Standard Plenoptic Cameras Mapping to Camera Arrays and Calibration based on DLT

Nuno Barroso Monteiro, Joao P. Barreto, José António Gaspar, *Member, IEEE*

***Abstract*—First prototypes of standard plenoptic cameras (SPCs) were based on arrays of pinhole cameras. Despite the array nature, viewpoint pinhole arrays are not intrinsically provided by current SPC calibration tools. In this work, we start by detailing the mapping between the SPC model and a camera array of viewpoints. Then, the mapping is used to propose a calibration procedure for the SPC based on a grid of corners. Calibration involves two steps, first a linear solution and then a nonlinear optimization minimizing the ray re-projection error. The proposed calibration methodology compares favourably with state of the art calibrations and the linear solution proposed for the initial stage of the calibration outperforms the state of the art.**

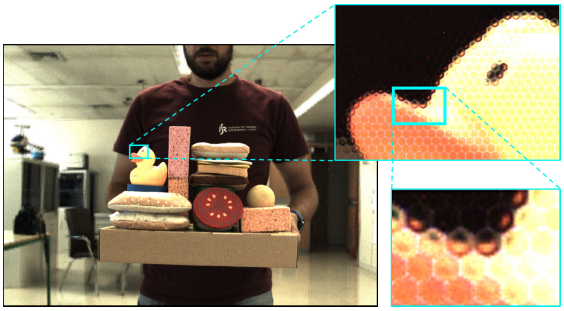
***Index Terms*—Standard Plenoptic Camera, Viewpoint Camera Array, DLT Calibration.**

I. Introduction

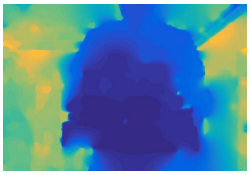
In a pinhole camera, different light rays reflected by a point in the object space are captured in a single pixel location in the image space. In a plenoptic camera, these different light ray directions are captured at different pixels, therefore creating a lightfield [2], [3]. Lightfields open new possibilities like single image depth estimation [4], [5] or refocusing [6]. Relevance and interest of these applications motivated the appearance of several types of lenslet based plenoptic cameras as the standard plenoptic camera (SPC) [7] or the focused plenoptic camera (FPC) [8]. Comprehensive introduction and review of the major lightfield concepts and capabilities can be found in overview articles as [9], [10].

In this work, we focus on the SPC which consists of a main lens, one single high definition imaging sensor, and a microlens array. The SPC geometry generates unfocused microlens images (MIs) (Figure 1.a) by placing the main lens focal plane on the microlens array plane [11].

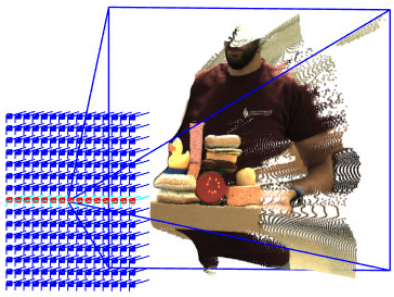
The geometry model most used for SPCs is the one proposed by Dansereau *et al.* [12]. This model maps rays in the image space indexed by pixels and microlenses indices to rays in the object space defined in metric units. The concept of viewpoint image (VI) defined by Ng *et al.* [6], obtained by selecting the same pixel for each microlens, allows to conveniently view the SPC as a camera array (Figure 1.c). Camera arrays help explaining the geometry of viewpoint cameras. However, the projection model for the viewpoint cameras is still to be fully formalized and there is no connection established with the camera model proposed by Dansereau *et al.* [12].



(a) SPC raw image and zoom of microlenses



(b) Reconstructed depth map



(c) 3D reconstruction and camera array (centers spaced 50×)

Fig. 1: Lightfield scene reconstruction and camera array. **(a)** Image captured on the sensor of a SPC with detail of the MIs formed in the sensor. **(b)** depicts the depth map obtained using [1]. **(c)** Viewpoint camera array obtained by calibration where the spacing among projection centers has been scaled 50 times to be perceptible on the 3D plot.

