Corner-Based Geometric Calibration of Multi-Focus Plenoptic Cameras

Sotiris Nousias[1,2]   Francois Chadebecq[1,2]   Jonas Pichat[2]   Pearse Keane[3]
Sebastien Ourselin[1,2]   Christos Bergeles[1,2]

[1]Wellcome/EPSRC Centre for Interventional and Surgical Sciences, UCL, London, UK
[2]Centre for Medical Image Computing, UCL, London, UK
[3]NIHR Biomedical Research Centre, Moorfields Eye Hospital, London, UK

sotiris.nousias.15@ucl.ac.uk

Abstract

We propose a method for geometric calibration of multi-focus plenoptic cameras using raw images. Multi-focus plenoptic cameras feature several types of micro-lenses spatially aligned in front of the camera sensor to generate micro-images at different magnifications. This multi-lens arrangement provides computational-photography benefits but complicates calibration. Our methodology achieves the detection of the type of micro-lenses, the retrieval of their spatial arrangement, and the estimation of intrinsic and extrinsic camera parameters therefore fully characterising this specialised camera class. Motivated from classic pinhole camera calibration, our algorithm operates on a checker-board's corners, retrieved by a custom micro-image corner detector. This approach enables the introduction of a reprojection error that is used in a minimisation framework. Our algorithm compares favourably to the state-of-the-art, as demonstrated by controlled and free-hand experiments, making it a first step towards accurate 3D reconstruction and Structure-from-Motion.

1. Introduction

Classic pinhole cameras project the multi-dimensional information of the light flowing through a scene into a single 2D snapshot. This projection limits the information that can be reconstructed from the 2D acquisition. Plenoptic (or light-field) cameras, on the other hand, capture a 4D slice of the plenoptic function [1], termed the "light-field". These cameras provide both spatial and angular information on the light flowing through a scene; multiple views are captured in a single photographic exposure [1].

Adelson and Wang [2] used a lenticular array placed in front of the sensor of a traditional pinhole camera to create one of the first plenoptic cameras. Micro-manufacturing and computing advances allowed Ng to replace the lentic-
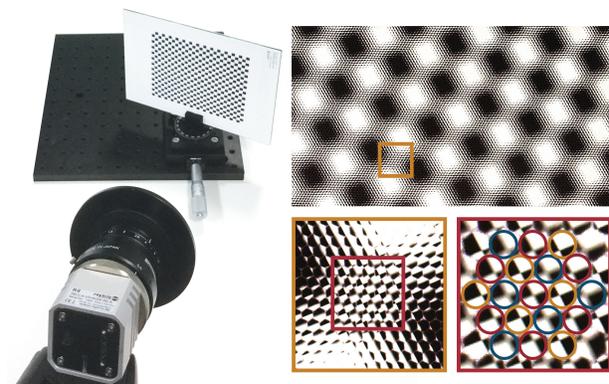

Figure 1: Illustration of a multi-focus plenoptic camera from Raytrix[TM] and a raw plenoptic image.

ular array with a micro-lens array positioned at the main lens' focal length (unfocused plenoptic camera), resulting in the first hand-held light-field camera [21]. Designs that balance the trade-off between angular and spatial resolution place the micro-lenses behind the focal plane of the main lens (Keplerian) [20], or in front (Galilean) [23, 10].

The multi-dimensional acquisitions of plenoptic cameras facilitate the recovery of 3D information. Using the encoded light-ray directions, images can be digitally refocused [21], while single-shot depth maps can be obtained by combining defocus and multiple-view cues [28]. Therefore, plenoptic cameras are increasingly considered in computer vision [23, 30, 16, 29, 18, 19], robotics [7, 5], and even life-sciences [24] and medical imaging [3, 26].

Using well-calibrated cameras will improve application and algorithm performance. As newer plenoptic cameras feature not a single but several different types of micro-lenses (multi-focus plenoptic cameras), calibrating the imaging system geometrically, *i.e.* end-to-end from 3D points to 2D image points, becomes critical.

Related Work: Johannsen *et al.* presented metric calibration and distortion correction for multi-focus plenoptic

cameras using a grid of circular patterns [17]. The thin-lens model [14] was used to project virtual depth values[1] from image space onto 3D space. Neither end-to-end image formation nor different micro-lens types were considered. Heinze *et al.* extended the work to include the tilt/shift of the main lens, accounting for micro-lens types [15] but without considering end-to-end image formation.

Strobl *et al.* utilized total focus[2] and virtual depth images in a calibration framework for multi-focus plenoptic cameras [27]. As total focus images implicitly rely on the estimated depth, this approach suffers from a causality dilemma between depth and calibration parameters. Further, the pixel size was assumed to be known. The paper concluded with the need for a calibration that uses raw plenoptic images while considering different micro-lens types.

Dansereau *et al.* in [6] were the first to deliver end-to-end geometric calibration for Keplerian light-field cameras with a single type of micro-lenses (Lytro[TM]). The pixel-to-ray correspondences were modelled using a thin-lens model (representing the main lens) and pinhole cameras (representing micro-lenses). Corners extracted from the sub-aperture images[3] of a checker-board (the calibration target) were used to obtain a 4D intrinsic camera parameter matrix. The limitation of this approach, also highlighted in [4], is that obtaining sub-aperture images (from different types of micro-lenses) assumes a calibrated camera.

The state-of-the-art in geometric calibration for light-field cameras using raw images with a single type of micro-lenses (Lytro[TM] and Illum[TM]), [4], is based on the thin-lens-and-pinhole model of [6]. Line features, representing the edges of the checkers of the calibration target are extracted and used to obtain a solution for the intrinsic and extrinsic camera matrices. It was argued that line features are required because the resolution of micro-images in light-field cameras is not suitable for corner detection.

