Blur Aware Calibration of Multi-Focus Plenoptic Camera

Mathieu Labussière¹, Céline Teulière¹, Frédéric Bernardin², Omar Ait-Aider¹
¹ Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal, F-63000 Clermont-Ferrand, France
² Cerema, Équipe-projet STI, 10 rue Bernard Palissy, F-63017 Clermont-Ferrand, France

mathieu.labu@gmail.com, firstname.name@{uca,cerema}.fr

Abstract

This paper presents a novel calibration algorithm for Multi-Focus Plenoptic Cameras (MFPCs) using raw images only. The design of such cameras is usually complex and relies on precise placement of optic elements. Several calibration procedures have been proposed to retrieve the camera parameters but relying on simplified models, reconstructed images to extract features, or multiple calibrations when several types of micro-lens are used. Considering blur information, we propose a new Blur Aware Plenoptic (BAP) feature. It is first exploited in a pre-calibration step that retrieves initial camera parameters, and secondly to express a new cost function for our single optimization process. The effectiveness of our calibration method is validated by quantitative and qualitative experiments.

1. Introduction

The purpose of an imaging system is to map incoming light rays from the scene onto pixels of photo-sensitive detector. The radiance of a light ray is given by the plenoptic function \( L(x, \theta, \lambda, \tau) \), introduced by Adelson et al. [1], where \( x \in \mathbb{R}^3 \) is the spatial position of observation, \( \theta \in \mathbb{R}^2 \) is the angular direction of observation, \( \lambda \) is the wavelength of the light and \( \tau \) is the time. Conventional cameras capture only one point of view. A plenoptic camera is a device that allows to retrieve spatial as well as angular information.

From Lumigraph [17] to commercial plenoptic cameras [19, 23], several designs have been proposed. This paper focuses on plenoptic cameras based on a Micro-Lenses Array (MLA) placed between the main lens and the photo-sensitive sensor (see Fig. 2). The specific design of such a camera allows to multiplex both types of information onto the sensor in the form of a Micro-Images Array (MIA) (see Fig. 1 (b)), but implies a trade-off between the angular and spatial resolutions [10, 16, 8]. It is balanced according to the MLA position with respect to the main lens focal plane (i.e., focused [23, 9] and unfocused [19] configurations).

Figure 1: The Raytrix R12 multi-focus plenoptic camera used in our experimental setup (a), along with a raw image of a checkerboard calibration target (b). The image is composed of several micro-images with different blurred levels and arranged in an hexagonal grid. In each micro-image, our new Blur Aware Plenoptic (BAP) feature is illustrated by its center and its blur radius (c).

1.1. Related Work

Unfocused plenoptic camera calibration. In this configuration, light rays are focused by the MLA on the sensor plane. The calibration of unfocused plenoptic camera [19] has been widely studied in the literature [6, 2, 25, 24, 11, 36]. Most approaches rely on a thin-lens model for the main lens and an array of pinholes for the micro-lenses. Most of them require reconstructed images to extract features, and limit their model to the unfocused configuration, i.e., set-
ting the micro-lens focal length at the distance MLA-sensor. Therefore those models cannot be directly extended to the focused or multi-focus plenoptic camera.

**Focused plenoptic camera calibration.** With the arrival of commercial focused plenoptic cameras [18, 23], new calibration methods have been proposed. Based on [14], Heinze et al. [13] have developed a new projection model and a metric calibration procedure which is incorporated in the RxLive software of Raytrix GmbH. Other calibration methods and models have been proposed, either to overcome the fragility of the initialization [26], or to model more finely the camera parameters using depth information [33, 32]. O'Brien et al. [22] introduced a new 3D feature called plenoptic disc and defined by its center and its radius. Nevertheless, all previous methods rely on reconstructed images meaning that they introduce error in the reconstruction step as well as in the calibration process.

To overcome this problem, several calibration methods [35, 34, 29, 20, 3, 21, 31] have been proposed using only raw plenoptic images. In particular, features extraction in raw micro-images has been studied in [3, 21, 20] achieving improved performance through automation and accurate identification of feature correspondences. However, most of the methods rely on simplified models for optic elements: the MLA is modeled as a pinholes array making it impossible to retrieve the focal lengths, or the MLA misalignment is not considered. Some do not consider distortions [35, 21, 31] or restrict themselves to the focused case [35, 34, 20].

