



HAL
open science

Light-field camera calibration from raw images

Charles-Antoine Noury, Céline Teulière, Michel Dhome

► **To cite this version:**

Charles-Antoine Noury, Céline Teulière, Michel Dhome. Light-field camera calibration from raw images. International Conference on Digital Image Computing: Techniques and Applications (DICTA'17), Nov 2017, Sydney, Australia. pp.1-8, 10.1109/DICTA.2017.8227459 . hal-01657735

HAL Id: hal-01657735

<https://uca.hal.science/hal-01657735v1>

Submitted on 29 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Light-field camera calibration from raw images

Charles-Antoine Noury

Université Clermont Auvergne,
CNRS, Institut Pascal

F-63000 Clermont-Ferrand, France
Email: charles_antoine.noury@uca.fr

Céline Teulière

Université Clermont Auvergne,
CNRS, Institut Pascal

F-63000 Clermont-Ferrand, France
Email: celine.teuliere@uca.fr

Michel Dhome

Université Clermont Auvergne,
CNRS, Institut Pascal

F-63000 Clermont-Ferrand, France
Email: michel.dhome@uca.fr

Abstract—This paper presents a new calibration method for lenslet-based plenoptic cameras. While most existing approaches require the computation of sub-aperture images or depth maps which quality depends on some calibration parameters, the proposed process uses the raw image directly. We detect micro-images containing checkerboard corners and use a pattern registration method to estimate their positions with subpixel accuracy.

We present a more complete geometrical model than previous work composed of 16 intrinsic parameters. This model relates 3D points to their corresponding image projections.

We introduce a new cost function based on reprojection errors of both checkerboard corners and micro-lenses centers in the raw image space. After the initialization process, all intrinsic and extrinsic parameters are refined with a non-linear optimization. The proposed method is validated in simulation as well as on real images.

I. INTRODUCTION

A plenoptic camera (or light-field camera) is a passive sensor that captures five dimensions of a light field: intensity, position and orientation of incoming light rays onto the sensor. Unlike regular cameras, each pixel of the sensor receives rays corresponding to a particular incidence. This property allows to acquire richer information about the scene with various applications such as depth estimation or refocusing.

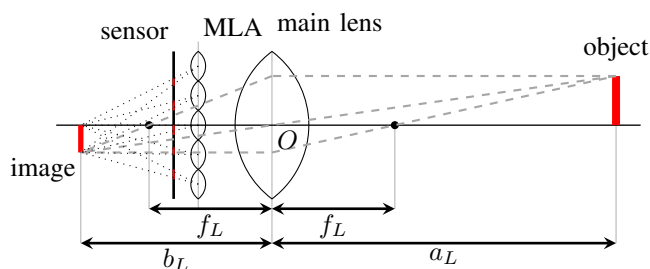


Figure 1. Galilean plenoptic camera model.

Light-field cameras are based on the *integral photography* concept introduced in 1908 by Lippmann [10]. He created a system composed of closely spaced lenses arrays to record multiple images of a scene. Since then, various versions were designed like camera arrays or hand-held models [7], [1].

This paper focuses on hand-held plenoptic cameras composed of a micro-lenses array (MLA) placed between the principal lens and the sensor (Fig. 1). Light rays going through the main lens are then projected through micro-lenses.

The first consumer hand-held plenoptic camera was conceived in 2005 by Ng [13], [12] and released by Lytro ¹. In this model called unfocused plenoptic camera, the MLA is placed in the focal plane of the main lens. The focal length of micro-lenses is equal to the distance between the MLA and the sensor.

In 2009, Georgiev and Lumsdaine [11] presented a modified version named focused plenoptic camera where the MLA is focused on the image formed by the main lens. The model is called Keplerian when the main lens focal plane is located in front of the MLA, and Galilean when it is placed behind (Fig. 1).



Figure 2. Zoom on a plenoptic raw image.

A plenoptic raw image is composed of micro-images grid (Fig. 2). Each micro-image represents rays captured by its corresponding micro-lens. Thanks to the information contained in a single raw image, plenoptic cameras can be used in many computer vision applications.

One of those applications is the image synthetization. A synthesized image (or sub-aperture image) is an image simulating a picture taken from a virtual standard camera. A plenoptic raw image gives access to multiple synthesized images corresponding to different points of view of the same scene [16].

Another popular application of light-field cameras is the possibility of refocusing the image on different scene details after the picture is recorded [13], [11]. Plenoptic cameras have also been used to compute depth maps, to perform visual odometry [3], [19] or Simultaneous Localization and Mapping (SLAM) algorithms [5].

In all of these applications, the accuracy of the results strongly depends on the knowledge of the camera calibration parameters. Thus the accurate calibration of light-field cameras is a key requirement for their use. In this paper we present

¹<https://lytro.com>

a novel calibration method for light-field cameras directly expressed in the sensor space. We present a 16-intrinsic-parameter model including the focal length of the main lens and its lateral distortions, the position of the MLA and the sensor with respect to the main lens as well as the mean space between micro-lenses. To optimize this large number of parameters, we introduce a robust cost function taking into account checkerboard observations as well as micro-lenses centers in the raw image space.