Different types of micro-lenses, *i.e.* multi-focus plenoptic cameras, cannot be handled in [4], which was, in addition, not developed for Galilean cameras. Further, even though lines are abundant in the raw images acquired by Lytro[TM]/Illum[TM], this is contrary to those acquired by multi-focus plenoptic cameras, such as Raytrix[TM], that have a higher spatial (micro-image) resolution at the expense of lower angular resolution (micro-lens-array cardinality) [23]. Finally, corner-based calibration overperforms line-based calibration as it enables the introduction of a 3D-to-2D reprojection error that is a representative performance measure of end-to-end imaging-system models.

Contributions: This manuscript considers the calibration of multi-focus plenoptic cameras, namely Raytrix[TM],

using raw images. The contributions are:
- Algorithm for corner-detection in micro-images of checker-boards.
- Micro-lens classification and spatial-arrangement retrieval based on focus measures.
- The equivalence of Galilean and Keplerian plenoptic camera models.
- Corner-based geometric calibration of all micro-lens types in a plenoptic camera.

Improving the state-of-the-art, our method minimizes the reprojection error in the raw plenoptic image. Further, we address the unsolved problem of multi-focus plenoptic camera calibration. We demonstrate precise calibration with a small number of images and validate the recovered extrinsics against known camera motion. Code is provided[4].

2. 3D-to-2D Corner Matching

A crucial step in any calibration algorithm is to pair 3D points on the calibration grid with their corresponding 2D image points. In a pinhole camera, a single 3D point corresponds to a single 2D point. Further, the spatial arrangement of checker-board corners in 3D space is assumed for the 2D corners on the image as well. In plenoptic images, a single 3D corner on the checker-board corresponds to multiple 2D projected corners, see Fig. 1.

Assuming that each micro-image contains at most one 2D corner, *i.e.* the checkers are sufficiently large, and that neighbouring micro-images contain 2D projections of the same 3D corner, we extract 3D-to-2D feature matches by:
1. Estimating micro-lens centres and micro-images.
2. Detecting corners in micro-images.
3. Retrieving 3D-to-2D correspondences.

Following correspondence retrieval, 2D corners are used to classify micro-lenses and retrieve their spatial arrangement prior to geometric camera calibration.

Micro-lens centre detection uses correlation methods such as in [6]. Given the centres, the vignetting caused by each micro-lens in the raw plenoptic image guides the segmentation into micro-images. The pixels contained in the square circumscribed to the circle delimited by the vignetting define the micro-image to be extracted. Hereafter, we assume retrieved micro-lens centres and micro-images.

2.1. Robust Corner Detection in Micro-Images

Early experimentation indicated that existing detectors cannot reliably identify checker-board corners in micro-images or report their absence, as they examine local patches that may contain strong gradient changes due to noise but do not necessarily represent corners.

We developed a dedicated corner detector. When a corner exists in a micro-image, it will be found at a "saddle/minimax" point in the intensity domain. The two axes,

---

[1]Virtual depth refers to relative depth values obtained from disparity.

[2]The total focus image is synthesized from the raw image provided the virtual depth. It corresponds to an all-in focus image of the scene [23].

[3]Each sub-aperture image displays light incident on the sensor from a small region of the aperture [21].

| (a) | (b) | (c) | (d) |
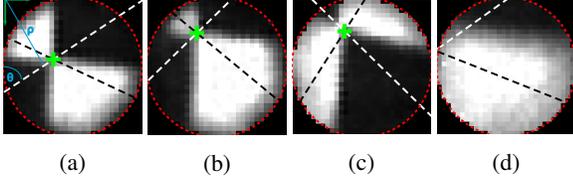
Figure 2: Retrieval of corners, (a)-(c), or their absence (d).

Method	Mean ± Std	Maximum	Minimum
Proposed	1.16 ± 1.54	6.97	0.001
Harris [12]	3.14 ± 2.56	8.99	0.001
FAST [25]	6.19 ± 2.26	9.99	0.187
Förstner [8]	7.9 ± 3.66	16.8	0.19
Hough based	4.5 ± 3.52	14.4	0.02

Table 1: Error [in pixels] of corner detectors.

not necessarily orthogonal, centred at the corner, define the directions of the saddle. Our algorithm relies on the observation that these two directions are lines[5] of maximum and minimum accumulated intensity traversing the image, *e.g.* black and white dotted lines in Fig. 2. Retrieval of these lines identifies two potential saddle axes, with the corner lying at their intersection.

More specifically, we assume the line is of the form $\rho = \cos(\theta)x + \sin(\theta)y$, where ρ is the distance between the line and the origin, and θ is the angle between the line and the y (vertical) axis (see Fig. 2a). The algorithm operates on a $K \times K$ micro-image $I : [1, K] \times [1, K] \to [0, 1]$ as follows.

Sets P, Θ of equidistant values $\rho \in \{0, \cdots, K\sqrt{2}\}$ and $\theta \in \{0, \cdots, \pi\}$ are created with ρ_{disc}, and θ_{disc}, discretisation steps, respectively.

Combinations $L = (P_i, \Theta_j) \in P \times \Theta$ generate a set of lines $\mathcal{L}$, sorted by accumulated intensities for all pixels along the line:

$$\text{v}(L, I) = \sum_{x=1}^{K}\{I_{\text{xy}}|I_{\text{xy}} \geq t_{\text{h}}\} - \sum_{x=1}^{K}\{I_{\text{xy}}|I_{\text{xy}} \leq t_{\text{l}}\}, \quad (1)$$

where I_{xy} is the image intensity at (x, y), y is given from the equation of the line when x, P_i, Θ_j, are known and $t_{\text{h}}, t_{\text{l}}$, are respectively maximum and minimum intensity thresholds. Equation (1) maximises the number of high intensity pixels along the line, while minimising the number of low intensity pixels.

