Calibrating Focused Light-Field Cameras Using Plenoptic Disc Features

Sean G. P. O'Brien, Jochen Trumpf, Viorela Ila, and Rob Mahony
Australian Centre for Robotic Vision,
Research School of Engineering,
Australian National University, Canberra, ACT, 2601
{sean.obrien, jochen.trumpf, viorela.ila, robert.mahony}@anu.edu.au

Abstract

This paper proposes a new method for estimating calibration parameters of plenoptic cameras by minimizing the nonlinear plenoptic reprojection error. Novel plenoptic feature types are proposed as data for the calibration method. These plenoptic disc features are in a natural one-to-one correspondence with physical points in front of the camera. We exploit the intrinsic geometry of plenoptic cameras in a novel projection model that relates the plenoptic disc features to physical points. The resulting calibration quality, as quantified by mean reprojection error and 3D reconstruction error, outperforms recently published results.

1. Introduction

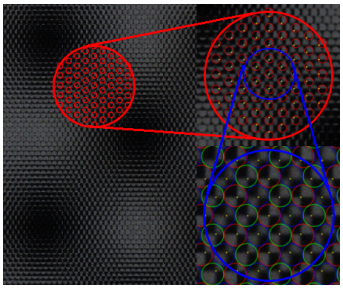


Figure 1: A zoomed view of a raw light-field image taken by a Raytrix R42 camera. In the large red circle is a plenoptic disc containing the set of lenslets that can see a specific feature. The yellow dots highlight the detected feature. Subimages corresponding to different lenslet types are shown in the blue circle in different colors.

Calibration estimation is a fundamental problem in computer vision. Accurate calibration, both intrinsic and extrinsic, is essential for the generation of metrically correct scene reconstructions, as well as being crucial in other preprocessing tasks. Due to the complexity of the lenslet geometry of plenoptic cameras, existing multi-camera calibration methods are not directly effective and there is a growing literature aimed at developing effective models for calibration [17, 16, 19, 18, 8, 5, 15, 7, 2, 13, 1].

Most existing calibration techniques, for both plenoptic and other sorts of cameras, consist of three main steps and are based on estimating a projective transformation that models the camera for a ray-based model of light. The first step takes in as data raw images, and estimates the locations of features in these images generating a list of correspondences between frames. The second step is initialisation, generating an initial estimate of the calibration parameters. A cost function, typically mean reprojection error, is then minimised in the third, optimisation step.

This paper proposes a new calibration method for plenoptic light-field cameras that outperforms existing methods on the key performance measures of mean reprojection error and 3D reconstruction error. The first step of most existing calibration methods requires matching of subimage feature points between subimages, and data association between these image points between different light-fields (image frames) of a known target. As there can be dozens of subimages in which a given feature is visible per frame (see Fig. 1), the identification of subimage features becomes an onerous task [13]. This process is made more difficult by the fact that the different focal lengths used for multi-focus light field cameras [10] mean that many of the subimages are out-of-focus for each of the images. Furthermore, since each subimage has a small resolution there is limited information available to make highly accurate feature extraction, especially when a feature point approaches the edge of a subimage. Reliable and robust extraction of accurate subimage feature points for lightfield calibration is a key limitation to existing calibration methods. Our proposed method avoids this problem by exploiting the intrinsic geometry of the light-field camera in order to derive a feature type that is unique to plenoptic cameras, namely plenoptic disc features. These features are similar in nature to 'circles of confusion' in conventional photography

[14], although this term can also refer to a certain type of optical aberration [6]. This feature type has similarities to the feature type suggested in [3], however that paper does not use such feature types for calibration. Furthermore, we show that our plenoptic disc feature parametrisation is 3-dimensional and these features are in one-to-one correspondence with positions of points in the body-fixed frame of the camera. We call the function that maps points to these features the *plenoptic projection*.

Exploiting this feature parametrisation, together with the simplified 3-intrinsic parameter model originally proposed by [2], vastly simplifies the optimisation problem used in the final step. The initialisation procedure is based on the work by [2], reformulated to use our plenoptic disc features. The cost function minimised in our final step is a *plenoptic reprojection error*, that minimises distance between the coordinates of the plenoptic disc features used as data and the expected coordinates of these features given camera parameter estimates.

The mean reprojection and reconstruction errors that we obtain from this approach outperform state-of-the-art results [2, 13]. In summary, the main contributions of this paper are:

- Full use of the plenoptic camera geometry to derive a plenoptic disc feature type that allows robust feature extraction from raw light-field images.
- A calibration method based on this geometry that outperforms existing state-of-the-art techniques in terms of accuracy and robustness.

1.1. Previous Work

Seminal work in the calibration of lenslet based plenoptic cameras was published in 2013 [3, 8]. Dansereau *et al*. used the Lytro plenoptic camera and derived a camera rectification formulation that allowed a simple optimisation algorithm for image calibration. A similar approach is undertaken more recently in [7]. An advantage of this approach is that the resulting calibration optimisation tends to be more robust, however, the parameters identified are less directly associated with physical parameters of the camera. Moreover, the approach is less well suited to multi-focal light field cameras.