II. RELATED WORK

Different plenoptic camera calibration approaches have been proposed. Processing light field cameras raw data is challenging, therefore first state of the art methods manipulate 3D features and minimize distances in 3D space.

In 2013, Dansereau et al. [4] presented a 15-parameter model (*pixel-to-ray*) to calibrate an unfocused plenoptic camera. In this work, a light field of the uncalibrated system is first generated, then synthesized images of the scene are extracted in order to detect checkerboard corners. Each observation is associated to a ray. The cost function is the squared 3D distance between each ray and its corresponding 3D point.

In 2013, Johannsen et al. [8] followed by Zeller, Noury et al. [18] presented in 2016 similar calibration methods of focused light-field cameras. In these works, synthesized images as well as depth maps are used to compute 3D observations. The cost function is then expressed in 3D space computing the difference between object points and their observations. A complete distortion model is presented including lateral and depth distortions. While in [8], depth distortion is applied on virtual image points projected by the main lens, in [18] it is applied directly on light rays crossing the MLA thereby reflecting the physical reality. Although they presented a complete distortion model there is no consideration on the MLA misalignment or the mean distance between micro-lenses.

These methods based on 3D features are appealing since they do not need to handle raw data. Indeed, feature extraction from raw image is not trivial due to the low micro-images resolution. Yet, the computation of accurate synthesized images and depth maps already requires a good estimation of intrinsic camera parameters. Here is the *Chicken-and-egg* problem: observations are needed to calibrate the camera while a calibrated camera is needed to collect observations.

To avoid this problem, one can use proprietary software such as *Lytro Desktop*² or *RxLive*³ which directly provide accurate synthesized images and depth maps to perform calibration. This however implies to rely on a software which already has all camera parameters that we wish to estimate.

In 2016, Zhang et al. [20] proposed a calibration method using bi-planar checkerboard observations from raw images. During the optimization process, checkerboard planes are reconstructed in 3D space and the difference between the computed plane inter-space and the ground truth is minimized.

These quoted methods rely on minimization in 3D space to calibrate light-field cameras but 3D features reconstruction operations (sub-aperture images and depth maps) contain interpolation phases introducing uncertainty in the process.

Closer to our work, some methods recently proposed to use only information contained in raw images so that the observations accuracy depends only on the quality of the detection. In 2014, Bok et al. [17] proposed a calibration method of an unfocused plenoptic camera based only on raw images avoiding reconstruction steps. In this work, checkerboard line features are used as observations. To detect those features they generate few sample images (with different edge orientations) depicting checkerboard borders. Then, they compare them with each selected micro-image performing a Normalized Cross-Correlation (NCC) and keep the best match. Camera parameters are optimized minimizing the squared distance between each line feature and its corresponding projected checkerboard corners in 2D space.

Despite an interesting approach, this model does not take into account all MLA parameters like its misalignment with the sensor and the mean space between micro-lenses. Since we extract checkerboard corners, the method of [17] is not suitable: considering that a corner is represented by two checkerboard edges, we would have to generate numerous samples corresponding to each line orientation and compare them with each selected micro-image. Also they worked on a Lytro camera where every micro-image has the same resolution. But some cameras (like Raytrix⁴ products) present raw images containing blurred micro-images due to different micro-lens types composing the MLA. So template matching techniques would not give accurate results.

In this paper we present a new calibration method of light-field cameras based only on raw images. To deal with the micro-image low resolution, we developed a new detector to find checkerboard observations with subpixelic accuracy in each micro-image. We propose a more complete model with 16 parameters including MLA misalignment, radial and tangential distortions of the main lens. We also optimize the mean space between the micro-lenses. To optimize all those parameters, we propose a new cost function to constrain the optimization.

In the rest of the paper, we first explain our projection model (Section III) then detail our calibration process (Section IV) and present the results on simulated and real scenes (Section V).

III. CAMERA MODEL

In this section, we describe our camera projection model. The main lens and the micro-lenses are respectively modeled as a thin lens (Section III-A) and pinholes (Section III-B).

In the following, we use homogeneous coordinates to represent 3D and 2D points. For example, homogeneous coordinates of a 3D point in the camera frame are written as $P_C = (X_C, Y_C, Z_C, 1)^T$.

²<https://illum.lytro.com/desktop>

³<https://www.raytrix.de/downloads>

⁴<https://www.raytrix.de>

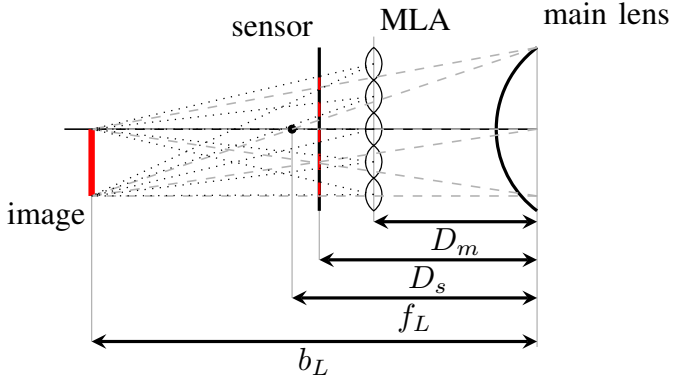


Figure 3. Zoom in Galilean plenoptic camera model. D_m and D_s respectively stand for the distance between the main lens and the MLA and the distance between the main lens and the sensor.