In this work, we build from the model of Dansereau *et al.* [12] and derive the mapping between a SPC and the viewpoint camera array. The viewpoint camera array representation is used to define a new calibration procedure. The accuracy of the mapping described is evaluated by a corner based calibration for commercially available SPCs. The code and datasets used

This work was supported by the Portuguese Foundation for Science and Technology (FCT) projects (grant numbers UID/EEA/50009/2019, PD/BD/105778/2014, and PINFRA/22084/2016) and the European Commission project ORIENT (ERC/2016/693400). J. P. Barreto thanks FCT and COMPETE2020 for generous funding under grant PTDC/EEIAUT/3024/2014.

N. B. Monteiro and J. A. Gaspar are with the Institute for Systems and Robotics, University of Lisbon, Portugal, {nmonteiro,jag}@isr.tecnico.ulisboa.pt

J. P. Barreto is with the Institute for Systems and Robotics, University of Coimbra, Portugal, jpbar@isr.uc.pt

Copyright © 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

are provided[1].

Contributions. The contributions of this work are two-fold: (i) formal definition of the projection model for a viewpoint camera, and mapping between the models of a SPC proposed by Dansereau et al. [12] and an array of viewpoint pinhole cameras, and (ii) definition of a linear solution for a SPC capable of estimating all parameters of the camera model based on the viewpoint camera array representation. This work can also be seen as an entry point to plenoptic cameras for researchers and developers acquainted with the pinhole camera model.

In terms of structure, we present in Section II a review of the camera array mappings and calibration procedures for SPCs. In Section III, we introduce the SPC model removing the redundancies with the extrinsic parameters. The mapping from the SPC model to the viewpoint camera array representation is described in Section IV. The proposed calibration procedure is presented in Section V with an emphasis on the linear solution obtained based on the viewpoint camera array representation. The results of applying the calibration proposed to calibrate comercially available SPCs are reported in Section VI and the major conclusions are presented in Section VII.

Notation: non-italic letters correspond to functions, italic letters correspond to scalars, lower case bold letters correspond to vectors, and upper case bold letters correspond to matrices. Vectors represented in homogeneous coordinates are denoted by $(\tilde{\cdot})$.

II. Related Work

The camera models proposed for SPCs [12], [13] consider the microlenses as pinholes and the main lens as a thin lens. Recent works define a projection matrix associated with each microlens of the microlens array. Namely, Bok et al. [13] describes the microlens array for a SPC using six parameters and the knowledge of the corresponding microlenses centers in the raw image. The microlenses centers are not assumed to be regularly spaced as in Dansereau et al. [12]. Zeller et al. [14] also describes the microlens camera array but for a FPC with the purpose of enabling visual odometry directly from the MIs.

Bok et al. [13] performs calibration based on line features extracted from the MIs on the raw image. This method is not adequate to calibrate the SPC when the calibration grid is placed near the world focal plane of the main lens because of the difficulty of detecting features on the unfocused MIs. On the other hand, the calibration procedure proposed by Dansereau et al. [12] is capable of calibrating the SPC even on this situation.

In Dansereau et al. [12], the light field in the image space defined in pixels (i, j) and microlenses (k, l) indices is mapped to the light field in the object space defined by a position (s, t) and a direction (u, v) in metric units (Figure 2). This mapping considers a 5×5 matrix with ten free intrinsic parameters that is obtained by propagating the rays from the sensor to the object space using ray transfer matrices. Nonetheless, there is not provided a direct connection with a projection matrix for either the microlens or viewpoint cameras. That connection, detailed in this work, allows adapting methods from the mainstream computer vision to plenoptic cameras.

The 5×5 matrix is used to calibrate a virtual SPC using corner points on VI as features. The linear solution for the calibration procedure described in [12] is based on estimating a homography for each viewpoint camera and pose of the calibration pattern. This solution estimates eight from the ten free intrinsic parameters, being the remaining two parameters estimated later in the nonlinear optimization.

The calibration with VIs requires a pre-processing step, denoted as decoding [12], to transform the 2D raw image (Figure 1.a) into a 4D light field. There are several approaches for the decoding process like the ones presented in [12], [15], [16]. In this work, we focus solely on the calibration of a SPC. The decoding originates a virtual SPC that assumes a light field that is obtained considering that the microlenses define a rectangular tiling instead of the actual hexagonal tiling (Figure 1.a).

The closer connection of the mapping proposed by Dansereau et al. [12] to a pinhole projection matrix is the one provided by Marto et al. [1] regarding the representation of a camera array composed of identical co-planar cameras. However, the mapping proposed in [1] does not explain the zero disparity in the epipolar plane images (EPIs) for points in the main lens world focal plane (box B in Figure 3.b) [17].