Johannsen *et al*. [8] formulated a general reprojection model in terms of the physical parameters of a Raytrix camera. This work considered a relatively simple model of lens distortion and required careful initialisation of the optimisation to converge. Strobl *et al*. [15] recognised the fragility of the calibration optimisation and proposed a step-wise calibration approach where first the focal length and optical centre of the main lens is determined (as well as some distortion parameters) before the internal offset of the Micro-Lens Array (MLA) from the sensor and main lens respectively are determined. Sun *et al*. [16] use a similar approach, where they hand determine the ratio of MLA distance to sensor with respect to MLA distance to the image plane for a specific point, allowing them to effectively identify the relative focal length of the main lens separately form the calibration process. Another recent contribution is proposed by Zeller [17, 18]. Although the focus of these papers is on visual odometry, they require a calibrated camera to provide metric reconstructions. The depth calibration proposed in [18] uses a separate optimisation process. Another direction stemming from this approach has lead to the consideration of more sophisticated models of the lens distortion and non-planarity of the MLA. Heinze *et al*. [5] consider more sophisticated models of the distortion of the main lens. Zhang *et al*. [19] consider a detailed model of the lenslet array geometry that calibrates for non-planarity of the array. Lenslet based plenoptic cameras, however, are constructed with careful attention to the coplanarity of the lenslet array and the image plane [12], and for cameras such as the Raytrix R42, this additional complexity is not required.

All the above papers require matching of point features across multiple images and multiple subimages. Although many of the methods use standard feature extraction methods to automate the matching process, there are necessarily errors in the identification and data association of these features. Bok *et al*. [2] introduced novel line features to improve the automation and accuracy of the feature identification. More recently Nousias *et al*. [13] developed corner based features along with an end-to-end calibration process. Both these papers have achieved improved performance through automation and accurate identification of feature correspondences. These papers provide a good benchmark for the evaluation of the present work, particularly since they minimize a mean reprojection error criterion as we do.

2. Problem Formulation

In this section, we introduce the projective model used in the proposed calibration method.

2.1. Projection Through a Thin Lens

In this sub-section we express all points in the body-fixed frame $\mathbf{C}$ of the camera. A point P expressed in this frame has coordinates (P^x, P^y, P^z). We model the focus lens positioned in front of the micro-lens array (MLA) as a thin lens. For thin lenses, every point P on one side of the lense corresponds to another point Q, for which all the rays of light passing through P pass through Q and vice-versa (see Fig. 2). The point Q is called the *image point* of P. For points P in front of the camera, so that $P^z > 0$, we have that the image point Q satisfies

$$\frac{1}{F} = \frac{1}{P^z} - \frac{1}{Q^z}, \tag{1}$$

where F is the focal length of the focus lens. Because the image point Q always lies on a line passing through P and

the optical centre, we can determine the position of Q to be given by

$$Q = \left(\frac{F}{F - P^z}\right) P. \tag{2}$$

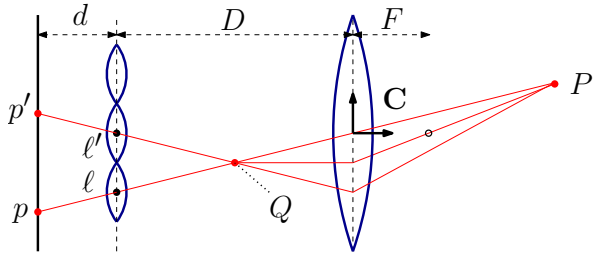


Figure 2: A point P with image point Q is shown. Two lenslets ℓ and ℓ' are shown with the pixels p and p' of the perspective projections of the point Q through each respective lenslet.

2.2. Projection Through a Micro-Lens Array

Lenslet-based light-field cameras are constructed by positioning an MLA in between a focus lense and an imaging plane. The raw data of a light-field camera appears as an hexagonal array of smaller circular *subimages* stitched together (Fig. 1). A subimage in the raw data is the image produced by a single lenslet ℓ. The pixel coodinates (p^u, p^v) of a pixel p are counted positively from the top left corner of the raw light-field image. We assign to each lenslet ℓ *lenslet coordinates* $(\ell^s, \ell^t) \in \mathbb{R}^2$ given by the pixel coordinates of the centre of the subimage of that lenslet (Fig. 5)

Lenslets ℓ and pixels p are also represented by their *physical coordinates* expressed in the body-fixed frame $\mathbf{C}$ of the camera, denoted $\ell = (\ell^x, \ell^y, \ell^z)$ and $p = (p^x, p^y, p^z)$, respectively. We assume that the MLA is parallel to the main focus lens of the camera, so that all the lenslets have a constant displacement $\ell^z = -D$, and so $\ell = (\ell^x, \ell^y, -D)$, see Fig. 2.

The physical coordinates ℓ of the optical centre of the lenslet are related to its lenslet coordinates (ℓ^s, ℓ^t) by the equation

$$\ell = \left(S\frac{D}{D+d}(\ell^s - c^u), S\frac{D}{D+d}(\ell^t - c^v), -D\right). \tag{3}$$

In this equation, S is the physical scale of the MLA in metres per pixel, (c^u, c^v) are the pixel coordinates of the optical centre of the camera, and d is the distance between the MLA and the imaging plane. The parameter $\frac{D+d}{S}$ is referred to as f in other papers [13, 2], and in the literature is called the "focal length" of a pinhole camera model for the lenslet. However, the physical meaning of this parameter should not be confused with that of the focal length of a thin-lens.

Under the assumption that pixels are at a constant distance d from the micro-lens array, the physical coordinates of a pixel p with pixel coordinates (p^u, p^v) are given by

$$p = (S(p^u - c^u), S(p^v - c^v), -D - d). \tag{4}$$

Given physical coordinates ℓ of a lenslet and an image point Q, the location of the pixel that images Q through the lenslet ℓ is found by determining where the line passing through Q and ℓ intersects the pixel plane (Fig. 2). Using a similar-triangles argument, p is given by

$$p = \frac{d}{D + Q^z}(\ell - Q) + \ell. \tag{5}$$

The pixel coordinates of p are then found by solving (4) for (p^u, p^v).