All frames follow the same convention: with the camera directed towards an object, the Z axis is pointing towards the object ($Z_C > 0$), the Y axis is pointing downwards and the X axis is pointing to the right.

The camera coordinate origin is located at the optical center of the main lens. The positions of camera elements are expressed in the main lens coordinate system. The origin of the MLA is placed at the matrix top left corner. The sensor's origin is located at the pixel $(0, 0)$.

A transformation matrix from a base A to B is written ${}^B T_A$ as a 4×4 matrix composed of a 3×3 rotation matrix R and a 1×3 translation matrix t .

A. Thin lens model

As depicted in Fig. 1, the image of an object through a thin lens follows the equation $\frac{1}{f_L} = \frac{1}{a_L} + \frac{1}{b_L}$ where f_L is the focal length of the lens, a_L is the distance between the object and the optical center O and b_L is the distance between the image and O .

The projection of a point $P_C = (X_C, Y_C, Z_C, 1)^\top$ to another point $P_V = (X_V, Y_V, Z_V, 1)^\top$ through the thin lens is defined as $P_V \propto K_L \cdot P_C$ with K_L :

$$\begin{pmatrix} X_V \\ Y_V \\ Z_V \\ 1 \end{pmatrix} \propto \begin{pmatrix} f_L & 0 & 0 & 0 \\ 0 & f_L & 0 & 0 \\ 0 & 0 & f_L & 0 \\ 0 & 0 & -1 & f_L \end{pmatrix} \cdot \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} \quad (1)$$

B. Pinhole model

The pinhole model describes the projection of a point $P_C = (X_C, Y_C, Z_C, 1)^\top$ to a point $p = (u, v, 1)^\top$ in image space where (u, v) are pixel coordinates. It defines the image creation using perspective projection on a plane considering that each light ray is passing through the optical center O .

The pinhole projection is expressed as $p \propto K_p \cdot P_C$ with the projection function K_p defined as:

$$K_p = \begin{pmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (2)$$

(f_x, f_y) are the pixel focal lengths with $f_x = f_p/d_x$ and $f_y = f_p/d_y$ composed of the metric focal length f_p and the metric dimensions of sensor pixels (d_x, d_y) along (x, y) . (u_0, v_0) are the principal point coordinates in the image space.

C. Distortion models

To accurately model the light projection, we consider optical aberrations applied to P_V generated by the main lens, in particular radial and tangential distortions. Due to the small size of each micro-image (20×20 pixels for a Raytrix camera), we ignore distortions generated by micro-lenses.

1) *Radial distortion*: the radial distortion is defined by a polynomial depending on $r = \sqrt{x_u^2 + y_u^2}$: the distance between the projected point (x_u, y_u) and the optical axis.

$$\Delta r_{rad} = A_0 r^3 + A_1 r^5 + A_2 r^7 + \dots \quad (3)$$

This results in the correction terms Δx_{rad} and Δy_{rad} written as:

$$\Delta x_{rad} = x_u \cdot \frac{\Delta r_{rad}}{r}, \quad \Delta y_{rad} = y_u \cdot \frac{\Delta r_{rad}}{r} \quad (4)$$

In our implementation we use a radial distortion model composed of three coefficients (A_0, A_1, A_2) .

2) *Tangential distortion*: the tangential distortion is generated by the misalignment between the sensor and the lens. We use the model defined in Brown [2] with the two first coefficients (B_0, B_1) :

$$\begin{aligned} \Delta x_{tan} &= B_0 \cdot (r^2 + 2x_u^2) + 2B_1 x_u y_u \\ \Delta y_{tan} &= B_1 \cdot (r^2 + 2y_u^2) + 2B_0 x_u y_u \end{aligned} \quad (5)$$

Once the optical aberrations are modeled, we define the distortion matrix K_D :

$$K_D = \begin{pmatrix} 1 & 0 & 0 & \Delta x_{rad} + \Delta x_{tan} \\ 0 & 1 & 0 & \Delta y_{rad} + \Delta y_{tan} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6)$$

D. Complete model

In the plenoptic camera model, light rays coming from an object are first projected through the main lens then through the MLA (Fig. 3). The projection of a scene point $P_W = (X_W, Y_W, Z_W, 1)^\top$ to a 2D point $p_l = (u, v, 1)^\top$ through the main lens and a micro-lens l follows the equation $p_l = K_l \cdot P_W$ with:

$$K_l = K_{pl} \cdot {}^l T_V \cdot K_D \cdot K_L \cdot {}^C T_W \quad (7)$$

IV. CALIBRATION METHOD

Our calibration method is based on correspondences between 3D points of a checkerboard and their observations in the image space. While this is the standard method to calibrate central cameras, its transposition to light-field cameras is not straightforward due to a different camera model with more parameters and to difficulties in robustly extracting features from small micro-images. Consequently, current approaches for plenoptic camera calibration often rely on sub-aperture

images or depth maps computed by some proprietary software [8], [15], [18] bypassing some potential interpolation steps.

We propose a calibration method entirely based on checkerboard corners detections in micro-images, the optimization being directly expressed in sensor space.