For the line L_1 with the highest accumulated intensity, a set $\widetilde{\mathcal{L}}$ of intersecting lines with parameters $\rho \in \{0, \cdots, K\sqrt{2}\}$ and $\theta = (\theta_{L_1} + \pi/2 + \beta)$, $\beta \in \{-\pi/5, \cdots, \pi/5\}$, (discretised by ρ_{disc} and θ_{disc}) is sorted by increasing $\text{v}(\widetilde{L}, 1 - I)$. $\widetilde{L}_1$ has the minimum accumulated intensity in the original micro-image I.

Then, the intersection of L_1 and $\widetilde{L}_1$ constitutes a valid corner if it is within the image and:

$$\left| \frac{\text{v}(L_1, I)}{\text{v}(\widetilde{L}_1, 1 - I)} - 1 \right| \leq \text{v}_{\text{thresh}}. \quad (2)$$

Figures 2a-2c show corners found in exemplary micro-images. No corner is found in Fig. 2d where (2) is violated. The black and white lines are L_1, and $\widetilde{L}_1$, respectively.

We compared our detector to Harris [12], FAST [25], Förstner [8] and Hough-lines intersection on 5014 micro-images containing manually annotated corners. Our

---
[5]The term "line" does not imply an edge or a "Hough line".

method, for $t_{\text{h}} = 0.7, t_{\text{l}} = 0.3, \text{v}_{\text{thresh}} = 1, \rho_{\text{disc}} = 0.25, \theta_{\text{disc}} = \pi/50$, detected 3187 corners, Harris 2807, FAST 1094, Förstner 3343 and intersection of Hough lines 3783. The detection precision, see Table 1, shows that our algorithm finds the most corners and has the best performance. Our parameters are empirically tuned but remain constant throughout the experiments. Perturbing the selected values by up to 20% increases the estimation error by up to 0.283 pixels, still outperforming other detectors.

2.2. 3D-to-2D Corner Correspondence Assignment

The corners in the raw plenoptic image are now detected, but the correspondence between each corner and a 3D point is still unknown. In this section, we demonstrate how to identify which micro-lens observes which corner of the checker-board using the central sub-aperture image.

This image is obtained by sampling the central pixel of each micro-image. It resembles one acquired by a pinhole camera, making it possible to extract corners using an off-the-shelf detector [9]. Each 2D corner corresponds to a single 3D corner and their spatial arrangement is retained.

Given the relationship between pixels and micro-images, the detected corners should approximately fall at the centre of their corresponding micro-images. Since neighbouring micro-lenses observe the 3D scene from multiple but closely related viewpoints, we can assume that if a 2D corner, detected in the sub-aperture image, belongs to micro-image under micro-lens (i, j), then N micro-lenses surrounding (i, j) also observe it.

This observation provides an estimate of the number of 2D corners that each 3D corner corresponds to, and allows the pairing of the 3D checker-board corners with the 2D corners detected in the previous step. By examining a large number of neighbouring micro-lenses and relying on the corner detection's results, the number of examined micro-images is refined to those only containing detected corners.

3. Micro-Lens Type/Arrangement Retrieval

Multi-focus plenoptic cameras contain micro-lenses of multiple focal lengths. The proposed classification is based on extracting focus measures from the micro-images and linking together the ones that present similar focus characteristics. The algorithm's input is a single plenoptic image
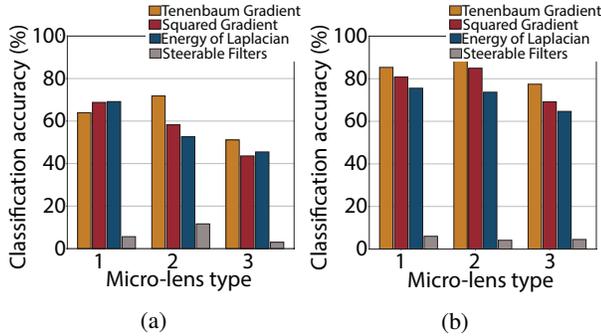
Figure 3: Micro-lens classification accuracy with 4 different metrics using (a) full micro-images, and (b) corner regions.

of the calibration target and the detected 2D corners along with the corresponding micro-image centres. In the following, only micro-images containing corners are considered.

3.1. Micro-Lens Classification

Since each 3D corner point lies on a single depth, only one micro-lens type generates in-focus micro-images. The remaining micro-lenses generate micro-images with varying degrees of defocus. Therefore, the degree of focus of each micro-image can be used to classify the micro-lenses into types. Classification is performed on sets of micro-images that view the same 3D corner, as they are grouped via the 3D-to-2D correspondences already obtained.

A circular region of radius r is extracted around the detected corner in each micro-image of the set under examination. The focus score of each region is computed relying on metrics presented in [22] and indicated in the legend of Fig. 3. The score is used to cluster micro-images into $k = 3$ types (made public by the manufacturer) with k-means algorithm. Such clustering method is reliable and robust, but alternative approaches that do no require the knowledge of the number of groups may be employed.

A comparison of focus metrics and their effect on classification is shown in Fig. 3 based on 330 clusters of 19 micro-lenses each. The Tenenbaum Gradient was the most robust focus metric. "Gradient" methods, in general, yield more correct classifications than "decomposition" methods, perhaps because they are less affected by the small resolution of the examined micro-image region.

Figure 3a and Fig. 3b highlight the importance of only considering a region around the detected 2D corner for classification. Assigning a single focus measure to the entire micro-image is sub-optimal because of vignetting. Further, when the checker-board corner is observed off-centre and obliquely, the black/white regions over the entire micro-image are imbalanced and affect the robustness of the focus metric. When using the entire micro-image, the interclass variance is small and leads to misclassifications.