3. Plenoptic Projections and Features

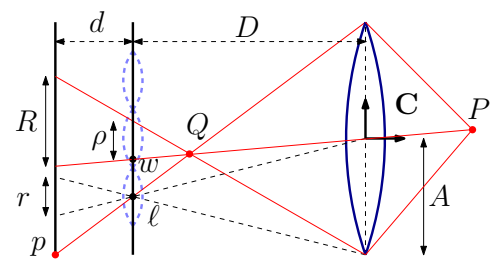


Figure 3: A plenoptic disc corresponding to a point P is entirely determined by the parameters w and ρ. A straight line passing through the optical centre of the focus lens and the image point Q is shown, and where this line intersects the pupilar plane is the plenoptic disc centre w. A lenslet on the boundary of W is labelled ℓ. The pixel p in the subimage of ℓ that images Q appears on the boundary of the subimage of ℓ. The radius ρ can be calculated from the aperture A using a similar-triangles argument. The figure also shows the subimage radius r.

In this section, we use an idealised model of a plenoptic camera called the two-plane parametrisation. In this model, we assume that there are an infinite number of lenslets positioned at every point of a plane, called the pupilar plane, behind the focus lens of the plenoptic camera.

In such a model, a point P is projected to a set of lenslet-pixel pairs (ℓ, p), where p is the projection of the image point Q through the lenslet ℓ, as given by (5). However, if the subimages of each of the lenslets are circular and equal in radius, the set of lenslet-pixel pairs contains no more information than the set W of lenslets ℓ for which the point P is visible to lenslet ℓ. We call the set W the *plenoptic disc* of a point P.

If the subimages of each of the lenslets are circular, the plenoptic disc will be circular, and thus can be represented

entirely by a centre $w = (w^x, w^y, -D)$ and a signed radius ρ. This is because the set of lenslets ℓ for which a virtual point Q projects into the subimage of ℓ forms a disc in the pupilar plane, see Fig. 3.

The relationship between an image point Q and the plenoptic disc data is as follows. Let A be the physical aperture radius of the circular focus lens of the camera. A lenslet physical coordinate ℓ is on the boundary of the set W if there is a line passing from ℓ through Q such that the intercept of this line with the focus lens has length A, see Fig. 3. By similar triangles, we find that

$$\rho = \frac{D + Q^z}{-Q^z} A. \tag{6}$$

The physical centre of W is given by

$$w = -\frac{D}{Q^z} Q. \tag{7}$$

Since the physical radius ρ is related to a radius R in the raw image by the relation $\rho = S\frac{D}{D+d}R$, then using $A = \frac{D}{d}(Sr)$, where r is the lenslet subimage radius expressed in pixels, together with (6) and (1), we find that

$$\frac{1}{P^z} = \frac{D - F}{DF} - \frac{d}{r(D+d)D} R, \tag{8}$$

where P is the real point corresponding to the image point Q. The parameters $\frac{D-F}{DF}$ and $-\frac{d}{r(D+d)D}$, are equal to $-\frac{K_1}{K_2}$ and $-\frac{1}{rK_2}$, respectively, using the notation of [2].

Because a physical camera may have different scales S^u and S^v for the pixels this will result in two different parameters $f^u = \frac{D+d}{S^u}$ and $f^v = \frac{D+d}{S^v}$. We obtain the relation of a point P and the plenoptic disc data (w^s, w^t) and R as

$$P = -\frac{rK_2}{rK_1 + R}\left(\frac{w^s - c^u}{f^u}, \frac{w^t - c^v}{f^v}, 1\right). \tag{9}$$

The parameters f^u, f^v, c^u, c^v, K_1 and K_2 are the intrinsics we estimate for from plenoptic disc data. They are sufficient to provide point estimates using (9). This relation is bijective, and determining the plenoptic disc data (w^s, w^t, R) determines entirely the point P corresponding to it, and vice-versa, if the extrinsics and intrinsics of the camera are known. The projection of a point P to the triple (w^s, w^t, R) is called the *plenoptic projection*, denoted Π, and is given by

$$\Pi(P) = \left(-f^u \frac{P^x}{P^z} + c^u, -f^v \frac{P^y}{P^z} + c^v, -\frac{rK_2}{P^z} - rK_1\right). \tag{10}$$

In summary, we model a plenoptic camera in terms of a projection that sends a point P to a triple (w^s, w^t, R), called the *plenoptic disc data*, where (w^s, w^t) are lenslet coordinates, called the *plenoptic disc centre*, and R is a signed radius called the *plenoptic disc radius*.

Because the triple (w^s, w^t, R) can be determined purely from raw light-field data, we can use this feature data, together with knowledge of the true positions of the feature points they correspond to, to estimate the intrinsics and extrinsics of the camera using (10).

3.1. Distortion Model

Although the focus lens used with the Raytrix R42 camera in the experiments has a negligible lens distortion of 0.02% total variation [9], Lytro cameras suffer from significant lens distortion, which affects only plenoptic disc centre coordinates in our projection model. We model the effect of plenoptic disc distortion with a first-order approximation

$$(w_u^s, w_u^t) = (1 + k_1\delta^2)((w_d^s, w_d^t) - (c^u, c^v)) + (c^u, c^v) \tag{11}$$

where (w_u^s, w_u^t) are the undistorted plenoptic disc coordinates and (w_d^s, w_d^t) are the distorted coordinates and δ is the distance in pixels from the distorted plenoptic disc centre to the optical centre $\delta = ||(w_d^s, w_d^t) - (c^u, c^v)||$. This one-parameter lens distortion model corrects the majority of the observed distortion in the raw Lytro images.