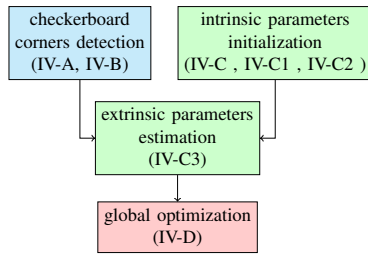


Figure 4. Calibration process of a plenoptic camera.

The calibration process is detailed in Fig. 4. Initializing micro-images grid parameters allows to estimate the exact position of each micro-image in the raw image space (Section IV-A). Once those positions are determined, each micro-image can be analyzed in order to detect corners (Section IV-B). We first initialize all intrinsic and extrinsic parameters (Section IV-C) and then refine them with a non-linear optimization approach (Section IV-D).

A. Micro-images grid parameters estimation

To analyze micro-images separately, their positions on the raw image have to be determined. We apply the method described in [4].

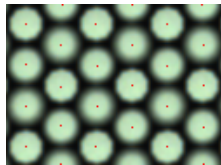


Figure 5. Zoom on a raw image with detected micro-images centers.

Using a raw image of a white scene (Fig. 5), each micro-lens center projected on the sensor is detected at the intensity maximum in the corresponding micro-image. Then all parameters of the micro-images grid can be determined. We estimate the mean space between micro-images, the translation and in-plane rotation offset of the grid in the raw image coordinate system.

Note: due to vignetting effects caused by the main lens aperture, micro-images placed at the raw image borders are not circular and their detected centers are not correctly estimated. To accurately estimate parameters of the micro-images grid, only micro-images inside an interest window are considered.

B. Checkerboard corners detection

Using the position of micro-images in the raw image space, we are able to analyze them in order to detect checkerboard corners.

The use of standard corner detectors like Harris [6] shows poor results on micro-images from light-field cameras. Many corners are detected along micro-image borders due to vignetting effect generated by micro-lenses borders. Moreover some light-field cameras, like Raytrix ones, present MLA with different types of micro-lenses, characterized by different focal lengths. Hence some micro-images are blurred avoiding to accurately extract corners. Those issues led us to design a specific method to detect corners in micro-images with subpixelic precision.

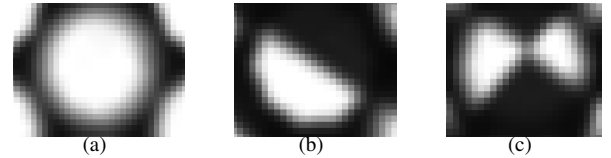


Figure 6. Types of micro-images: (a) full, (b) border et (c) corner.

Our detection process is divided in two steps: we first classify micro-images with respect to their content type (*full*, *border*, *corner*, see Fig. 6) and then we find corner locations.

To classify the content of a micro-image we propose to use histograms of gradients: the gradients in polar coordinates are computed to obtain magnitudes and orientations. A histogram is then obtained cumulating magnitude values for each orientation (Fig. 7).

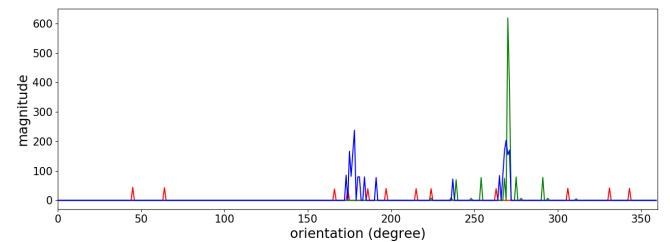


Figure 7. Histogram samples of (a) *full* (red), (b) *border* (green) and (c) *corner* (blue) micro-images. The abscissa of a peak reveals the orientation of a checkerboard border contained in the micro-image (from 0 to 360 degrees).

There are three main histogram patterns: A flat histogram for the *full* type. *border* and *corner* types respective histograms present one and two peaks. Those peaks give information on the amount of checkerboard borders and their orientations in each micro-image.

For each *corner* micro-image I (Fig. 9(a)), the corner position c_I has to be found with subpixelic precision. To this end a model image M , containing a known corner position c_M , is generated using colors intensity of I and line orientations from the corresponding histogram (Fig. 7). The corner position c_M in M is then refined with subpixelic precision using a dense optimization. In that aim, we use a Levenberg-Marquardt algorithm. We optimize a 6 parameters vector H representing a 2D affine transformation: translation (2 degrees of freedom (dof)), rotation (1 dof), stretch (1 dof), shear (1 dof) and scale (1 dof).

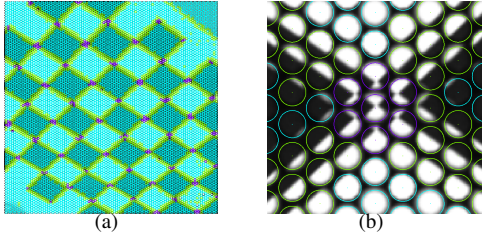


Figure 8. Characterization of micro-images types *full* (blue), *border* (green) et *corner* (purple) on Raytrix images: (a) raw image, (b) zoom.