Finally, few have considered the multi-focus case [13, 3, 21, 31] but dealt with it in separate processes, leading to different intrinsic and extrinsic parameters according to the type of micro-lenses.

**1.2. Contributions**

This paper focuses on the calibration of micro-lenses-based Multi-Focus Plenoptic Camera (MFPC). To the best of our knowledge, this is the first method proposing a single optimization process that retrieves intrinsic and extrinsic parameters of a MFPC directly from raw plenoptic images. The main contributions are the following:

- We present a new Blur Aware Plenoptic (BAP) feature defined in raw image space that enables us to handle the multi-focus case.
- We introduce a new pre-calibration step using BAP features from white images to provide a robust initial estimate of internal parameters.
- We propose a new reprojection error function exploiting BAP features to refine a more complete model, including in particular the multiple micro-lenses focal lengths. Our checkerboard-based calibration is conducted in a single optimization process.

A visual overview of our method is given in Fig. 3. The remainder of this paper is organized as follows: first, the camera model and BAP feature are presented in Section 2. Then, the feature detection is detailed in Section 4 and the calibration process in Section 5. Finally, our results are presented and discussed in Section 6. The notations used in this paper are shown in Fig. 2. Pixel counterparts of metric values are denoted in lower-case Greek letters.

2. Camera model and BAP feature

2.1. Multi-focus Plenoptic Camera

We consider multi-focus plenoptic cameras as described in [9, 23]. The main lens, modeled as a thin-lens, maps object point to virtual point behind (resp., in front of) the image sensor in Galilean (resp., Keplerian) configuration. Therefore, the MLA consists of $I$ different lens types with focal lengths $f(i), i \in \{1, \ldots, I\}$ which are focused on $I$ different planes behind the image sensor. This multi-focus setup corresponds to the Raytrix camera system described in [23] when $I = 3$. The micro-lenses are modeled as thin-lenses allowing to take into account blur in the micro-image. Our model takes into account the MLA misalignment, freeing all six degrees of freedom.

The tilt of the main lens is included in the distortion model and we make the hypothesis that the main lens plane $\Pi_1$ is parallel to the sensor plane $\Pi_s$. Furthermore, we choose the main lens frame as our camera reference frame, with $O$ being the origin, the $z$-axis coinciding with the optical axis and pointing outside the camera, and the $y$-axis pointing downwards. Only distortions of the main lens are considered. We use the model of Brown-Conrady [4, 5] with three coefficients for the radial component and two for the tangential.
Furthermore, we take into account the deviation of the image center and the optical center for each micro-lens because it tends to cause inaccuracy in decoded light field. Therefore, the principal point \( c_{0,0}^{k,l} \) of the micro-lens indexed by \( (k,l) \) is given by

\[
e_{0,0}^{k,l} = 
\begin{bmatrix}
  u_{0}^{k,l} \\
  v_{0}^{k,l}
\end{bmatrix} = \frac{d}{D + d} \left( \begin{bmatrix}
  u_{0} \\
  v_{0}
\end{bmatrix} - e_{k,l} \right) + c_{k,l},
\]

where \( c_{k,l} \) is the center in pixels of the micro-image \((k,l)\), \([u_{0} \ v_{0}]^\top\) is the main lens principal point, \(d\) is the distance MLA-sensor and \(D\) is the distance main lens-MLA, as illustrated in Fig. 2.

Finally, each micro-lens produces a micro-image onto the sensor. The set of these micro-images has the same structural organization as the MLA, i.e. in our case an hexagonal grid, alternating between each type of micro-lens. The data can therefore be interpreted as an array of micro-images, called by analogy the MIA. The MIA coordinates are expressed in image space. Let \( \Delta e \) be the pixel distance between two arbitrary consecutive micro-images centers \( c_{k,l} \). With \( s \) the metric size of a pixel, let \( \Delta C = s \cdot \Delta e \) be its metric value, and \( \Delta C \) be the metric distance between the two corresponding micro-lenses centers \( C_{k,l} \). From similar triangles, the ratio between them is given by

\[
\frac{\Delta C}{\Delta e} = \frac{D}{d + D} \implies \Delta C = \Delta e \cdot \frac{D}{d + D}.
\]

We make the hypothesis that \( \Delta C \) is equal to the micro-lens diameter. Since \( d \ll D \), we can make the following approximation:

\[
\frac{D}{D + d} = \lambda \approx 1 \implies \Delta C = \Delta e \cdot \frac{D}{D + d} \approx \Delta e.
\]

This approximation will be validated in the experiments.