4. Plenoptic Camera Calibration

Calibration of a plenoptic camera has three main blocks, see Fig. 4: the first is a feature estimation block. In this paper, we estimate the plenoptic disc data corresponding to corners of a checkerboard (cf. Sections 3 and 4.1). The second block is an initialisation block, that produces a calibration parameter estimate (cf. Section 4.2). The third block is a non-linear optimisation routine that refines the initial estimate produced by the second block (cf. Section 4.3).

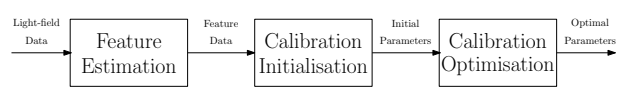


Figure 4: A block diagram of a generic calibration method.

4.1. Feature Estimation

We propose a novel plenoptic feature estimation method that avoids problems associated with identifying features in the low-resolution subimages by instead using higher-resolution sub-aperture images [4] that are computed from raw light-field data. Other papers [2, 13] extract features from the low-resolution subimages. An additional advantage of the proposed method is lower memory usage. Our high-resolution Raytrix data is ill-suited for the method proposed in [2], as the template size used in that paper grows geometrically with subimage radius, resulting in that algorithm failing to terminate in our experiments.

We assume that the calibration grid consists of M interior corners, and that a corner point has body-fixed-frame coordinates P_i indexed by $i = 1, \ldots, M$. The first step of

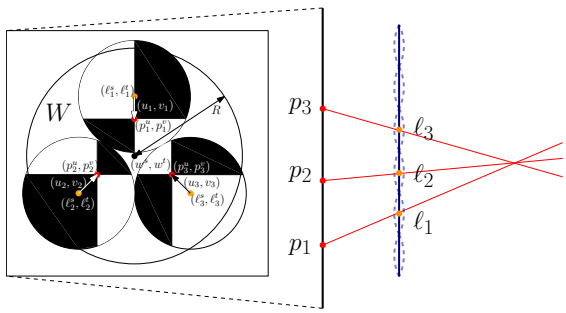


Figure 5: The centre (w^s, w^t) and radius R of a plenoptic disc W is shown on a raw light-field image, cf. also Fig. 7. The three lenslets in W are labelled by their lenslet coordinates (ℓ_1^s, ℓ_1^t), (ℓ_2^s, ℓ_2^t), (ℓ_3^s, ℓ_3^t) that are in the centres of the subimages of these lenslets. The pixels (p_1^u, p_1^v), (p_2^u, p_2^v), and (p_3^u, p_3^v) within the subimages of each of these lenselets corresponding to the same feature point are depicted. These pixels have offsets (u_1, v_1), (u_2, v_2), and (u_3, v_3) from the subimage centre, repectively.

calibration is estimation of the plenoptic disc feature data (w_i^s, w_i^t, R_i) corresponding to each of the corner points P_i of the calibration grid for the given raw light-field image. This step is repeated for each raw light-field image.

Our feature estimation process starts by obtaining a set of N sub-aperture images I_k indexed by $k = 1, \ldots, N$ by selecting from each lenslet subimage the pixel with constant offset (u_k, v_k) from the subimage centre and stitching the resulting image together. Because the lenslets are arranged in an hexagonal lattice, generating a rectangular sub-aperture image from the constant offset pixels of each subimage requires interpolation. The dimensions of the subimage are $\frac{U}{r}$ and $\frac{V}{r}$, where $U \times V$ is the dimension of the raw light-field image and r is the lenslet subimage radius.

For the interpolation, the colour $I_k(q^u, q^v)$ assigned to pixel (q^u, q^v) in the sub-aperture image I_k is given by finding the three nearest lenslet coordinates $(\ell_1^s, \ell_1^t), (\ell_2^s, \ell_2^t), (\ell_3^s, \ell_3^t)$, of lenslets in the subimage to $r(q^u, q^v)$. Then, $r(q^u, q^v)$ is expressed as a convex sum of these coordinates so that $r(q^u, q^v) = \alpha_1(\ell_1^s, \ell_1^t) + \alpha_2(\ell_2^s, \ell_2^t) + \alpha_3(\ell_3^s, \ell_3^t)$, where $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Then we assign the colour to pixel q that respects this convex sum, so that $I_k(q^u, q^v) = \alpha_1 L(\ell_1^s + u_k, \ell_1^t + v_k) + \alpha_2 L(\ell_2^s + u_k, \ell_2^t + v_k) + \alpha_3 L(\ell_3^s + u_k, \ell_3^t + v_k)$.

Running a standard checkerboard detector on each of the resulting sub-aperture images I_k gives a list of detected checkerboard corner features $F_k = \{(q_{i,k}^u, q_{i,k}^v)\}_{i=1}^M$ appearing in the image. Note that the set F_k of point features is indexed by the offset index k, whereas individual point features within F_k are indexed by both the point index i and the offset index k.