We optimize translation parameters in order to move c_M to c_I . Rotation, stretch and shear parameters are optimized to adjust the orientation of the two edges. As explained, in case of multiple micro-lenses focal lengths, some micro-images can be blurred resulting in smoothed edges. For good fitting results and thus good corner localization on such micro-images, we consider a scale parameter which controls the level of edge blurring through interpolation.

The cost function $\varepsilon(H)$ is the sum of the quadratic errors between each pixel p of the warped model and the current micro-image I .

$$\varepsilon(H) = \sum_{p \in P} \|I(p) - M(\omega(H)(p))\|^2 \quad (8)$$

where P stands for the pixels of I . $\omega(H)$ is the warping function transforming c_M to c_I .

At the end of the optimization, c_I is computed as $c_I = \omega(H) \cdot c_M$ (Fig. 9(c)).

Note: due to vignetting effects, micro-images are bordered with a black halo. To remove those black borders from the cost function, we apply a mask to the raw image using an image of a white scene (Fig. 5).

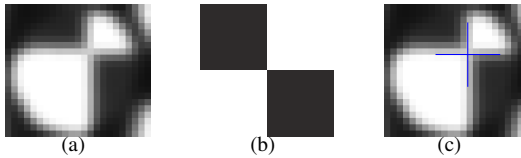


Figure 9. The corner position c_I contained in micro-image I (a) is computed optimizing the transformation between I and a model image M (b). After optimization, c_I is computed with subpixelic precision (c).

C. Parameters initialization

This section describes the intrinsic and extrinsic parameters initialization. We first present the method to set the mean space between micro-lenses (Section IV-C1) and then the other parameters initialization (Section IV-C2) followed by the poses estimation (Section IV-C3).

1) *Mean space between micro-lenses*: While a good initialization of intrinsic camera parameters can generally be obtained from the camera manufacturer, some parameters may not always be available. We explain here how to coarsely

estimate the d_μ mean space between the micro-lenses using a semi automatic method: removing the main lens, d_μ is estimated by projecting a point P far from the sensor.

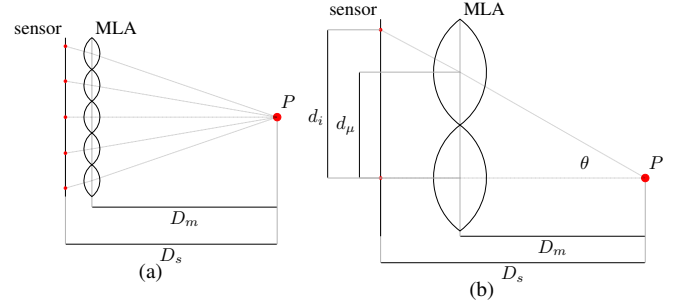


Figure 10. A point P is projected through the MLA: (a) global view, (b) zoom on two micro-lenses.

Fig. 10 depicts a plenoptic camera without the main lens, which is the configuration used to estimate d_μ . d_μ and d_i respectively stand for the distance between two micro-lenses centers (in μm) and the projection of a point P on the sensor (in metric space) (Fig. 10(b)). D_m is the distance between P and the MLA and D_s is the distance between P and the sensor plane. Here, only d_i is known since it can be estimate using the detection of micro-images centers (Section IV-A).

Let θ be the angle formed by two micro-lens centers at the point P . It is expressed as:

$$\tan \theta = \frac{d_i - d_\mu}{D_s - D_m} \quad (9)$$

When P is far, then $\theta \approx 0$. Using the equation (9) the value of d_μ is set: $d_\mu = d_i$.

2) *Other parameters initialization*: The main lens focal length is set according to the objective datasheet, distortion coefficients are set to zero and the sensor and MLA are centered around the optical axis.

3) *Pose estimation*: To optimize camera parameters, we link checkerboard corners with their observations.

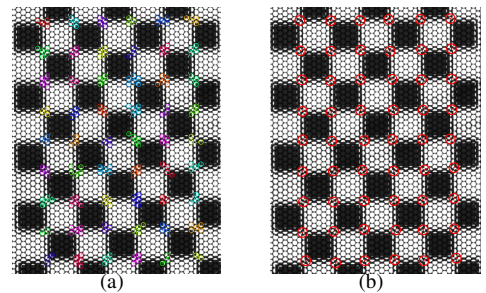


Figure 11. Raw image containing observations clusters (a) and their barycenters (b).

We estimate the camera poses using information in raw image: we compute the barycenter of each observation set corresponding to a checkerboard corner (Fig. 11(a)). Those barycenters correspond to the projection of the corners through

the main lens using the pinhole model. We then estimate the camera poses using a Perspective-n-Point (PnP) ([9]).

D. Global optimization

Once all camera projection parameters are set, we optimize them in a Levenberg-Marquardt algorithm.

parameters	dof
f_L (mm)	1
(A_0, A_1, A_2) (B_0, B_1)	3
d_u (μm)	1
MLA pose: R_m (rad), t_m (μm)	(3, 3)
sensor pose: t_s (μm)	3
camera poses: R_c (rad), t_c (mm)	(3, 3) \times F

Table I
CAMERA PARAMETERS TO OPTIMIZE.

Parameters to optimize are detailed in Table I. F stands for the number of frames used during the calibration. Hence we optimize 16 intrinsic parameters and 6 extrinsic parameters per frame.