3.2. Spatial Arrangement Retrieval

Assume N sets of micro-lenses, each set observing a different 3D corner. As labelling into types is performed locally, and different types of micro-lenses come into and out of focus depending on the checker-board's pose, the assigned types/labels are not consistent among the N examined sets. Also, there may be misclassified micro-lenses, as, for example, when all micro-images are in-focus.

A rectification strategy is implemented to compensate for the misclassifications and locality of labelling by identifying the classification that is the most consistent over the N sets under micro-lens label permutations. First, the lattices that represent the local micro-lens type arrangement are computed for every set. Second, the lattices are extended to create N candidate global grids that cover the entire micro-lens array. Third, a global configuration, the *candidate grid* is randomly chosen and compared against the $N - 1$ remaining ones and their label permutations. This candidate grid receives a single vote for every other configuration that it matches, and these matched configurations are now "fused" with the candidate grid. The process repeats considering only the unmatched grids as candidates, until no unmatched grids remain. The highest-voted candidate grid (most matches) is the global configuration.

An estimated spatial arrangement is illustrated in Fig. 1, where coloured circles represent different micro-lens types.

4. Projection Model and Calibration

Here, the plenoptic projection model for Galilean cameras is derived and used in calibration. The main lens is modelled as a thin lens and the micro-lenses as pinhole cameras [6, 4]. We do not consider radial distortion, which is absent from our images due to small field-of-view and high quality of lenses, but could be introduced according to [31].

4.1. Forward Projection Function

With reference to Fig. 4, the (red solid/dashed) rays refracted by the main lens converge to a 3D point that is behind the image sensor. In other words, the main lens creates a virtual image. This virtual image is then reprojected by the micro-lenses onto the image sensor. This is a unique characteristic of the Galilean camera model [17, 15] and, as discussed, is contrary to the Keplerian camera model wherein the main lens' focus is in front of the micro-lens array. In this section, the 3D point that the rays converge to is termed "image" (see Fig. 4). Since a multi-focus plenoptic camera contains a number of micro-lens types, it can be assumed that the model comprises an equal number of different pinhole cameras, three in our case. The following analysis is applied to each type of micro-lens independently, and our experiments will show that the approach provides different intrinsic parameters but the same extrinsics.
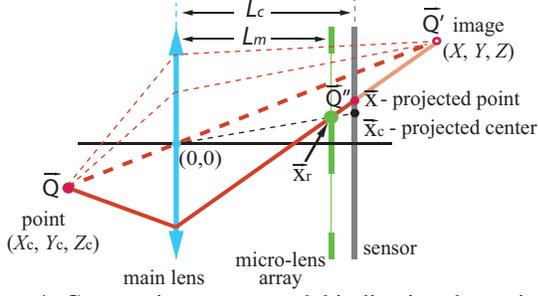
Figure 4: Geometric camera model indicating the main lens (thin lens), micro-lens array, virtual image, and acquired image (adapted from [4] for the Galilean design). Projected locations lie in the normalized coordinate system.

Following the notation of Bok *et al.*, the origin, and the z-axis of the camera coordinate system, are the centre, and the optical axis of the main lens, respectively. The relationship between the 3D point $\overrightarrow{Q} = (X_c, Y_c, Z_c)$ and its image $\overrightarrow{Q'} = (X, Y, Z)$ is derived from the thin lens equation [14]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{F}{F - Z_c} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}, \quad (3)$$

where F is the focal length of the main lens. As the micro-lenses are approximated as pinhole cameras, the projected location of $\overrightarrow{Q'}$ on the sensor, $\overrightarrow{Q''}$, can be found as the intersection of the sensor and a ray passing the micro-lens centre. This ray is called the "principal ray" of the micro-lens. With reference to Fig. 4, $\overrightarrow{Q''}$ is the intersection of the solid red ray and the solid black sensor. The ray is not highlighted after its intersection with the sensor, because it does not exist in reality, hence the term "virtual".

The micro-lenses', and sensor's, location in the camera coordinate system are L_m, and L_c, respectively. The micro-lens' real location in the camera coordinate system, $\overrightarrow{x}_r$, is computed via its projected location $\overrightarrow{x}_c = (x_c, y_c)$:

$$\overrightarrow{x}_r \equiv \begin{bmatrix} x_r \\ y_r \\ L_m \end{bmatrix} = L_m \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix}. \quad (4)$$

Projected locations are expressed in a normalised coordinate system, *i.e.* $z = 1$. Similarly, the location of the projected point in the camera coordinate system, $\overrightarrow{x}_i$, is:

$$\overrightarrow{x}_i \equiv \begin{bmatrix} x_i \\ y_i \\ L_c \end{bmatrix} = L_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (5)$$

Since the micro-lens centre $\overrightarrow{x}_r$, the projected point $\overrightarrow{Q'}$, and the image point $\overrightarrow{Q''}$ are collinear, Thales theorem gives:

$$\frac{x_i - x_r}{X - x_i} \equiv \frac{L_c x - x_r}{X - L_c x} = \frac{L_c - L_m}{Z - L_c}. \quad (6)$$

Inserting the physical micro-lens centre location, x_r, (4), in (6) and solving for the projected point location x:

$$x = \frac{L_m - L_c}{(L_m - Z)L_c}X + \frac{L_m(L_c - Z)}{(L_m - Z)L_c}x_c. \quad (7)$$

Normalised coordinates are used for x. Then, for the Galilean system we are examining:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{(L_m - Z)L_c} \begin{bmatrix} (L_m - L_c)X + L_m(L_c - Z)x_c \\ (L_m - L_c)Y + L_m(L_c - Z)y_c \end{bmatrix}, \quad (8)$$

which is identical to the projection equation for the Keplerian system examined in [4] proving the geometric equivalence of the two camera models.