### 2.2. BAP feature and projection model

Using a camera with a circular aperture, the blurred image of a point on the image detector is circular in shape and is called the blur circle. From similar triangles and from the thin-lens equation, the signed blur radius of a point in an image can be expressed as

\[
\rho = \frac{1}{s} \cdot \frac{A}{2} \left( \frac{1}{f} - \frac{1}{a} - \frac{1}{d} \right),
\]

with \( s \) the size of a pixel, \( d \) the distance between the considered lens and the sensor, \( A \) the aperture of this lens, \( f \) its focal length, and \( a \) the distance of the object from the lens.

This radius appears at different levels in the camera projection: in the blur introduced by the thin-lens model of the micro-lenses and during the formation of the micro-image while taking a white image. To leverage blur information, we introduce a new BlurAware Plenoptic (BAP) feature characterized by its center and its radius, i.e., \( p = (u, v, \rho) \).

Therefore, our complete plenoptic camera model allows us to link a scene point \( p_w \) to our new BAP feature \( p \) through each micro-lens \((k,l)\)

\[
\begin{bmatrix}
  u \\
  v \\
  \rho
\end{bmatrix} \propto P(i,k,l) \cdot T_p(k,l) \cdot \varphi(K(F) \cdot T_e \cdot p_w),
\]

where \( P(i,k,l) \) is the blur aware plenoptic projection matrix through the micro-lens \((k,l)\) of type \( i \), and computed as

\[
P(i,k,l) = P(k,l) \cdot K(f^{(i)})
\]

\[
\begin{bmatrix}
  d/s & 0 & u_{0}^{k,l} \\
  0 & d/s & v_{0}^{k,l} \\
  0 & 0 & s \frac{\Delta C}{d}
\end{bmatrix}
\begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  0 & 0 & 1 \\
  0 & 1/f^{(i)} & 0
\end{bmatrix}.
\]

\( P(k,l) \) is a matrix that projects the 3D point onto the sensor. \( K(f) \) is the thin-lens projection matrix for the given focal length. \( T_p \) is the pose of the main lens with respect to the world frame and \( T_p(k,l) \) is the pose of the micro-lens \((k,l)\) expressed in the camera frame. The function \( \varphi(\cdot) \) models the lateral distortion.

Finally, the projection model defined in Eq. (5) consists of a set \( \Xi \) of \((16 + I)\) intrinsic parameters to be optimized, including the main lens focal length \( F \) and its 5 lateral distortion parameters, the sensor translation, encoded in \((u_{0}, v_{0})\) and \(d\), the MLA misalignment, i.e., 3 rotations \((\theta_x, \theta_y, \theta_z)\) and 3 translations \((t_x, t_y, D)\), the micro-lens inter-distance \( \Delta C \), and the \( I \) micro-lens focal lengths \( f^{(i)} \).

### 3. Pre-calibration using raw white images

Drawing inspiration from depth from defocus theory [27], we leverage blur information to estimate parameters (e.g., here our blur radius) by varying some other parameters (e.g., the focal length, the aperture, etc.) in combination with known (i.e., fixed or measured) parameters. For instance, when taking a white picture with a controlled aperture, each type of micro-lens produces a micro-image (MI) with a specific size and intensity, providing a way to distinguish between them. In the following all distances are given with reference to the MLA plane. Distances are signed according to the following convention: \( f \) is positive when the lens is convergent; distances are positive when the point is real, and negative when virtual.