Formally, if S is the vector of camera parameters (Table I), the calibration problem is expressed as the minimization of a cost function $\varepsilon(S)$. This function is composed of two sums of quadratic errors: the first part is equal to the sum of quadratic errors between each checkerboard corner projection through the micro-lens l using the function $K_l(S)$ and its observation $p_{f,k,l}$ in its corresponding micro-image. The second term is defined as the sum of quadratic distances between each l micro-image center $c_{i,l}$ and its corresponding projected micro-lens center.

$$\varepsilon(S) = \sum_{f=1}^F \sum_{k=1}^K \sum_{l=1}^{L_{f,k}} \|p_{f,k,l} - K_l(S) \cdot P_k\|^2 + \sum_{l=1}^L \|c_{i,l} - K_{pl}(S) \cdot c_m\|^2 \quad (10)$$

where $f \in \{1, \dots, F\}$ is the index of the current frame. P_k , $k \in \{1, \dots, K\}$ are checkerboard 3D points. $L_{f,k}$ stands for the micro-lens indices where the projection of P_k is observed in the f frame. K_L refers to the plenoptic projection function (7). c_m stands for the center of the main lens. And K_{pl} refers to the pinhole projection function through the micro-lens l (2).

During the optimization, each corner is projected through its corresponding micro-lenses (where it was observed). Considering only the first part of (10), camera parameters would be optimized without considering that the MLA is still aligned with the sensor and then the final solution would be a singularity. Hence the parameters optimization cannot only rely on checkerboard corners observations but also on micro-images centers. The second equation term enforces projected micro-lens centers to get closer to their corresponding micro-image centers.

V. EVALUATION

In this section, we evaluate the proposed method both quantitatively in a controlled environment and qualitatively with real images when ground truth parameter values are not available.

A. Experimental setup

We first describe the experimental setup for both synthetic and real images. Tests were performed on a i7 quad-core CPU (3.60 GHz).

1) *Simulation setup*: To evaluate our calibration, we have developed a simulator to generate synthetic images using our plenoptic camera projection model (Section III). This software is able to project 3D points from a virtual scene to the camera image space and to generate plenoptic raw images using ray tracing. This simulator is used to quantitatively evaluate the calibration performance in a controlled environment. Results are presented in Table III.

2) *Real scene setup*: Experiments were performed with a Raytrix R5 composed of a Baumer HGX40 sensor and a Tamron 23FM16SP lens with a 16 mm focal length. We used a 20 images dataset of a checkerboard plane at a mean distance of 60 cm. Camera poses were chosen in order to record different orientations of the checkerboard and to cover every part of the image.

All parameters were optimized using the Levenberg-Marquardt algorithm from the open source LMA C++ library ⁵ [14]. Optimized parameters of Table I are presented in Table IV.

B. Simulation

Camera poses were generated randomly according to a normal distribution at a mean distance of 80 cm from the checkerboard verifying that each corner is projected at least in one micro-image.

Once all poses and corresponding observations are generated, we introduce an initial error following a 0 mean normal distribution on the projection parameters to generate the initial parameter values. Corresponding standard deviations are given in Table II.

The calibration is tested on perfect and noised observations. Latter, we applied noises following normal distributions (set to 0 mean) on checkerboard corners and micro-image centers observations with respectively 1 pixel and 0.5 pixel standard deviations.

Results are given in Table III. On a perfect observation dataset, the final RMSE is around $1.10^{-13} \approx 0$ pixel and errors on camera parameters are negligible ($\propto 10^{-14}$).

Using noisy observations, the final RMSE is equal to 0.53 pixel and errors on final parameters are low. Here there are some example values of errors (in %) between optimized parameters and ground truth: $\delta f_L = 0.09\%$, $\delta d_\mu = 0.03\%$, $\delta D_m = 0.01\%$ and $\delta D_s = 0.04\%$. This validates the convergence of our calibration procedure.

⁵<https://github.com/bezout/lma-include>

parameters	standard deviation
f_L (mm)	1
$(A_0, A_1, A_2) \times 10^{-3}$	(1, 0, 0)
$(B_0, B_1) \times 10^{-3}$	(1, 0)
d_μ (μm)	5
MLA pose: R ($\times 10^{-3} \text{rad}$)	(1, 1, 1)
t (μm)	(100, 100, 100)
sensor pose: t (μm)	(100, 100, 100)
camera poses: R ($\times 10^{-1} \text{rad}$)	(1, 1, 1)
t (mm)	(50, 50, 50)

Table II

STANDARD DEVIATION OF THE INITIAL NOISE ON EACH PARAMETER. R STANDS FOR 3 ROTATION PARAMETERS AND t FOR 3 TRANSLATION PARAMETERS.

obs. \ images	10	20
perfect:		
initial RMSE (pixel)	47.0	41.3
final RMSE (pixel)	2.4×10^{-13}	1.7×10^{-13}
noisy:		
initial RMSE (pixel)	55.1	25.2
final RMSE (pixel)	0.53	0.53

Table III

FINAL RMSE WITH PERFECT AND NOISY OBSERVATIONS (OBS.) ACCORDING TO THE NUMBER OF IMAGES.