The relation between normalised coordinates, (x, y), (x_c, y_c) and the image coordinates (u, v), (u_c, v_c) for the micro-lens pinhole camera is given by a 2×3 matrix [13]:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (9)$$

where skew is assumed zero, f_x, f_y, are the focal lengths along the $x-$, and $y-$ axes, respectively, and c_x, c_y are the principal points. Combining (8) with (9) leads to the 3D-to-2D projection model:

$$\begin{bmatrix} u - c_x \\ v - c_y \end{bmatrix} = \begin{bmatrix} \dfrac{f_x(L_m - L_c)X + L_m(L_c - Z)(u_c - c_x)}{(L_m - Z)L_c} \\ \dfrac{f_y(L_m - L_c)Y + L_m(L_c - Z)(v_c - c_y)}{(L_m - Z)L_c} \end{bmatrix}. \quad (10)$$

Next, we discuss retrieving the equation's parameters.

4.2. Calibration

The 3D corners $\overrightarrow{X}_w = (X_w, Y_w, Z_w)$ of the calibration checker-board are co-planar and, without loss of generality, $Z_w = 0$ can be assumed. $\overrightarrow{X}_w$ is expressed in the global coordinate frame, and is transformed to $\overrightarrow{Q}$ in the camera coordinate frame through an Euclidean transformation of rotation $\mathbf{R} \in SO(3)$ and translation $\overrightarrow{t} \in \mathbb{R}^{3 \times 1}$:

$$\overrightarrow{Q} = \mathbf{R}\overrightarrow{X}_w + \overrightarrow{t}. \quad (11)$$

In the following, we assume that the transformed 3D corners $\overrightarrow{Q}$ have been matched with 2D corners (u, v).

4.2.1 Initial Solution

For an initial solution of the camera parameters, let $(c_x, c_y) \equiv (0, 0)$, and $f_x \equiv f_y \equiv f$. By subtracting the imaged micro-lens centre, $\overrightarrow{u}_c$, and substituting (3) in (10), we obtain as in [4]:

$$\begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} \equiv \begin{bmatrix} u - u_c \\ v - v_c \end{bmatrix} = \frac{1}{K_1 Z_c + K_2} \begin{bmatrix} fX_c - Z_c u_c \\ fY_c - Z_c v_c \end{bmatrix}, \quad (12)$$

where:

$$K_1 = -\frac{(L_m + F)L_c}{(L_m - L_c)F}, \quad (13)$$

$$K_2 = \frac{L_m L_c}{L_m - L_c}. \quad (14)$$

Inserting (11) in (12) and rearranging, we obtain the system $\mathbf{A} \overrightarrow{x} = 0$:

$$\begin{bmatrix} -\overrightarrow{\tilde{\mathrm{X}}}_w & \overrightarrow{0}_{1\times3} & \overrightarrow{\tilde{\mathrm{X}}}_w u_c & \overrightarrow{\tilde{\mathrm{X}}}_w \Delta u \\ \overrightarrow{0}_{1\times3} & -\overrightarrow{\tilde{\mathrm{X}}}_w & \overrightarrow{\tilde{\mathrm{X}}}_w v_c & \overrightarrow{\tilde{\mathrm{X}}}_w \Delta v \end{bmatrix} \begin{bmatrix} fr_{11} \\ fr_{12} \\ ft_1 \\ fr_{21} \\ fr_{22} \\ ft_2 \\ r_{31} \\ r_{32} \\ t_3 \\ K_1 r_{31} \\ K_1 r_{32} \\ K_1 t_3 + K_2 \end{bmatrix} = 0$$

$$(15)$$

where $\overrightarrow{\tilde{\mathrm{X}}}_w = \begin{bmatrix} X_w & Y_w & 1 \end{bmatrix}$. The terms r_{ij} refer to the j-th element of i-th row of the rotation matrix $\mathbf{R}$.

Each 2D corner creates two constraints. Stacking all correspondences into matrix $\mathbf{A}$, the initial solution is the right singular vector, namely $\overrightarrow{\nu}$, corresponding to the smallest singular value. Imposing the orthogonality constraint on the columns of $\mathbf{R}$, the initial parameters are:

$$f = \sqrt{\frac{-\nu_1\nu_2 - \nu_4\nu_5}{\nu_7\nu_8}}, \quad (16)$$

$$\lambda = \sqrt{(\nu_1/f)^2 + (\nu_4/f)^2 + \nu_7^2}, \quad (17)$$

$$\overrightarrow{r}_1 = \frac{1}{\lambda f} \begin{bmatrix} \nu_1 & \nu_4 & f\nu_7 \end{bmatrix}^T, \quad (18)$$

$$\overrightarrow{r}_3 = \frac{\overrightarrow{r}_1 \times \begin{bmatrix} \nu_2 & \nu_5 & f\nu_8 \end{bmatrix}^T}{\| \overrightarrow{r}_1 \times \begin{bmatrix} \nu_2 & \nu_5 & f\nu_8 \end{bmatrix}^T \|}, \quad (19)$$

$$\overrightarrow{r}_2 = \overrightarrow{r}_1 \times \overrightarrow{r}_3, \quad (20)$$

$$\overrightarrow{t} = \frac{1}{\lambda f} \begin{bmatrix} \nu_3 & \nu_6 & f\nu_9 \end{bmatrix}^T, \quad (21)$$

$$K_1 = \frac{\nu_{10} + \nu_{11}}{\lambda(r_{31} + r_{32})}, \quad (22)$$

$$K_2 = \frac{\nu_{12}}{\lambda} - K_1 t_3. \quad (23)$$

4.2.2 Non-Linear Optimisation

Following the estimation of the camera parameters assuming common micro-lens focal length f, and zero principal points, Levenberg-Marquardt optimisation minimises the reprojection error from 3D corners to 2D correspondences