#### 3.1. Micro-image radius derivation

Taking a white image is equivalent for the micro-lenses to image a white uniform object of diameter \( A \) at a distance
We relate the micro-image (MI) radius to the plenoptic camera parameters. From optics geometry, the image of this object, i.e. the resulting MI, is equivalent to the image of an imaginary point constructed as the vertex of the cone passing through the main lens and the considered micro-lens (noted V in Fig. 4). Let \( a' \) be the distance of this point from the MLA plane, given from similar triangles and Eq. (2) by

\[
\frac{a'}{A} = \frac{\Delta C}{A - \Delta C} = -D \left( A \left( \frac{d + D}{\Delta c D} \right) - 1 \right)^{-1},
\]

with \( A \) the main lens aperture. Note the minus sign is due to the fact that the imaginary point is always formed behind the MLA plane at a distance \( a' \), and thus considered as a virtual object for the micro-lenses. Conceptually, the MI formed can be seen as the blur circle of this imaginary point. Therefore, using Eq. (4), the metric MI radius \( R \) is given by

\[
R = \frac{\Delta C}{2} d \left( \frac{1}{f} - \frac{1}{a'} - \frac{1}{d} \right).
\]

From the above equation, we see that the radius depends linearly on the aperture of the main lens. However, the main lens aperture cannot be computed directly whereas we have access to the \( f \)-number value. The \( f \)-number of an optical system is the ratio of the system’s focal length \( F \) to the diameter of the entrance pupil, \( A \), given by \( N = F/A \).

Finally, we can express the MI radius for each micro-lens focal length type \( i \) as

\[
R(N^{-1}) = m \cdot N^{-1} + q_i \tag{9}
\]

with

\[
m = \frac{dF}{2D} \quad \text{and} \quad q_i = \frac{1}{f(i)} \left( \frac{\Delta c D}{d + D} \right) \cdot \frac{d}{2} - \frac{\Delta c}{2} \tag{10}
\]

Let \( q'_i \) be the value obtained by \( q'_i = q_i + \frac{\Delta c}{2} \).

### 3.2. Internal parameters estimation

The internal parameters \( \Omega = \{m, q'_1, \ldots, q'_I\} \) are used to compute the radius part of the BAP feature and to initialize the parameters of the calibration process. Given several raw white images taken at different apertures, we estimate the coefficients of Eq. (9) for each type of micro-image. The standard full-stop \( f \)-number conventionally indicated on the lens differs from the real \( f \)-number calculated with the aperture value \( AN \) as \( N = \sqrt{2N} \).

From raw white images, we are able to measure each micro-image (MI) radius \( \rho = |R|/s \) in pixels for each distinct focal length \( f(i) \) at a given aperture. Due to vignetting effect, the estimation is only conducted on center micro-images which are less sensitive to this effect. Our method is based on image moments fitting. It is robust to noise, works under asymmetrical distribution and is easy to use, but needs a parameter \( \alpha \) to convert the standard deviation \( \sigma \) into a pixel radius \( \rho = \alpha \cdot \sigma \). We use the second order central moments of the micro-image to construct a covariance matrix. Finally, we choose \( \sigma \) as the square root of the greater eigenvalue of the covariance matrix. The parameter \( \alpha \) is determined such that at least 98% of the distribution is taken into account. According to the standard normal distribution \( Z \)-score table, \( \alpha \) is picked up in \([2.33, 2.37]\). In our experiments, we set \( \alpha = 2.357 \).

![Figure 4: Formation of a micro-image with its radius \( R \) through a micro-lens while taking a white image at an aperture \( A \). The point \( V \) is the vertex of the cone passing by the main lens and the considered micro-lens. \( V' \) is the image of \( V \) by the micro-lens and \( R \) is the radius of its blur circle.](image-url)
4. BAP feature detection in raw images

At this point, the internal parameters $\Omega$, used to express our blur radius $\rho$, are available. The process (illustrated in Fig. 3) is divided into three phases: 1) using a white raw image, the MIA is calibrated and micro-image centers are extracted; 2) checkerboard images are processed to extract corners at position $(u, v)$; and 3) with the internal parameters and the associated virtual depth estimate for each corner, the corresponding BAP feature is computed.