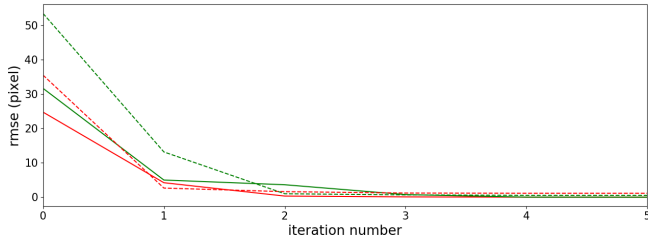


Figure 12. RMSE evolution according to the iteration number using datasets of 10 (red) and 20 (green) images with perfect (plain) and noisy (dotted) observations.

Fig. 12 depicts the RMSE evolution according to the number of iterations during the calibration. The final RMSE value is reached in less than 5 iterations. The optimization process stops in approximately 250s after 40 iterations.

Note: the plenoptic camera has to be initialized in the correct configuration (Galilean or Keplerian) because the cost function prevents from reconfiguration while optimizing.

In this paper we do not present an exhaustive study of the convergence zone. Nevertheless we present some significant clues of the robustness of our method. We initialized d_μ to different values (from $0.0\mu\text{m}$ to 1mm) and completely decentered the MLA with the sensor (from $t : (0.0, 0.0, 14.0)\text{mm}$ to $t : (-10.0, -10.0, 14.0)\text{mm}$). The optimization converges every time to the solution presented in Table IV in less than 5 iterations. This indicates that our method is robust to wrong parameters initialization especially concerning those of the MLA.

C. Real images

1) *Intrinsic parameters initialization*: Following the initialization method described in Section IV-C1, the main lens was unmounted and an image of a laser point located 4 m from the camera was taken. Thus the distance between the MLA and the sensor is negligible compared to D_m . We obtained $d_\mu = 125.84\mu\text{m}$. f_L was initialized to 16 mm and distortion coefficients to zero. Those parameters were then optimized minimizing the cost function presented in (10). Concerning the MLA and sensor poses: Rotation matrices were set to identity and translations were initialized centered around the main lens optical axis. We use a Galilean model (Fig. 1) so we initialized the MLA position behind the main lens focal plane at an arbitrary distance of 15.2 mm from the main lens. Finally the sensor is placed at 15.6 mm from the main lens.

2) *Global optimization*: During the global optimization, we optimized parameters written in Table I.

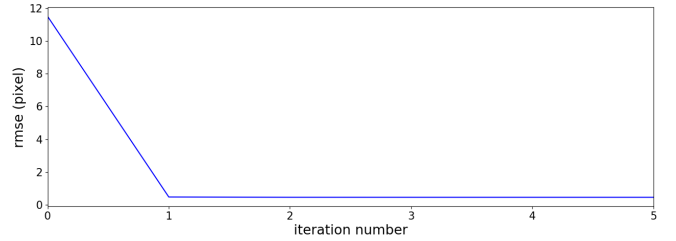


Figure 13. RMSE evolution according to the iteration number using a 20 images dataset.

The fig. 13 depicts the evolution of the RMSE during the optimization. Using our dataset, it starts at 11.5 pixels. At the first iteration it is equal to 46.99×10^{-2} pixel. The optimization process stops after the 54 iterations (364s) with a final RMSE equals to 45.19×10^{-2} pixel.

parameters	initial values	optimized values
f_L (mm)	16.00	17.02
$(A_0, A_1, A_2) \times 10^{-3}$	(0.0, 0.0, 0.0)	(-0.4, 0.0, 0.0)
$(B_0, B_1) \times 10^{-3}$	(0.0, 0.0)	(0.0, 0.2)
d_μ (μm)	125.84	124.76
MLA pose: R ($\times 10^{-3} \text{rad}$)	(0.0, 0.0, 0.0)	(0.32, 0.31, -0.56)
t ($\times 10^3 \mu\text{m}$)	(-5.56, -5.54, -15.2)	(-5.61, -5.75, -15.97)
sensor pose: t ($\times 10^3 \mu\text{m}$)	(-5.63, -5.63, -15.6)	(-5.80, -5.93, -16.37)

Table IV

INTRINSIC PARAMETERS OF THE PLENOPTIC CAMERA BEFORE AND AFTER OPTIMIZATION.

The comparison of Tables II and IV shows that the initial camera state is closer to the solution than virtual states produced in simulation. This indicates that the initial state computed with our initialization method is in the convergence zone.

3) *Validation*: Since there is no ground truth on the real camera parameters we evaluate the validity of our real camera calibration through reprojection. First, we compute 3D corners projections and show they are close to their observations (Fig. 14) with a reprojection error equals to 1.10 pixel using a 20 images dataset.

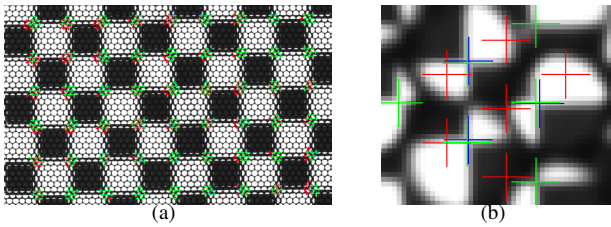


Figure 14. Total (a) and zoomed (b) raw image containing corner observations (blue) and their projections before (red) et after (green) optimization.