Parameter	Type 1	Type 2	Type 3
K_1	-21.625	-17.556	-14.805
K_2	1.2593×10^4	1.1313×10^4	9963.38
f_x [pix]	2.3770×10^4	2.3738×10^4	2.3825×10^4
f_y [pix]	2.3769×10^4	2.3727×10^4	2.3825×10^4
c_x [pix]	1920.22	1923.27	1921.94
c_y [pix]	1079.93	1080.64	1080.71

Table 2: Plenoptic camera calibration: intrinsic parameters.

to retrieve f_x, f_y, c_x, c_y (9) and refine K_1, K_2 (22) - (23), and the extrinsic parameters $\mathbf{R}$, $\overrightarrow{t}$. The reprojection error is minimised independently for each type of micro-lens:

$$f(K_{1,2}^i, f_{x,y}^i, c_{x,y}^i, \mathbf{R}, \overrightarrow{t}) = \sum \| \overrightarrow{u} - \pi(\overrightarrow{X}_w) \|^2, \quad (24)$$

where the superscript i refers to the different type of micro-lenses, and $\overrightarrow{u}, \pi(\overrightarrow{X}_w)$, are the detected corner in the micro-lens image, and the projection of the 3D point, respectively.

After optimisation, different extrinsics are obtained per micro-lens type. As our experiments will demonstrate, the extrinsics are sufficiently close, indicating a single camera body, whose representative extrinsics are calculated by averaging the extrinsics of the individual types [11].

5. Experiments and Results

We used Raytrix™ R8 with a 50 mm lens from Kowa™. The size of the raw plenoptic images is 3840×2160 pixels. The lens/camera system has field-of-view and depth-of-field of approximately $9\,\mathrm{cm} \times 7\,\mathrm{cm}$, and $19.5\,\mathrm{cm}$, respectively. First, we calculate the intrinsics via free-hand calibration of two checker-boards with different checker sizes, demonstrate algorithm convergence, and compare with [4]. Then, knowing the intrinsics, we recover a known checker-board motion, demonstrating the accuracy of the extrinsics.

5.1. Free-Hand Calibration

A checker-board with checker dimensions of $4\,\mathrm{mm} \times 4\,\mathrm{mm}$ was used. Ten images of the checker-board under various poses were captured; typical raw images can be seen in Fig. 1. The corner detection identified 11321 features to be used for calibration across all images. Nineteen micro-images per 3D corner were considered for the 3D-to-2D correspondences. Using the 2D corners, the type and spatial alignment of the micro-lenses were obtained. The calibration parameters calculated by the initial linear solution were refined by non-linear optimisation. The reconstructed poses of the ten images used for calibration are shown in Fig. 7a.

From [15], different intrinsic calibration parameters per micro-lens type are expected. Indeed, Table 2 shows that every micro-lens type has unique K_1 and K_2, (22) - (23). These refer to the internal configuration of the multi-focus
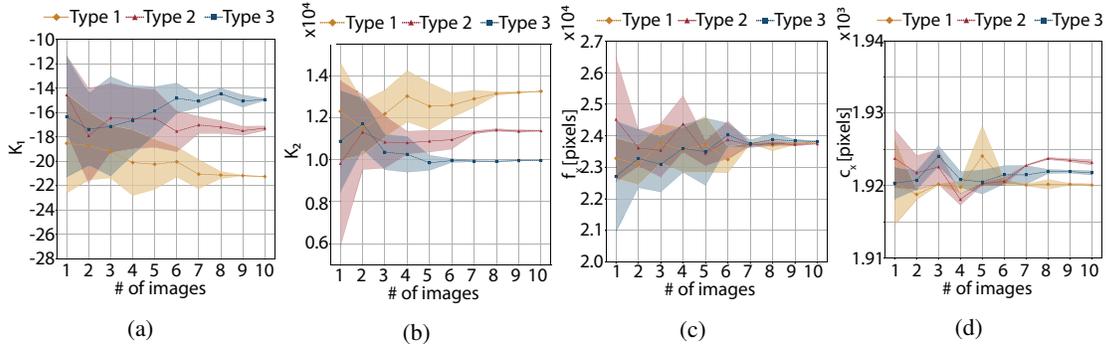
Figure 5: Parameter convergence for sets of different number of images. Behaviour of: (a) K_1, (b) K_2, (c) f_x, and (d) c_x.

plenoptic camera, as they directly depend on L_m, L_c, and F. Practically, this implies that each micro-lens type has a different distance from the image sensor, as also reported in [15]. Therefore, calibration retrieves this camera detail.

As discussed, the extrinsic parameters of each micro-lens type are first calculated independently. The experiments showed that, for all checker-board poses, almost identical extrinsic parameters were estimated for the micro-lens types. The average Euclidean distance for the types' translation is 1.42 ± 0.90 mm, corresponding to 0.3% of the distance to the checker-board. The difference in rotation is $0.14° \pm 0.09°$ solid angle. Achieving similar extrinsics through a per-micro-lens-type optimisation supports the fact that the multi-focus plenoptic camera encompasses different types of micro-lenses within a single camera body.

Figure 6 gives the average reprojection error for each image and micro-lens type. The errors are consistently small, with an average value of approximately 0.83 pixels.

Next, we estimated the precision/variance in the estimated parameters when different number of images are used. Calibration was performed using k-combinations from a set of 10 images, with $k = 1, \cdots, 10$. Even if calibration can be performed with a single image, this experiment provides a rule-of-thumb on the number of images required for precise (in terms of standard deviation) calibration. The obtained average values and standard deviation for K_1, K_2, f_x, c_x are shown in Fig. 5, which illustrates that $6 - 8$ images are required for precise parameter estimation.

Comparison with the state-of-the-art: We compared our method with [4]. The first observation was that [4] was not able to detect a sufficient number of lines on a checker-board with 4 mm checkers. The second observation was that that calibration method was not converging if all micro-lenses were treated as of the same type, i.e. without classification. The proposed micro-lens classifier is necessary regardless of the calibration algorithm used. Thus, for a fair comparison, we used a checker-board with 8 mm checkers and adapted [4] to treat each micro-lens type separately.

