conclusions and Future Work

There are two reasons why the plane + parallax method is useful for light fields. First, prior alignment of images on a reference plane is already a requirement for most light field representations. Second, for planar cameras and a parallel reference plane we can write the planar parallax as a product of camera displacement and relative depth (Eq (1)). Specifically, parallax is a bilinear function of camera parmaters and relative depth, which enables the rank-1 factorization described in section 3. This leads to a simple calibration procedure which is robust and yields better results than a full metric calibration for applications of light fields described above. We are currently investigating extensions of this technique for cameras in general positions and arbitrary reference planes. For such configurations, changing the reference frame requires a *planar homology* [2], rather than just an image translation. We believe we can still use parallax measurement to compute the parameters of homologies needed to reproject images onto new reference planes.

Synthetic aperture photography can be used to recover scene geometry using a shape-from-focus algorithm. Although we have only considered focussing on planes, one could (given enough calibration) project the images onto focal surfaces of arbitrary shape. This suggests investigating algorithms that construct an evolving surface which converges to the scene geometry. It would be interesting to extend techniques for shape estimation, segmentation and tracking from multiple frames to light fields of scenes where objects of interest may be severely occluded. As an example, we could try to use synthetic aperture photography to track a person in a crowd using a focal surface that follows the target person.

Acknowledgments

We would like to thank our colleagues for assistance in our outdoor light field acquisitions. This work was partially supported by grants NSF IIS-0219856-001 and DARPA NBCH 1030009.

References

[1] R. Collins. A Space Sweep Approach to True Multi-Image Matching. In *Proc. of IEEE CVPR,* 1996.

[2] A. Criminisi, I. Reid, and A. Zisserman. Duality, Rigidity and Planar Parallax. In *Proc. of ECCV*, 1998.

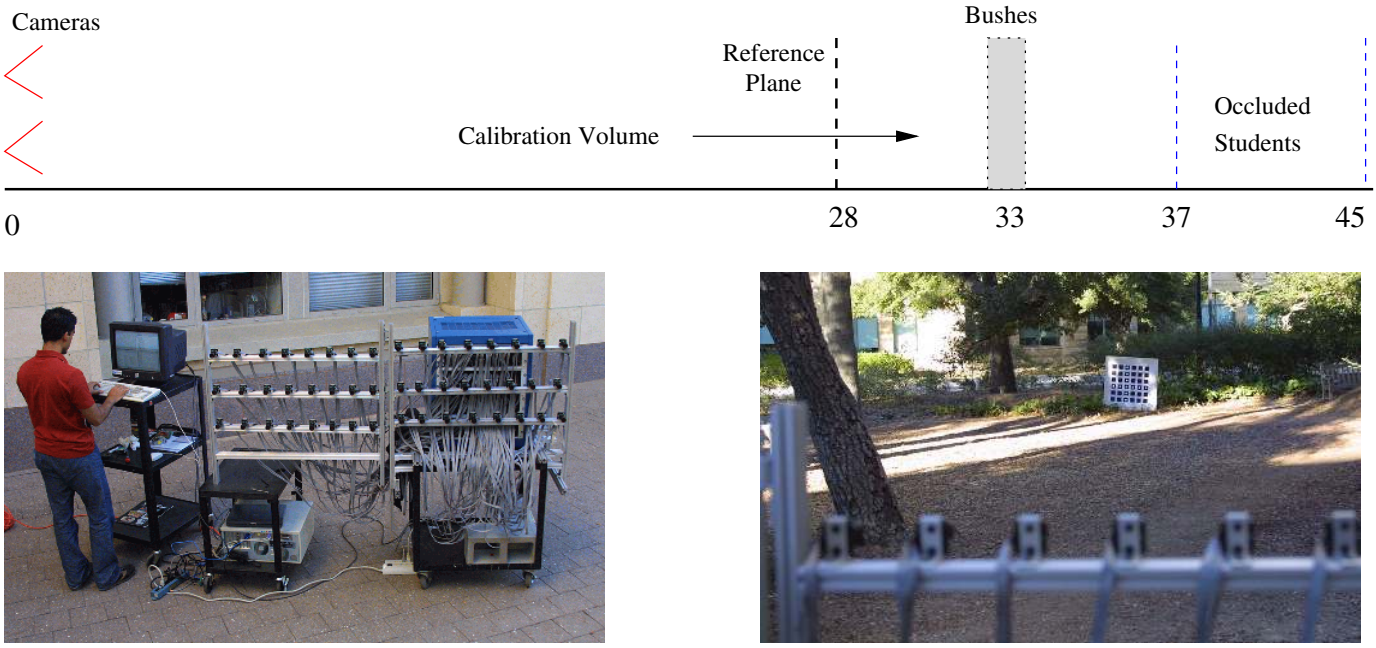


Figure 3: Experimental setup for synthetic aperture photography. Top: scene layout and distances from camera array (meters). Left: Our 45-camera array on a mobile cart, controlled by a standalone 933 MHz Linux PC. Right: The view from behind the array, showing the calibration target in front of bushes. The goal was to try and see people standing behind the bushes using synthetic aperture photography.