Scaling any detected corner feature $(q_{i,k}^u, q_{i,k}^v) \in F_k$ by r gives the lenslet coordinates $(\ell_{i,k}^s, \ell_{i,k}^t)$ of the lenslet $\ell_{i,k}$ for which P_i is visible in the subimage of $\ell_{i,k}$ with offset (u_k, v_k) from the subimage centre $(\ell_{i,k}^s, \ell_{i,k}^t)$. As P_i is visible in the subimage of $\ell_{i,k}$, P_i projects to some $p_{i,k}$ defined by equations (2) and (5). This pixel has pixel coordinates $(p_{i,k}^u, p_{i,k}^v) = (\ell_{i,k}^s, \ell_{i,k}^t) + (u_k, v_k)$, because $q_{i,k}$ was found in a sub-aperture image generated with constant offset (u_k, v_k).

Therefore, from a raw light-field image we can obtain a collection of lenslet-pixel pairs $\{((\ell_{i,k}^s, \ell_{i,k}^t), (p_{i,k}^u, p_{i,k}^v))\}_{i=1,k=1}^{M,N}$ corresponding to the N corners as seen in M sub-aperture images.

Now, for any two of these obtained lenslet-pixel pairs $((\ell_{i,k}^s, \ell_{i,k}^t), (p_{i,k}^u, p_{i,k}^v))$ and $((\ell_{i,k'}^s, \ell_{i,k'}^t), (p_{i,k'}^u, p_{i,k'}^v))$ corresponding to some P_i, we note that

$$\begin{pmatrix} p_{i,k}^u - p_{i,k'}^u \\ p_{i,k}^v - p_{i,k'}^v \end{pmatrix} = \left(1 + \frac{r}{R_i}\right) \begin{pmatrix} \ell_{i,k}^s - \ell_{i,k'}^s \\ \ell_{i,k}^t - \ell_{i,k'}^t \end{pmatrix}. \tag{12}$$

Note that when $(u_{k'}, v_{k'}) = (0, 0)$, we have $(\ell_{i,k'}^s, \ell_{i,k'}^t) = (w_i^s, w_i^t)$, because a point feature will appear in the centre of the subimage of the plenoptic disc centre. Therefore, (12), with $(u_{k'}, v_{k'}) = (0, 0)$ provides a linear system of equations that can be used to estimate the plenoptic disc feature data (w_i^s, w_i^t, R_i) by solving

$$\begin{pmatrix} -1 & 0 & -\frac{u_k}{r} & \ell_k^s \\ 0 & -1 & -\frac{v_k}{r} & \ell_k^t \end{pmatrix} \begin{pmatrix} w_i^s \\ w_i^t \\ R_i \\ 1 \end{pmatrix} = 0, \tag{13}$$

where there are 2 rows in the data matrix for each offset index k for which the corner P_i is successfully detected in the sub-aperture image I_k. In practice, we only use sub-aperture images I_k for which all M checkerboard corners are successfully detected.

In summary, the method for obtaining the plenoptic disc feature data estimates for a single raw light-field image is as follows:

1. For each pixel offset (u_k, v_k), generate the sub-aperture image I_k.

2. For each sub-aperture image I_k, run a standard corner detector to obtain a set of sub-aperture image features $q_{i,k}$.

3. Compute the corresponding lenslet-pixel pairs $((\ell_{i,k}^s, \ell_{i,k}^t), (p_{i,k}^u, p_{i,k}^v))$.

4. Using the lenslet-pixel pairs corresponding to a given point feature P_i, find the least-squares estimate of (w_i^s, w_i^t, R_i) by solving (13).

These steps are applied to every raw light-field image in the dataset.

4.2. Calibration Initialisation

Initialisation parameters are found by deriving a linear system of equations that perfect data from a single light-field image must satisfy, and solving the system for gathered data in a least-squares sense. Let $X_j \in SE(3)$ denote the pose of the camera with respect to the fixed frame $\mathbf{O}$ when it captures the raw light field image frame j. Let ${}^{\mathbf{O}}P_i$ be the position of the corner of a checkerboard expressed in the coordinates of the fixed frame $\mathbf{O}$. At frame j, the corner P_i has coordinates $P_{i,j}$ in the body-fixed frame of the camera. Using (10) with $f^u = f^v = f$, we obtain the relations

$$w^s P^z_{i,j} + f P^x_{i,j} = 0 \tag{14}$$

$$w^t P^z_{i,j} + f P^y_{i,j} = 0 \tag{15}$$

$$(rK_1 + R)P^z_{i,j} + rK_2 = 0. \tag{16}$$

Denoting $X_j^{-1} = (\Omega_j, \tau_j)$ with Ω_j the rotational part and τ_j the translational part, we have

$$\begin{pmatrix} P^x_{i,j} \\ P^y_{i,j} \\ P^z_{i,j} \end{pmatrix} = \begin{pmatrix} \Omega_j^{11} & \Omega_j^{12} & \Omega_j^{13} \\ \Omega_j^{21} & \Omega_j^{22} & \Omega_j^{23} \\ \Omega_j^{31} & \Omega_j^{32} & \Omega_j^{33} \end{pmatrix} \begin{pmatrix} {}^{\mathbf{O}}P^x_i \\ {}^{\mathbf{O}}P^y_i \\ {}^{\mathbf{O}}P^z_i \end{pmatrix} + \begin{pmatrix} \tau^x_j \\ \tau^y_j \\ \tau^z_j \end{pmatrix}. \tag{17}$$