There are few works referring to the geometry of the viewpoint cameras. In Hahne et al. [18], the location of the viewpoints projection centers is defined using the same ray propagation strategy from sensor to object space as in Dansereau et al. [12]. Nonetheless, the complete viewpoint projection matrix is not defined and no association with a SPC model is made since the optical settings of the main lens and microlens array are assumed to be known. This work is tailored to aid in the design of new plenoptic cameras. In Bok et al. [13], a first attempt is made to define the projection matrix for the viewpoints. However, the intrinsic matrix is assumed to be common among all viewpoint cameras and is defined based on the diameter (in pixels) of a MI and the knowledge of the parameters used to describe the microlens array rather than using a geometrically approach. Furthermore, the geometry proposed does not allow to explain the zero disparity for points in the world focal plane of the main lens.

In this work, we consider the pinhole viewpoint camera constraint to represent the mapping introduced in [12] using eight free intrinsic parameters. This is accomplished by shifting the rays parameterization plane along the optical axis of the camera [19] to the plane containing the viewpoint projection centers. Additionally, we provide the mapping between the virtual SPC and the viewpoint camera array that is consistent with the zero disparity in the EPI for points in the world focal plane of the main lens. The viewpoint camera array representation is used to define a calibration for the SPC based on corner features from VIs. The linear solution proposed starts from the estimation of a single generalized homography for all viewpoints per pose of the pattern, and extending techniques from pinhole camera calibration recover the eight

[1] www.isr.tecnico.ulisboa.pt/nmonteiro/articles/plenoptic/tcsvt2019/

(b)

(a) (c)

Fig. 2: Geometry of a SPC whose main lens focal plane corresponds to plane (a). The light field in the image space is parameterized using pixel (i, j) and microlenses (k, l) indices while the light field in the object space is parameterized using a point (s, t) defined on the parameterization plane and a direction (u, v). (b) shows the raw image of a calibration grid placed on the mains lens world focal plane and (c) exhibits the details of the microlenses in red box A.

free intrinsic parameters of the camera model. Generalization is obtained considering a camera array composed of different co-planar cameras with parameterized principal point shift and baseline among viewpoints.

III. Standard Plenoptic Camera

A SPC can be represented by a 5×5 matrix H [12] which maps back-projection rays on sensor coordinates to rays in object (metric) space coordinates. In formal terms, H is a mapping of rays $\tilde{\ } = [i; j; k; l; 1]^T$ in the image space to rays $\tilde{\ } = [s; t; u; v; 1]^T$ in the object space:

$$\tilde{\ } = H \tilde{\ } \qquad (1)$$

where rays $$ are parameterized using pixel (i, j) and microlenses (k, l) indices and rays $$ are parameterized using a position (s, t) on a plane $$ and a direction (u, v) defined in metric units [17] (Figure 2.a). The mapping defined by Dansereau et al. [12] has 12 non-zero entries, however choosing the plane $$ to coincide with the plane containing the viewpoint projection centers (Supp. Material B) and removing the redundancies with the translational components of the extrinsic parameters (Supp. Material C) allows to define the mapping H with 8 non-zero entries

$$H = \begin{bmatrix} h_{si} & 0 & 0 & 0 & 0 \\ 0 & h_{tj} & 0 & 0 & 0 \\ h_{ui} & 0 & h_{uk} & 0 & h_u \\ 0 & h_{vj} & 0 & h_{vl} & h_v \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} : \qquad (2)$$

In order to help establishing the relationship between the SPC and the pinhole camera model, in the following we denominate H as the light field intrinsics matrix (LFIM). We note that LFIM is a simplified term, as H effectively contains intrinsic parameters information, however, it also contains baseline information, as detailed in Section IV. Conventional extrinsic parameters, as found in pinhole camera models, defining a world coordinate system, are in fact not contained in H.