[3] P. Favaro and S. Soatto. Seeing Beyond Occlusions (and other marvels of a finite lens aperture). In *Proc. of IEEE CVPR*, 2003.

[4] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The Lumigraph. In *Proc. of ACM SIGGRAPH*, 1996.

[5] M. Irani, P. Anandan and M. Cohen. Direct Recovery of Planar-Parallax from Mulitple Frames. In *Vision Algorithms: Theory and Practice*, Springer Verlag LNCS vol 915, 1999.

[6] M. Irani, P. Anandan and D. Weinshall. From Reference Frames to Reference Planes: Multi-View Parallax Geometry and Applications. In *Proc. of ECCV*, 1996.

[7] A. Isaksen, L. McMillan and S. Gortler. Dynamically Reparametrized Light Fields. In *Proc. of ACM SIGGRAPH*, 2000.

[8] M. Levoy and P. Hanrahan. Light Field Rendering. In *Proc. of ACM SIGGRAPH*, 1996.

[9] S. Seitz, C. Dyer. Complete Scene Structure from Four Point Correspondences. In *Proc. of IEEE ICCV,* 1995.

[10] P. Sturm. Critical Motion Sequences for the Self-Calibration of Cameras and Stereo Systems with Variable Focal Length. In *Proc. of BMVC, 1999.*

[11] R. Szeliski. Shape and Appearance Models from Multiple Images. In *Workshop on Image Based Rendering and Modelling (http://research.microsoft.com/∼szeliski/IBMR98/web/)*, 1998.

[12] B. Triggs. Plane + Parallax, Tensors and Factorization. In *Proc. of ECCV*, 2000.

[13] V. Vaish. Light Field Camera Calibration. *http://graphics.stanford.edu/projects/array/geomcalib/*

[14] B. Wilburn, M. Smulski, H. Kelin Lee, M. Horowitz. The Light Field Video Camera. In *Proc. SPIE Electronic Imaging*, 2002.

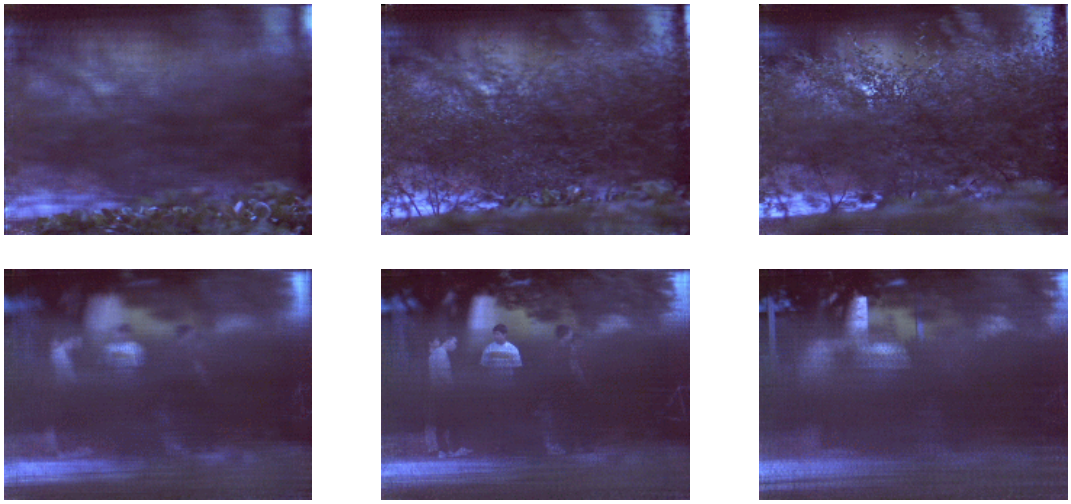
[15] J. Yang, M. Everett, C. Buehler, L. McMillan. A Real-Time Distributed Light Field Camera. In *Proc. of Eurographics Workshop on Rendering*, 2002.

[16] L. Zelnik-Manor, M. Irani. Multi-view Subspace Constraints on Homographies. In *Proc. of IEEE ICCV,* 1999.