4.1. Blur radius derivation through micro-lens

To respect the f-number matching principle [23], we configure the main lens f-number such that the micro-images fully tile the sensor without overlap. In this configuration the working f-number of the main imaging system and the micro-lens imaging system should match. The following relation is then verified for at least one type of micro-lens:

$$\frac{\Delta \mathbf{a}}{a} = \frac{1}{d} \cdot \left( \frac{\Delta \mathbf{c} D}{d + D} \right) = \frac{F}{ND}. \quad (11)$$

We consider the general case of measuring an object $p$ at a distance $a_l$ from the main lens. First, $p$ is projected through the main lens according to the thin lens equation, $1/F = 1/a_l + 1/b_l$, resulting in a point $p'$ at a distance $b_l$ behind the main lens, i.e. at a distance $a = D - b_l$ from the MLA. The metric radius of the blur circle $r$ formed on the sensor for a given point $p'$ at distance $a$ through a micro-lens of type $i$ is expressed as

$$r = \left( \frac{\Delta \mathbf{c} D}{d + D} \right) \cdot \frac{d}{2} \cdot \left( \frac{1}{f(i)} - \frac{1}{a} - \frac{1}{d} \right) = \left( \frac{\Delta \mathbf{c} D}{d + D} \right) \cdot \frac{d}{2} \cdot \left( \frac{1}{d} - \frac{1}{a} \right) = \frac{\Delta \mathbf{c} D}{2a} \cdot \frac{d}{2} = \frac{\Delta \mathbf{c}}{2} \cdot \frac{d}{a}.$$

$$r = \left( \frac{\Delta \mathbf{c} D}{d + D} \right) \cdot \frac{d}{2} \cdot \left( q_i - \frac{\Delta \mathbf{c}}{2} \right). \quad (12)$$

In practice, we do not have access to the value of $\Delta \mathbf{C}$ but we can use the approximation from Eq. (3). Moreover, $a$ and $d$ cannot be measured in the raw image space, but the virtual depth can. Virtual depth refers to relative depth value obtained from disparity. It is defined as the ratio between the object distance $a$ and the sensor distance $d$:

$$\nu = \frac{a}{d}. \quad (13)$$

The sign convention is reversed for virtual depth computation, i.e. distances are negative in front of the MLA plane. If we re-inject the latter in Eq. (12), taking caution of the sign, we can derive the radius of the blur circle of a point $p'$ at a distance $a$ from the MLA by

$$r = \frac{\lambda \Delta \mathbf{c}}{2} \cdot \nu^{-1} + \left( q_i - \frac{\Delta \mathbf{c}}{2} \right). \quad (14)$$

This equation allows to express the pixel radius of the blur circle $\rho = r/s$ associated to each point having a virtual depth without explicitly evaluating $A, D, d, F$ and $f(i)$.

4.2. Features extraction

First, the Micro-Images Array has to be calibrated. We compute the micro-image centers observations $\{e_{k,l}\}$ by intensity centroid method with sub-pixel accuracy [30, 20, 28]. The distance between two micro-image centers $\delta a$ is then computed as the optimized edge-length of a fitted 2D grid mesh with a non-linear optimization. The pixel translation offset in the image coordinates, $(\tau_x, \tau_y)$, and the rotation around the $(-z)$-axis, $\vartheta_z$, are also determined during the optimization process. From the computed distance and the given size of a pixel, the parameter $\lambda \Delta \mathbf{c}$ is computed.

Secondly, as we based our method on checkerboard calibration pattern, we detect corners in raw images using the detector introduced by Noury et al. [20]. The raw images are desigotted by dividing them by a white raw image taken at the same aperture.

With a plenoptic camera, contrarily to a classical camera, a point is projected into more than one observation onto the sensor. Due to the nature of the spatial distribution of the data, we used the DBSCAN algorithm [7] to identify the
clusters. We then associate each point with its cluster of observations.

Once each cluster is identified, we can compute the virtual depth \( \nu \) from disparity. Given the distance \( \Delta C_{1-2} \) between the centers of the micro-lenses \( C_1 \) and \( C_2 \), i.e., the baseline, and the Euclidean distance \( \Delta i = |i_1 - i_2| \) between images of the same point in corresponding micro-images, the virtual depth \( \nu \) can be calculated with the intercept theorem:

\[
\nu = \frac{\Delta C_{1-2}}{\Delta C_{1-2} - \Delta i} = \frac{\eta \Delta C}{\eta \Delta C - \Delta i} = \frac{\eta \lambda \Delta e}{\eta \lambda \Delta e - \Delta i}, \tag{15}
\]

where \( \Delta C = \lambda \Delta e \) is the distance between two consecutive micro-lenses and \( \eta \geq 1.0 \). Due to noise in corner detection, we use a median estimator to compute the virtual depth of the cluster taking into account all combinations of point pairs in the disparity estimation.