One ray $= [s; t; u; v]^T$ can be represented as one parametric 3D line [20], namely $[x; y; z]^T = [s; t; 0]^T + [u; v; 1]^T$ for $\in \mathbb{R}$. Therefore, the LFIM matrix (2) allows to define the relationship between an arbitrary point $[x; y; z]^T$ in the object space and the ray in the image space [17] as

$$\begin{pmatrix} x \\ y \end{pmatrix} = H^{st}_{ij} \begin{pmatrix} i \\ j \end{pmatrix} + z \left(H^{uv}_{ij} \begin{pmatrix} i \\ j \end{pmatrix} + H^{uv}_{kl} \begin{pmatrix} k \\ l \end{pmatrix} + h_{uv} \right) \qquad (3)$$

where the LFIM is partitioned in three 2×2 sub-matrices and one 2×1 vector $h_{uv} = [h_u; h_v]^T$. The sub-matrices follow the notation $H^{(\ )}_{(\ )}$ where the subscript selects the columns and the superscript selects the lines, i.e., for example, H^{st}_{ij} selects the first two columns, denoted by ij, and the first two lines, denoted by st. Equation (3) shows that given one ray in image coordinates, the LFIM H allows defining a back-projection ray in the object space or, equivalently, a 3D point at a specific depth z.

IV. Viewpoint Camera Array

In this section, we represent a SPC as a camera array of viewpoints. The array representation is mapped from the SPC model defined by Dansereau et al. [12] for SPCs, namely, from the LFIM (2).

Let the projection matrix P^{ij}, parameterized by the coordinates $(i; j) \in \mathbb{Z}^2$, represent the SPC as an array

$$P^{ij} = K^{ij} \left[I_{3\times3} \quad t^{ij} \right] {}^cT_w \qquad (4)$$

where K^{ij} denotes the intrinsic matrix, $I_{3\times3}$ is a 3×3 identity matrix, t^{ij} is the projection center and ${}^cT_w = \begin{bmatrix} {}^cR_w & {}^ct_w \\ 0_{1\times3} & 1 \end{bmatrix}$ defines the rigid body transformation between the world and camera coordinate systems with rotation ${}^cR_w \in SO(3)$ and translation ${}^ct_w \in \mathbb{R}^3$, and $0_{1\times3}$ is the 1×3 null matrix.

Note that while ${}^cT_w$ defines one coordinate system for all viewpoints, the intrinsic matrix and the projection center are different for each viewpoint camera $(i; j)$. In the following, let the camera model for the viewpoint cameras (4) take into account that the principal point and the projection center are different for each viewpoint while the scale factor remains the same:

$$K^{ij} = \begin{bmatrix} k_u & 0 & u_0 + i \\ 0 & k_v & v_0 + j \\ 0 & 0 & 1 \end{bmatrix} \text{ and } t^{ij} = \begin{bmatrix} i \\ j \\ 0 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 0 \end{bmatrix} \qquad (5)$$

where the scalars k_u and k_v denote focal lengths and conversion from metric units to pixels (denominated as scale factors in the remainder of the paper). The vector $[u_0; v_0]^T$ defines the principal point for viewpoint $(i; j) = (0; 0)$, and the vectors $[u_0; v_0]^T$ and $[x_0; y_0; 0]^T$ denote principal point shift and baseline between consecutive viewpoint cameras, respectively.

A. Mapping from LFIM to Viewpoint Projection Matrices

Considering that the rays of one viewpoint camera converge to a unique point $(s;t)$ (Supp. Material B), one may set constant the values $(i;j)$ and solve equation (3) relatively to $(k;l)$. This gives an equation of a viewpoint pixel $(k;l)$ imaging the 3D point $(x;y;z)$ that can be rewritten as a pinhole model, equations (4) and (5), with the intrinsic matrix and the projection center de ned as

$$K^{ij} = \begin{bmatrix} \frac{1}{h_{uk}} & 0 & \frac{h_u}{h_{uk}} \\ 0 & \frac{1}{h_{vl}} & \frac{h_v}{h_{vl}} \\ 0 & 0 & 1 \end{bmatrix} \text{ and } t^{ij} = \begin{bmatrix} i\,h_{si} - \frac{h_{ui}}{h_{uk}} \\ j\,h_{tj} - \frac{h_{vj}}{h_{vl}} \\ 0 \end{bmatrix} : \quad (6)$$

This allows to obtain the mappings to the representations in (5). Namely, comparing (6) with (5), we identify a common component $[u_0;v_0]^T = [h_u/h_{uk};h_v/h_{vl}]^T$ and a differential (shift) component $[\Delta u_0;\Delta v_0]^T = [h_{ui}/h_{uk};h_{vj}/h_{vl}]^T$ on the principal point. The scale factors are de ned as $k_u = 1/h_{uk}$ and $k_v = 1/h_{vl}$, and the baseline is de ned as $[\Delta x_0;\Delta y_0;0]^T = [h_{si};h_{tj};0]^T$.