Substituting the expressions for $P_{i,j}$ into (14) - (16), we obtain the system

$$\left(\begin{pmatrix} w^s_{i,j} & 1 & 0 & 0 \\ w^t_{i,j} & 0 & 1 & 0 \\ R_{i,j} & 0 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} {}^{\mathbf{O}}P^x_i & {}^{\mathbf{O}}P^y_i & 1 \end{pmatrix} \right) \begin{pmatrix} \Omega_j^{31} \\ \Omega_j^{32} \\ \tau^z_j \\ -f\Omega_j^{11} \\ -f\Omega_j^{12} \\ -f\tau^x_j \\ -f\Omega_j^{21} \\ -f\Omega_j^{22} \\ -f\tau^y_j \\ rK_1\Omega_j^{31} \\ rK_1\Omega_j^{32} \\ rK_2 + rK_1\tau^z_j \end{pmatrix} = 0.$$

Here, $\otimes$ denotes the Kronecker product. For each frame j, this system is solved for the vector which satisfies the equation in a least sqaures sense. Extracting the initial intrinsics and extrinsics from the resulting singular vectors follows the same procedure as in [2].

4.3. Calibration Optimisation

The plenoptic disc feature data estimates are used as data in a non-linear optimisation routine, where the intrinsics and extrinsics of the camera are the parameters being estimated. Initial parameter estimates are given using the linear solution obtained in Section 4.2. The separate f-parameters f^u and f^v are initialised with the same f given in the initialisation step, the initial estimate of the optical centre (c^u, c^v) is the centre of the light-field image, and the initial lens distortion parameter k_1 is 0.

The error function that is minimised in this routine comes from (10), and (17), where the plenoptic disc feature data $(w^s_{i,j}, w^t_{i,j}, R_{i,j})$ have been estimated using the method discussed in Section 4.1, and world-frame point locations ${}^{\mathbf{O}}P_i$ are known. Let $\Lambda = (K_1, K_2, f^u, f^v, c^u, c^v, k_1)$ be the intrinsics of the camera, $\Xi = \{X_j\}$ be the set of extrinsics of the camera, and $\Phi = \{(w^s_{i,j}, w^t_{i,j}, R_{i,j}, {}^{\mathbf{O}}P_{i,j})\}$ be the known data. We minimise the *plenoptic reprojection error*, given by

$$\epsilon(\Lambda, \Xi; \Phi) = \sum_{i,j} (\Pi_\Lambda(P_{i,j}) - (w^s_{i,j}, w^t_{i,j}, R_{i,j}))^2 \tag{18}$$

where $P_{i,j} = X_j^{-1}\, {}^{\mathbf{O}}P_i$, given by (17), and Π_Λ denotes the plenoptic projection (10) with lens distortion modelled by (11), and parameters given by the intrinsic parameter estimate Λ.

5. Results

In this section we compare the proposed calibration algorithm to existing state-of-the-art methods proposed in [2, 3, 13]. Our code is publicly available[1].

5.1. Experimental Data

For our obtained datasets, a Raytrix R42 camera was used with a Kowa LM35SC 35mm focus lens [9]. The approximate focus distance was set to 0.25m, 0.5m, and 1m. We call these datasets R-A, R-B, and R-C, respectively.

A standard checkerboard was used as a calibration grid for the various experiments. As at shorter focal distances, the camera needed to be closer to the calibration target, the grid sizes for datasets R-A, R-B, and R-C were 4mm, 6mm, and 15.5mm, respectively. These datasets contained 24, 22, and 18 images, respectively. For datasets R-A and R-C the checkerboards contained 15 by 10 feature points. For dataset R-B, the checkerboard contained 6 by 8 feature points. The obtained resolutions of the raw light-field images were 7716 pixels (width) by 5364 pixels (height). Debayering and colour-correction was conducted upon capture using Raytrix software.

Lenslet types were identified in the raw light-field images using a standard method. Since the pinhole models for each lenslet type are identical, we consider only calibration for the lenslet types where the calibration grid is in best focus. Exploiting the multi-focal arrangement would likely improve the feature-extraction process but was not considered in this paper.

We also compared results on the datasets given in [3]. In those datasets, referred to here as L-B, L-D, and L-E, a Lytro camera was used. These datasets were chosen because each used a different focal distance and contained light-fields at a wide variety of poses and raw light-field images of varying degrees of focus.

[1] Available at: https://github.com/sgpobrien/PlenCalToolbox

The cells in Table 1 for Nousias *et al*. [13] comparing against the Lytro datasets were left blank in Table 1 because the small resolution of the Lytro subimages produces poor results for the feature-detection method of [13]. As the method proposed in [13] relies on detecting features in subimages, it is expected that their method would not fare well with these datasets. It is noted that the focal lenses of Lytro cameras have far higher lens distortion than the focal lenses used by Raytrix cameras. Since the method in [13] does not model lens distortion, comparing on Lytro data would be an unfair comparison. As dataset R-A also contained many in-focus lightfields where features are not visible in subimages, we did not compare [13] on this dataset.

As the method proposed by Dansereau *et al*. [3] relies on a particular formatting of the raw light-field images, it does not produce estimates for the Raytrix datasets, and as such the corresponding cells in Table 1 are left blank. On our version of Matlab, running the code of [3] on dataset L-D resulted in an exception being thrown to do with inconsistent orientations of the detected checkerboard, so these cells are left blank. For the same formatting reason, comparison of Bok *et al*. [2] on the Raytrix datasets was not conducted. We did not run the code of [2] on the dataset L-D because, as the authors note in that paper, their feature detection method does not work for in-focus light-fields, of which there are many in dataset L-D.