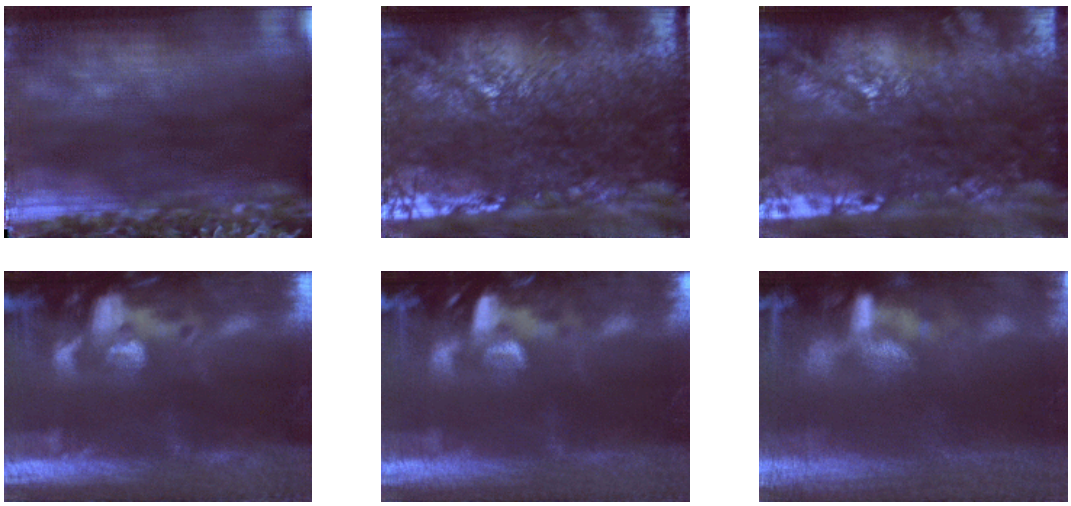
[17] L. Zelnik-Manor and M. Irani. Multi-frame Alignment of Planes. In *Proc. of IEEE CVPR,* 1999.

[18] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. *Technical Report MSR-TR-98-71*, Microsoft Research, 1998.

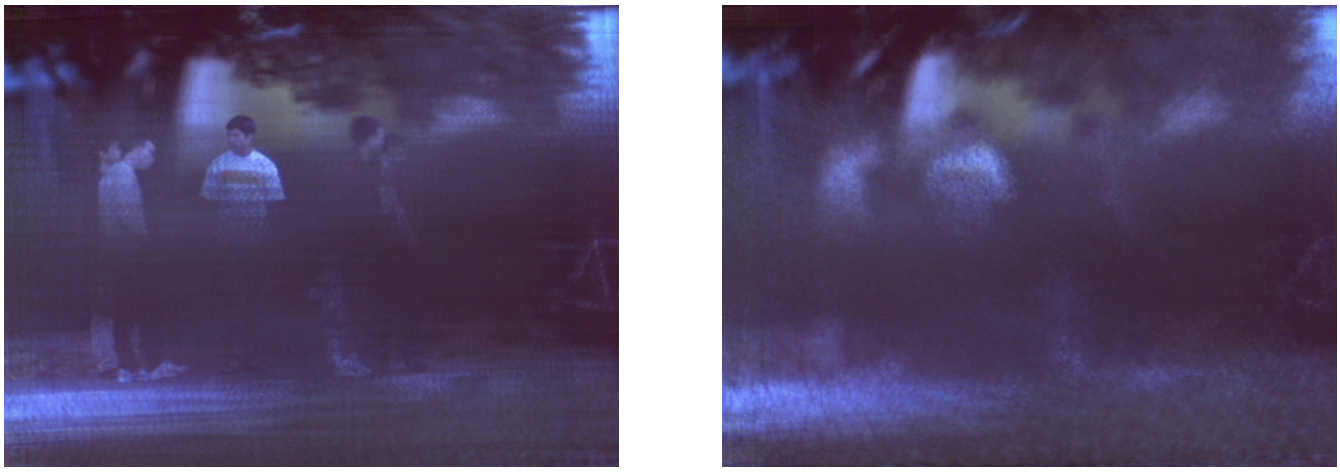
(a) Two images from a light field of students standing behind bushes.



(b) Synthetic aperture sequence, with the focal plane moving away from the array computed using our parallax based calibration method. We get sharp focus at different depths, ranging from the bushes to the building facade.



(c) Synthetic aperture sequence computed using full metric calibration. This does not produce well focussed images as the focal plane moves beyond the bushes.



(d) Synthetic aperture images from the light field, using parallax based calibration (left) and metric calibration (right). For both methods, we varied the focal plane to get the sharpest possible image of students behind the the bushes. Clearly, the parallax based calibration produces better focussed images.

(e) Results from another light field, of a cyclist behind the bushes. Left: an image from the original light field. Right: Synthetic aperture image at the focal plane through the cyclist, computed using our parallax method.

Figure 4: Results of our synthetic aperture photography experiments, comparing our parallax based calibration method with full metric calibration.