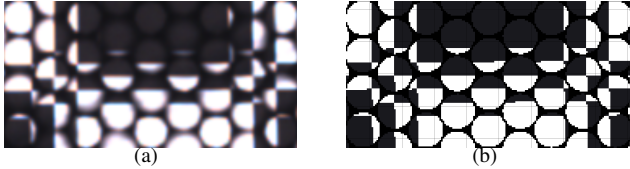


Figure 15. Zoom on a real image (a) and on a simulated image computed with the simulator using optimized projection parameters (b).

Then, in order to qualitatively evaluate the validity of our model and calibration results, we simulate a scene composed of a checkerboard and a plenoptic camera with the intrinsic and extrinsic parameters optimized with our method. Using these parameters, our simulator computes a synthetic raw image (Fig. 15(b)) similar to the real image (Fig. 15(a)).

In this section we presented quantitative and qualitative results indicating the effectiveness of our calibration method. They confirms that our proposed model fits the real camera.

VI. CONCLUSION

This article presented a new calibration method of a focused plenoptic camera based only on raw images. The problem of this calibration is formalized as the minimization of the reprojection error directly in sensor space. We proved that calibration of plenoptic cameras can be performed without synthesized images.

Due to the low resolution of micro-images we developed a new detector to estimate checkerboard observations with subpixelic accuracy.

We presented a 16-intrinsic-parameter model linking feature points to their corresponding 3D points. To optimize all those parameters we have formalized a robust cost function computing reprojection errors of checkerboard corners as well as micro-lenses centers in raw image space.

This method was validated on simulated and real images using a Raytrix camera. Future work aims to analyze the benefit of this calibration method on depth estimation using a plenoptic camera.

ACKNOWLEDGMENT

This study was possible thanks to Christophe Cudel from the MIPS Laboratory of the Université de Haute-Alsace who lent us the camera.

This research is funded by the French government research program "Investissements d'avenir" through the IMobS3 Laboratory of Excellence (ANR-10-LABX-16-01) and the FUI n°17 CLEAN Robot.

REFERENCES

- [1] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *PAMI*, 14(2):99–106, 1992.
- [2] D. C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3):444–462, 5 1966.
- [3] D. Dansereau, I. Mahon, O. Pizarro, and S. Williams. Plenoptic flow: Closed-form visual odometry for light field cameras. In *IROS*, pages 4455–4462, Sept 2011.
- [4] D. Dansereau, O. Pizarro, and S. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *CVPR*, pages 1027–1034, June 2013.
- [5] F. Dong, S.-H. Ieng, X. Savatier, R. Etienne-Cummings, and R. Benosman. Plenoptic cameras in real-time robotics. *IJRR*, 32(2):206–217, 2013.
- [6] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50, 1988.
- [7] H. E. Ives. Parallax panoramagrams made with a large diameter lens. *OSA*, 20(6):332–340, Jun 1930.
- [8] O. Johannsen, C. Heinze, B. Goldluecke, and C. Perwaß. On the calibration of focused plenoptic cameras. In *GCPR*, 2013.
- [9] V. Lepetit, F. Moreno-Noguer, and P. Fua. Epan: An accurate o(n) solution to the pnp problem. *IJCV*, 81(2), 2009.
- [10] G. Lippmann. Epreuves reversibles. photographies integrales. *Comptes Rendus De l'Academie Des Sciences De Paris*, 146:446–451, 1908.
- [11] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *ICCP*, pages 1–8, San Francisco, CA, April 2009.
- [12] R. Ng. *Digital light field photography*. PhD thesis, Stanford University, Stanford, USA, July 2006.
- [13] R. Ng, M. Levoy, M. Brédif, G. Guval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. Technical report, Stanford University, Computer Sciences, CSTR, 05 2005.
- [14] D. Ramadasan, M. Chevaldonné, and T. Chateau. Lma: A generic and efficient implementation of the levenberg–marquardt algorithm. *Software: Practice and Experience*, 2017.
- [15] K. H. Strobl and M. Lingenauber. Stepwise calibration of focused plenoptic cameras. *CVIU*, 145:140–147, 2016.
- [16] S. Xu, Z.-L. Zhou, and N. Devaney. *Multi-view Image Restoration from Plenoptic Raw Images*, chapter ACCV, pages 3–15. Springer International Publishing, Cham, 2015.
- [17] H. Yunsu Bok and I. Kweon. Geometric calibration of micro-lens-based light-field cameras using line features. *ICCV*, pages 287–300, 2014.
- [18] N. Zeller, C. Noury, F. Quint, C. Teulière, U. Stilla, and M. Dhome. Metric calibration of a focused plenoptic camera based on a 3d calibration target. *ISPRS Annals*, pages 449–456, 2016.
- [19] N. Zeller, F. Quint, and U. Stilla. Narrow field-of-view visual odometry based on a focused plenoptic camera. *ISPRS Annals*, II-3/W4:285–292, 2015.
- [20] C. Zhang, Z. Ji, and Q. Wang. Decoding and calibration method on focused plenoptic camera. *Computational Visual Media*, 2(1):57–69, 2016.