5.2. Performance Measures

The aim of these experiments was to test the accuracy of our calibration under a variety of conditions. However, the sparsity of Table 1 indicates that there is no standard performance measure to verify calibration methods for plenoptic cameras. Comparisons were made difficult by the wide variety of types of errors reported between each of the calibration methods. Our method calculates the widest variety of errors, rather than only the error being optimised. Note that any method that optimises a given error should have a natural advantage when that error is used as a performance measure. As such, we compare against a variety of errors rather than our optimised plenoptic reprojection error. Mean 3D reconstruction error (M3DE) is calculated by taking the average of the distances of point estimates from the actual points and dividing by the depth of the actual point in the estimated camera coordinate system. Note that plenoptic cameras allow single-image 3D reconstructions, making this a sensible measure. Mean reprojection error (MRE) is the average of the distances between extracted subimage corner feature coordinates and reprojected corner feature coordinates on the raw image, an example of which is given in Fig. 7. Mean sub-aperture reprojection error (MSRE) is the average of distances between extracted feature coordinates in sub-aperture images and reprojected feature coordinates onto those sub-aperture images.

As the code provided by each of the other methods [13, 3, 2] does not produce 3D reconstructions we derived several methods based on the projection models used in these papers and the obtained feature data of these methods. For Nousias, the reconstruction method finds the point that best fits their projection model [13], given a set of lenslet-pixel pairs known to correspond with that point. The reconstruction method used for Dansereau *et al*. [3] is based on calculating for all the lenslet-pixel pairs corresponding to a point P, the ray corresponding to that lenslet-pixel pair and finding the point that minimises the sum of distances to all of these rays. We were unable to find a reliable reconstruction method for the data obtained by Bok *et al*. [2].

The MSRE method used for Dansereau *et al*. [3] in Table 1 is based on solving their projection model given known ideal checker positions and pixel offsets for the unknown lenslet coordinates, which has a direct solution in their projection model, then applying their distortion model.

5.3. Discussion

In Table 1 we show the results of our calibration method compared against other existing state-of-the-art methods. Our method runs on the widest variety of datasets, and most consistently produces the smallest errors. The code we compared our method to were supplied by Nousias *et al*. [13], Dansereau *et al*. [3], and Bok *et al*. [2].

In Table 1 we compare each of the methods on different performance measures where it was possible to do so with the feature data produced by these methods. We first compared the methods on the measure of M3DE. As we were not able to provide reconstructions for the method of Bok *et al*. [2], this column is left blank. On this metric our method outperforms all the other methods except on dataset L-E, where the method of Dansereau *et al*. [3] performs better. This is likely due to their method implementing better preconditioning and higher-order lens distortion. One of our reconstructions together with extrinsics is shown in Fig. 6. The high M3DEs for Nousias *et al*. [13] are likely due to a flaw in their implementation discussed in the following paragraph.

In Table 1 we then compare our method with the method of Nousias *et al*. [13] on the measure of MRE. Although it was not possible to calculate this error using the feature data provided by the other methods [3, 2], our results for this measure are still shown. One likely factor affecting the accuracy of Nousias *et al*. [13] is preconditioning. It is noted that appropriate centering and scaling of parameters is often essential in order for Matlab-based optimisation algorithms to converge [11]. The algorithm implemented in [13] does not implement any centering or scaling, and the results for [13] reported in Table 1, can be improved for each of the datasets using both our feature data and scaling factors. These scaling factors did not significatly improve their results when using their feature data, however, suggesting that there may also be inaccuracies in their feature

Table 1: Table of error results. Best results per row are shown in bold. Measures that could not be computed are left blank.

Dataset	Ours	Nousias [13]	Dansereau [3]	Bok [2]
Mean 3D Reconstruction Error (%)				
R-A	**0.5206**			
R-B	**0.4482**	28.5775		
R-C	**1.4274**	53.7746		
L-B	**1.8642**		2.0419	
L-D	**4.2736**			
L-E	8.7459		**5.9599**	
Mean Reprojection Error (pixels)				
R-A	**0.9743**			
R-B	**0.2619**	2.0104		
R-C	**0.3832**	4.6925		
L-B	**0.3467**			
L-D	**1.2443**			
L-E	**0.2802**			
Mean Sub-aperture Reprojection Error (pixels)				
R-A	**0.5750**			
R-B	**1.0751**			
R-C	**0.6588**			
L-B	0.3427		**0.1775**	1.3125
L-D	**0.3061**			
L-E	**0.3514**		0.7383	0.3552

estimation step. A comparison between the intrinsics obtained using our method versus Nousias *et al*. is shown in Table 2. The accuracy of our feature estimation step and reprojections are demonstrated in Fig. 7.

In Table 1 we compare our results to the other methods on the measure of MSRE. Although it was not possible to calculate this error using the feature data of Nousias *et al*. [13], our results for these datasets are still shown. Our method outperforms the other proposed methods on this metric with the exception of dataset L-B, where it is beaten by Dansereau *et al*. [3], likely due to their higher-order approximation of lens distortion. Note that although the MSRE is smaller than ours for this cell, its M3DE is larger, demonstrating the non-transitive relation between these measures.

6. Conclusions and Future Work

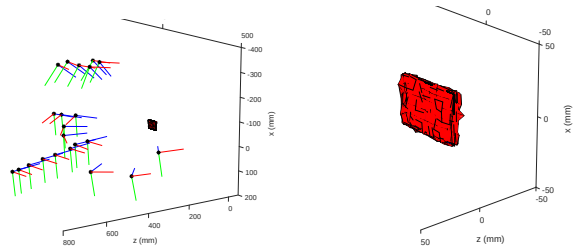
In this paper we use a novel plenoptic feature type, namely plenoptic disc features, for use in plenoptic camera calibration. Our method produces both superior mean reprojection errors onto the raw light-field images and better mean reconstruction errors.