Finally, from internal parameters \( \Omega \) and with the available virtual depth \( \nu \) we can compute the BAP features using Eq. (14). In each frame \( n \), for each micro-image \( (k, l) \) of type \( i \) containing a corner at position \((u, v)\) in the image, the feature \( p_{k,l}^i \) is given by

\[
p_{k,l}^i = (u, v, \rho), \quad \text{with} \quad \rho = r/s. \tag{16}
\]

In the end, our observations are composed of a set of micro-image centers \( \{c_{k,l}\} \) and a set of BAP features \( \{p_{k,l}^i\} \) allowing us to introduce two reprojection error functions corresponding to each set of features.

5. Calibration process

To retrieve the parameters of our camera model, we use a calibration process based on non-linear minimization of a reprojection error. The calibration process is divided into three phases: 1) the intrinsics are initialized using the internal parameters \( \Omega \); 2) the initial extrinsics are estimated from the raw checkerboard images; and 3) the parameters are refined with a non-linear optimization leveraging our new BAP features.

5.1. Parameters initialization

Optimization processes are sensitive to initial parameters. To avoid falling into local minima during the optimization process, the parameters have to be carefully initialized not too far from the solution.

First, the camera is initialized in Keplerian or Galilean configuration. First, given the camera configuration, the internal parameters \( \Omega \), the focus distance \( h \), and from the Eq. (17) of [23], the following parameters are set as

\[
d \leftarrow \frac{2mH}{F + 4m} \quad \text{and} \quad D \leftarrow H - 2d, \tag{17}
\]

where \( H \) is given by

\[
H = \left| \frac{h}{2} \left( 1 - \sqrt{1 + \frac{F^2}{h^2}} \right) \right|, \tag{18}
\]

with positive (resp., negative) sign in Galilean (resp., Keplerian) configuration.

The focal length, and the pixel size \( s \) are set according to the manufacturer value. All distortion coefficients are set to zero. The principal point is set as the center of the image. The sensor plane is thus set parallel to the main lens plane, with no rotation, at a distance \(- (D + d)\).

Seemingly, the MLA plane is set parallel to the main lens plane at a distance \(- D \). From the pre-computed MIA parameters the translation takes into account the offsets \((- s\tau_x, - s\tau_y)\) and the rotation around the \( z \)-axis is initialized with \(- \theta_z\). The micro-lenses inter-distance \( \Delta C \) is set according to Eq. (2). Finally, from internal parameters \( \Omega \), the focal lengths are computed as follows:

\[
f^{(i)} \leftarrow \frac{d}{2\cdot q_i} \cdot \Delta C. \tag{19}
\]

5.2. Initial poses estimation

The camera poses \( \{T_c^n\} \), i.e., the extrinsic parameters are initialized using the same method as in [20]. For each cluster of observation, the barycenter is computed. Those barycenters can been seen as the projections of the checkerboard corners through the main lens using a standard pinhole model. For each frame, the pose is then estimated using the Perspective-n-Point (PnP) algorithm [15].

5.3. Non-linear optimization

The proposed model allows us to optimize all the parameters into one single optimization process. We propose a new cost function \( \Theta \) taking into account the blur information using our new BAP feature. The cost is composed of two main terms both expressing errors in the image space: 1) the blur aware plenoptic reprojection error, where, for each frame \( n \), each checkerboard corner \( p_{w}^i \) is reprojected into the image space through each micro-lens \((k, l)\) of type \( i \) according to the projection model of Eq. (5) and compared to its observations \( p_{k,l}^i \); and 2) the micro-lens center reprojection error, where, the main lens center \( O \) is reprojected according to a pinhole model in the image space through each micro-lens \((k, l)\) and compared to the detected micro-image center \( c_{k,l} \).