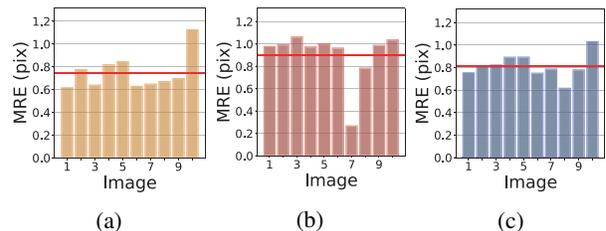Seven checker-board images were used. Table 3 depicts

Figure 6: Mean reprojection errors (MREs) for micro-lenses of (a) Type 1, (b) Type 2, and (c) Type 3. The red line indicates the mean error across all images.

the intrinsic parameters estimated from both algorithms for the 8 mm checker-board. The intrinsics retrieved by our algorithm are consistent with the ones from the free-hand calibration (smaller checkers), shown in Table 2, within approximately 10% difference. This implies that the retrieved parameters accurately describe the 3D-to-2D image-formation process, and, further, that our method is less restrictive in terms of checker size. On the contrary, the principal points obtained by [4] demonstrate a large discrepancy as they are estimated at approximately 350 pixels away from the image centre. Even if this can be explained for consumer systems such as Lytro™ and Illum™ that may contain misalignments in the optical system, it is unrealistic for systems such as Raytrix™ that are manufactured for industrial metrology.

The extrinsic camera parameters were extracted by both algorithms. Table 4 shows the maximum translation/rotation difference among the extrinsics for different types of micro-lenses for all checker-board poses. The translation difference among types i, j, is t_{ij}, while a_{ij} is the rotation difference (solid angle). In our case, the maximum difference in translation is 5.9mm, whereas it is 32.1mm for [4]. When the proposed calibration is used, the rotation difference has a maximum of $0.2°$, which is an order of magnitude less than the $2.5°$ obtained with [4].

5.2. Controlled Motion Experiment

To further evaluate our calibration method, we used the already calculated intrinsics to estimate a known motion

Parameter	Proposed method			Bok et al. [4]		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
K_1	-21.161	-17.239	-12.206	-17.684	-15.755	-16.102
K_2	1.4590×10^4	1.2424×10^4	9418.57	1.2198×10^4	1.1231×10^4	1.2405×10^4
f_x	2.3841×10^4	2.3803×10^4	2.4221×10^4	2.4014×10^4	2.3964×10^4	2.422×10^4
f_y	2.3840×10^4	2.3802×10^4	2.4223×10^4	2.4488×10^4	2.4615×10^4	2.4762×10^4
c_x	1920.33	1920.9	1922.13	2264.52	2306.76	2234.58
c_y	1080.14	1080.5	1080.23	1703.58	2047.48	1621.52

Table 3: Comparison with state-of-the-art: intrinsics.

	Transl. difference (mm)				Rot. difference ($^\circ$)			
	Proposed		[4]		Proposed		[4]	
#	t_{12}	t_{13}	t_{12}	t_{13}	α_{12}	α_{13}	α_{12}	α_{13}
1	2.2	3.8	27.8	29.4	0.04	0.03	2.3	2.5
2	2.5	3.6	27.6	28.9	0.2	0.02	2.3	2.4
3	1.5	5.9	30.7	32.1	0.1	0.07	2.2	2.5
4	1.5	2.8	25.9	27.1	0.2	0.1	2.3	2.2
5	2.2	2.6	27.9	28.6	0.07	0.1	2.4	2.5
6	1.9	3.5	29.4	31.0	0.07	0.08	2.4	2.5
7	1.1	3.0	25.9	27.2	0.1	0.2	2.3	2.2

Table 4: Comparison with state-of-the-art: extrinsics.

(a)   (b)
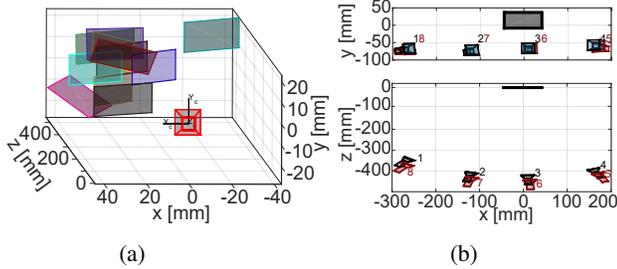
Figure 7: (a) Free-hand motion, and (b) Controlled motion.

	Rot. (x, y, z in degrees)			Transl. (x, y, z in mm)		
T_{12}	1.2	20.5	1.4	1.6	0.4	-4.0
T_{23}	-0.5	19.0	1.4	0.4	0.5	-2.6
T_{34}	-0.9	20.7	1.4	-0.9	0.5	-2.5
T_{45}	0.5	0.07	-0.09	-3.0	-3.0	22.7
T_{56}	0.4	-20.4	-1.4	1.0	-0.6	2.4
T_{67}	0.3	-18.5	-1.3	-0.3	-0.5	2.6
T_{78}	-0.9	-20.1	-1.3	-1.5	-0.4	3.7

Table 5: Extracted relative motion parameters.

ments, the mean error in rotation is 0.65°, 0.6°, and 1.16^o, along the x, y, and z axes, respectively. The mean errors in translation are 1.25mm, 0.86mm, and 2.87mm. Therefore, the motion is reconstructed successfully. Figure 7b illustrates the obtained motion, assuming for illustration purposes that the checker-board was stationary.