To the authors' knowledge, the proposed calibration implementation is the first that successfully and reliably runs with both Raytrix and Lytro data with only minor preprocessing required.

Along with better performance, our method provides a novel projection model that allows an easy translation between plenoptic disc features and physical 3D points, making it better suited for 3D reconstruction than ray-based approaches.

Table 2: Intrinsic Parameters for Dataset R-B.

Var.	Ours	Nousias
K_1	-13.1706	-10.23
K_2	1.14×10^4	1.13×10^4
f^u	3.21×10^4	3.18×10^4
f^v	3.21×10^4	3.18×10^4
c^u (pix)	2675	2681
c^v (pix)	4415	3857
k_1	-1.7×10^{-10}	0



(a) Extrinsics with reconstructions. (b) Reconstructions overlayed.

Figure 6: Calibration grid reconstruction and poses for dataset R-B. Camera faces forwards along blue axis.

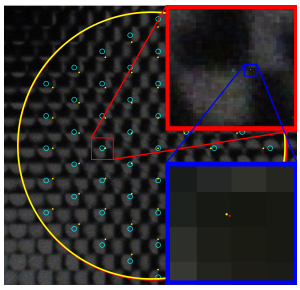


Figure 7: An example of a plenoptic disc feature and plenoptic reprojection of a point on a raw light-field image from dataset R-B. In the cyan circles are the lenslet coordinates (ℓ^s, ℓ^t) within a plenoptic disc W shown with red boundary and a reprojected plenoptic disc shown with yellow boundary. The red dots are the estimated subimage features corresponding to the plenoptic disc feature, and the yellow dots are the reprojected features.

7. Acknowledgements

This research was supported by the Australian Research Council through the ARC Discovery Project DP160100783 "Sensing a complex world: Infinite dimensional observer theory for robots".

References

[1] F. Bergamasco, A. Albarelli, L. Cosmo, A. Torsello, E. Rodolà, and D. Cremers. Adopting an unconstrained

ray model in light-field cameras for 3d shape reconstruction. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3003–3012, June 2015. 1

[2] Y. Bok, H. G. Jeon, and I. S. Kweon. Geometric calibration of micro-lens-based light field cameras using line features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):287–300, 2017. 1, 2, 3, 4, 6, 7, 8

[3] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1027–1034, June 2013. 2, 6, 7, 8

[4] C. Hahne. The standard plenoptic camera. http://www.plenoptic.info/, 2018. 4

[5] C. Heinze, S. Spyropoulos, S. Hussmann, and C. Perwass. Automated robust metric calibration algorithm for multifocus plenoptic cameras. *IEEE Transactions on Instrumentation and Measurement*, 65(5):1197–1205, May 2016. 1, 2

[6] R. E. Jacobson, N. Axford, S. Ray, and G. G. Attridge. *Manual of Photography: Photographic and Digital Imaging*. Butterworth-Heinemann, Newton, MA, USA, 9th edition, 2001. 2

[7] Z. Ji, C. Zhang, and Q. Wang. Light field camera self-calibration and registration. *SPIE Photonics Asia*, 10020:10, 2016. 1, 2

[8] O. Johannsen, C. Heinze, B. Goldluecke, and C. Perwaß. *On the Calibration of Focused Plenoptic Cameras*, pages 302–317. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. 1, 2

[9] Kowa. *LM35SC Data Sheet*, 2018. 4, 6

[10] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *2009 IEEE International Conference on Computational Photography*, pages 1–8, 2009. 1

[11] MathWorks. *When the Solver Fails*, 2018. 7

[12] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report*, 2(11):144–162, 2005. 2

[13] S. Nousias, F. Chadebecq, J. Pichat, P. Keane, S. Ourselin, and C. Bergeles. Corner-based geometric calibration of multi-focus plenoptic cameras. In *IEEE International Conference on Computer Vision*, 2017. 1, 2, 3, 4, 6, 7, 8

[14] M. Potmesil and I. Chakravarty. Synthetic image generation with a lens and aperture camera model. *ACM Trans. Graph.*, 1(2):85–108, Apr. 1982. 2

[15] K. H. Strobl and M. Lingenauber. Stepwise calibration of focused plenoptic cameras. *Computer Vision and Image Understanding*, 145:140 – 147, 2016. 1, 2

[16] J. Sun, C. Xu, B. Zhang, S. Wang, M. M. Hossain, H. Qi, and H. Tan. Geometric calibration of focused light field camera for 3-d flame temperature measurement. In *IEEE International Instrumentation and Measurement Technology Conference*, pages 1–6, 2016. 1, 2

[17] N. Zeller, F. Quint, and U. Stilla. Depth estimation and camera calibration of a focused plenoptic camera for visual odometry. *Journal of Photogrammetry and Remote Sensing*, 118:83 – 100, 2016. 1, 2

[18] N. Zeller, F. Quint, and U. Stilla. From the calibration of a light-field camera to direct plenoptic odometry. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1004–1019, 2017. 1, 2

[19] C. Zhang, Z. Ji, and Q. Wang. Decoding and calibration method on focused plenoptic camera. *Computational Visual Media*, 2(1):57–69, 2016. 1, 2