Let \( S = \{\Xi, \{T_c^n\}\} \) be the set of intrinsic \( \Xi \) and extrinsic \( \{T_c^n\} \) parameters to be optimized. The cost function \( \Theta(S) \) is expressed as

\[
\sum ||p_{k,l}^i - \Pi_{k,l}(p_{w}^i)||^2 + \sum ||c_{k,l} - \Pi_{k,l}(O)||^2. \tag{20}
\]

The optimization is conducted using the Levenberg-Marquardt algorithm.
Datasets. We calibrate our camera for three different focus distance configurations and build three corresponding datasets: R12-A for \( h = 450 \) mm, R12-B for \( h = 1000 \) mm, and R12-C for \( h = \infty \). Each dataset is composed of:

- white raw plenoptic images acquired at different apertures \( \{N \in \{4, 5, 6, 8, 11, 31, 16\}\} \) with augmented gain to ease circle detection in the pre-calibration step,
- free-hand calibration targets acquired at various poses (in distance and orientation), separated into two subsets, one for the calibration process and the other for the qualitative evaluation,
- a white raw plenoptic image acquired in the same luminosity condition and with the same aperture as in the calibration targets acquisition,
- and calibration targets acquired with a controlled translation motion for quantitative evaluation, along with the depth maps computed by the Raytrix software (RxLive v4.0.50.2).

We use a \( 9 \times 5 \) of 10mm side checkerboard for R12-A, a \( 8 \times 5 \) of 20mm for R12-B, and a \( 7 \times 5 \) of 30mm for R12-C. Datasets and our source code are publicly available\(^1\).

### 6. Experiments and Results

We evaluate our calibration model quantitatively in a controlled environment and qualitatively when ground truth is not available.

#### 6.1. Experimental setup

For all experiments we used a Raytrix R12 color 3D-light-field-camera, with a MLA of F/2.4 aperture. The mounted lens is a Nikon AF Nikkor F/1.8D with a 50mm focal length. The MLA organization is hexagonal, and composed of \( 176 \times 152 \) (width \( \times \) height) microlenses with \( I = 3 \) different types. The sensor is a Basler beA4000-62KC with a pixel size of \( s = 0.0055 \)mm. The raw image resolution is \( 4080 \times 3068 \).

#### 6.2. Results

**Internal parameters.** Note that the pixel MI radius is given by \( \rho = |R|/s \), and \( R \) is either positive if formed after the rays inversion (as in Fig. 4), or negative if before. With our camera \( f^{(i)} > d [12] \), so \( R < 0 \) implying that \( m \) and \( c_i \) are also negative. In practice, it means we use the value \(-\rho\) in the estimation process.

---

\(^1\)https://github.com/comsee-research
In our experiments, we set $\lambda = 0.9931$. We verified that the error introduced by this approximation is less than the metric size of a pixel (i.e., less than 1% of its optimized value). Fig. 5 shows the estimated lines (see Eq. (9)), with a cropped white image taken at $N = 8$, where each type of micro-lens is identified with its radius. An histogram of the radii distribution is also given for each type. From the radii measurements, the computed internal parameters are estimated and given for each dataset in Tab. 2.

As expected, the internal parameters $\Delta c$ and $m$ are different for each dataset, as $D$ and $\delta c$ vary with the focus distance $h$, whereas the $q'_i$ values are close for each dataset. Using intrinsics, we compare the expected coefficient $m^*$ (from the closed form in Eq. (10) with the optimized parameters) with its calculated $m$ value. The mean error over all datasets is $\bar{\varepsilon}_m = 1.36\%$ which is smaller than a pixel.

**Free-hand camera calibration.** The initial parameters for each dataset are given in Tab. 1 along with the optimized parameters obtained from our calibration and from the method in [20]. Some parameters are omitted for compactness (i.e., distortions coefficient and MLA rotations which are negligible). The main lens focal lengths obtained with the proprietary RxLive software are: $F_{h=450} = 47.709$ mm, $F_{h=1000} = 50.8942$ mm, and $F_{h=\infty} = 51.5635$ mm.