6. Conclusions and Discussion

We proposed a method for geometric calibration of multi-focus plenoptic cameras using the raw images. Our method considers the different type of micro-lenses of the camera and obtains their type and spatial arrangement using focus metric. A novel micro-image corner detector enables the use of a 3D-to-2D reprojection-error metric across many raw images. The obtained intrinsics are consistent among different datasets, and are validated through a controlled-camera-motion experiment.

Our approach bridges pinhole camera and multi-focus plenoptic camera calibration. It allows treating the latter as a generalized camera, *i.e.* assigning a ray to each pixel. We believe that this opens up exciting research directions in 3D reconstruction and Structure-from-Motion using multi-focus plenoptic cameras.

of the checker-board. If the relative checker-board/camera motion can be reconstructed using the calculated intrinsics, then the camera has been correctly calibrated. The extrinsic parameters were obtained by rearranging (15) to use the known intrinsics, and then refined via (24).

We performed a known checker-board motion and captured a sequence of 8 images. The motion was controlled via a rotary and a micro-manipulation stage from Thorlabs Gmbh, DE (see Fig. 1). The motion profile was: a) incremental rotation by 20° around the y-axis for frames $2-4$; b) translation by 25.0 mm along the positive z-axis, *i.e.* moving away from the camera, in frame 5; c) -20° incremental rotation along the y-axis for frames $6-8$.

The first frame is the reference frame. The relative frame i to frame j transformations, T_{ij}, are shown in Table 5. It can be seen that rotation around the y-axis is approximately 20° for frames $2-4, 6-8$, while frame 5 exhibits 23 mm translation. Compared to the ground-truth motion incre-

References

[1] E. H. Adelson and J. Bergen. The plenoptic function and the elements of early vision. *Computational models of visual Processing*, pages 3–20, 1991. 1

[2] E. H. Adelson and J. Y. A. Wang. Single Lens Stereo with a Plenoptic Camera. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):99–106, 1992. 1

[3] N. Bedard, T. Shope, A. Hoberman, M. A. Haralam, N. Shaikh, J. Kovacevic, N. Balram, and I. Tosic. Light field otoscope design for 3D in vivo imaging of the middle ear. *Biomedical Optics Express*, 8(1):9529–9534, 2017. 1

[4] Y. Bok, H. G. Jeon, and I. S. Kweon. Geometric calibration of micro-lens-based light field cameras using line features. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 39(2):287–300, 2017. 2, 4, 5, 6, 7, 8

[5] D. G. Dansereau, I. Mahon, O. Pizarro, and S. B. Williams. Plenoptic flow: Closed-form visual odometry for light field cameras. In *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pages 4455–4462, 2011. 1

[6] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 1027–1034, 2013. 2, 4

[7] F. Dong, S.-H. Ieng, X. Savatier, R. Etienne-Cummings, and R. Benosman. Plenoptic cameras in real-time robotics. *Int. J. Robotics Research*, 32(2):206–217, 2013. 1

[8] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Intercommission Conf. on Fast Processing of Photogrammetric Data*, pages 281–305, 1987. 3

[9] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster. Automatic camera and range sensor calibration using a single shot. In *IEEE Int. Conf. Robotics and Automation*, pages 3936–3943, 2012. 3

[10] T. Georgiev and A. Lumsdaine. Depth of Field in Plenoptic Cameras. *Eurographics*, (1):5–8, 2009. 1

[11] C. Gramkow. On averaging rotations. *Int. J. of Computer Vision*, 42(1-2):7–16, 2001. 6

[12] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, volume 15, pages 147–151, 1988. 3

[13] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press,, second edition, 2004. 5

[14] E. Hecht. *Optics*. San Francisco, CA, Addison Wesley, fourth edition, 2002. 2, 5

[15] C. Heinze, S. Spyropoulos, S. Hussmann, and C. Perwass. Automated Robust Metric Calibration Algorithm for Multi-focus Plenoptic Cameras. In *IEEE Int. Conf. Instrumentation and Measurement Technology*, pages 2038–2043, 2015. 2, 4, 6, 7

[16] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. So Kweon. Accurate depth map estimation from a lenslet light field camera. In *IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 1547–1555, 2015. 1

[17] O. Johannsen, C. Heinze, B. Goldluecke, and C. Perwaß. On the calibration of focused plenoptic cameras. In *GCPR Workshop on Imaging New Modalities*, 2013. 2, 4

[18] O. Johannsen, A. Sulc, and B. Goldluecke. On linear structure from motion for light field cameras. In *IEEE Int. Conf. Computer Vision*, pages 720–728, 2015. 1

[19] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. In *IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2014. 1

[20] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *IEEE Int. Conf. Computational Photography*, pages 1–8, 2009. 1

[21] R. Ng. *Digital light field photography*. PhD thesis, Stanford, 2006. 1, 2

[22] S. Pertuz, D. Puig, and M. A. Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, 2013. 4

[23] C. Perwaß and L. Wietzke. Single lens 3D-camera with extended depth-of-field. In *Proceedings of SPIE*, volume 8291, 2012. 1, 2

[24] R. Prevedel, Y.-G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden, et al. Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy. *Nature methods*, 11(7):727–730, 2014. 1

[25] E. Rosten, R. Porter, and T. Drummond. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(1):105–119, 2010. 3

[26] A. Shademan, R. S. Decker, J. Opfermann, S. Leonard, P. C. W. Kim, and A. Krieger. Plenoptic cameras in surgical robotics: Calibration, registration, and evaluation. In *IEEE Int. Conf. Robotics and Automation*, pages 708–714, 2016. 1

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

[28] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. *IEEE Int. Conf. Computer Vision*, 2:673–680, 2013. 1

[29] T.-C. Wang, A. A. Efros, and R. Ramamoorthi. Depth estimation with occlusion modeling using light-field cameras. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 38(11):2170–2181, 2016. 1

[30] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 36(3):606–619, 2014. 1

[31] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000. 4