With our method and [20], the optimized parameters are close to their initial value, showing that our method provides a good initialization for our optimization process. The $F$, $d$ and $\Delta C$ are consistent across the datasets with our method. In contrast, the $F$ and $d$ obtained with [20] show a larger discrepancy. This is also the case for the focal lengths obtained by RxLive.

**Poses evaluation.** Tab. 3 presents the relative translations and their errors with respect to the ground truth for the controlled environment experiment. Even if our relative errors are similar with [20], we are able to retrieve more parameters. With our method, absolute errors are of the order of the mm (R12-A: $0.37 \pm 0.15$ mm, R12-B: $1.66 \pm 0.58$ mm and R12-C: $1.38 \pm 0.85$ mm) showing that the retrieved scale is coherent. Averaging over all the datasets, our method presents the smallest relative error, with low discrepancy between datasets, outperforming the estimations of the RxLive software.

<table>
<thead>
<tr>
<th>R12-A</th>
<th>R12-B</th>
<th>R12-C</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon_x$</td>
<td>$\sigma_x$</td>
<td>$\varepsilon_x$</td>
<td>$\sigma_x$</td>
</tr>
<tr>
<td>Ours</td>
<td>$3.73$</td>
<td>$1.48$</td>
<td>$3.32$</td>
</tr>
<tr>
<td>[20]</td>
<td>$6.83$</td>
<td>$1.17$</td>
<td>$1.16$</td>
</tr>
<tr>
<td>RxLive</td>
<td>$4.63$</td>
<td>$2.51$</td>
<td>$4.26$</td>
</tr>
</tbody>
</table>

**Reprojection error evaluation.** For each evaluation dataset, the total squared pixel error is reported with its computed RMSE in Tab. 4. The error is less than 1px per feature for each dataset demonstrating that the computed intrinsics are valid and can be generalized to images different from the calibration set.

<table>
<thead>
<tr>
<th>R12-A (#11424)</th>
<th>R12-B (#3200)</th>
<th>R12-C (#9568)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total RMSE</td>
<td>Total RMSE</td>
<td>Total RMSE</td>
</tr>
<tr>
<td>$\varepsilon_{\text{all}}$</td>
<td>$8972.91$</td>
<td>$0.886$</td>
</tr>
<tr>
<td>$\varepsilon_{u,v}$</td>
<td>$8908.65$</td>
<td>$0.883$</td>
</tr>
<tr>
<td>$\varepsilon_{\rho}$</td>
<td>$64.257$</td>
<td>$0.075$</td>
</tr>
</tbody>
</table>

**Table 2:** Internal parameters (in µm) computed during the pre-calibration step for each dataset. The expected value for $m$ is given by $m^*$ and the relative error $\varepsilon_m = (m^* - m)/m^*$ is computed.

**Table 3:** Relative translation error along the $z$-axis with respect to the ground truth displacement. For each dataset, the mean error $\bar{\varepsilon}_z$ and its standard deviation $\sigma_z$ are given. Results are given for our method and compared with [20] and the proprietary software RxLive.

**Table 4:** Reprojection error for each evaluation dataset with their number of observations. For each component of the feature, the total squared pixel error is reported with its computed RMSE.

**7. Conclusion.**

To calibrate the Multi-Focus Plenoptic Camera, state-of-the-art methods rely on simplifying hypotheses, on reconstructed data or require separate calibration processes to take into account the multi-focal aspect. This paper introduces a new pre-calibration step which allows us to compute our new BAP feature directly in the raw image space. We then derive a new projection model and a new reprojection error using this feature. We propose a single calibration process based on non-linear optimization that enables us to retrieve camera parameters, in particular the micro-lenses focal lengths. Our calibration method is validated by qualitative experiments and quantitative evaluations. In the future, we plan to exploit this new feature to improve metric depth estimation.

**Acknowledgments.** This work has been supported by the AURA Region and the European Union (FEDER) through the MMII project of CPER 2015-2020 MMaSyF challenge. We thank Charles-Antoine Noury and Adrien Coly for their insightful discussions and their help during the acquisitions.
References

[18] Andrew Lumsdaine and Todor Georgiev. The focused plenoptic camera. In IEEE International Conference on Computational Photography (ICCP), pages 1–8, apr 2009. 2