An example of the pinhole model parameters for a viewpoint camera array obtained from a calibrated Lytro Illum camera can be found in Table V. This array is con gured for a focused depth of about 1.09 meters and describes 15×15 cameras, $i \in \{1;\dots;15\}$, equal for j, whose VIs have 625×433 $(k;l)$ pixels.

B. Properties of Viewpoint Projection Matrices

Considering equation (3), one can obtain the EPI geometry that relates the depth of a point with the disparity on the VIs $\left[\frac{k}{i};\frac{l}{j}\right]^T$

$$\frac{k}{i} = \frac{h_{si}}{h_{uk}}\frac{1}{z} - \frac{h_{ui}}{h_{uk}} \quad \text{and} \quad \frac{l}{j} = \frac{h_{tj}}{h_{vl}}\frac{1}{z} - \frac{h_{vj}}{h_{vl}} : \quad (7)$$

The mapping (6) allows to rewrite the EPI geometry de ned in equation (7) as

$$\frac{k}{i} = k_u\frac{x_0}{z} + \Delta u_0 \quad \text{and} \quad \frac{l}{j} = k_v\frac{y_0}{z} + \Delta v_0 : \quad (8)$$

The EPI geometry shows that despite the parallel optical axis, the zero disparity plane, also known as the optical focal plane [6] of the SPC main lens is at a nite depth due to the principal point shift (box B in Figure 3.b). Considering the geometry of the camera in Figure 2.a, the zero disparity plane corresponds to the plane with $z = k_u\frac{x_0}{\Delta u_0} = k_v\frac{y_0}{\Delta v_0}$.

Contrarily, if we consider the principal point shift equal to zero, i.e. cameras with same principal point and therefore same intrinsic matrix K^{ij}, one recovers the EPI geometry de ned in [21] that de nes points at in nity as the points of zero disparity [17]. Looking at the EPIs obtained from a light eld in Figure 3, one can see that the lines corresponding to different points in the object space have a range of positive and negative slopes. Namely, objects in the background (box A) have a negative slope while objects in the foreground (box C and D) have a positive slope. The disparity zero, in this scene, corresponds to the position of the person holding the objects (box B).

Notice also that the eld of view is similar for all viewpoint cameras. Scene regions observed by the different viewpoint cameras change slightly for depths other than the zero disparity plane depth z (Figure 4.d). This is a consequence of the array of projection centers and array of principal points modeling viewpoint cameras. For the zero disparity plane depth $z = \frac{h_{si}}{h_{ui}} = \frac{h_{tj}}{h_{vj}}$, the in uence of the different projection centers is cancelled by the principal point shift and the scene region observed is the same for all viewpoint cameras (Figure 4.c).

(a)

(b)

(c)

Fig. 3: The viewpoint cameras identi ed in red in Figure 1.c are used to obtain EPIs from the light eld at rows 185 (red) (b) and 265 (green) (c) on the central viewpoint (a).

V. Standard Plenoptic Camera Calibration

The calibration proposed considers the corners of a planar calibration grid of known dimensions as features. In the following, we assume that the corners in the world coordinate system have been matched with the imaged corners. Let us consider that we have a 4D light eld obtained from the raw image (Figure 2.a) after the decoding process [12], [22]. An imaged corner is de ned by a ray $= [i;j;k;l]^T$ in the image space. The $(k;l)$ coordinates correspond to the pixel coordinates of the detected corners on the VIs. The $(i;j)$ coordinates correspond to the viewpoint coordinates.

A. Linear Initialization

In this section, we will consider the mapping in Section IV to de ne a linear solution for the LFIM H associated with a plenoptic camera and the extrinsic parameters for each pose of the calibration grid.

Homography Estimation. Considering the viewpoint projection matrix (4), a point $m = [x;y;z]^T$ in the object space is projected to a point in the image plane q by

$$q \sim P^{ij}\,m = K^{ij}\,{}^cR_w\,{}^ct_w + t^{ij}\,m \quad (9)$$

where the symbol $\sim$ denotes equal up to a scale factor. The coplanar grid points allow to de ne a world coordinate system such that the z-coordinate is zero. In this context, denoting $m = [x;y;1]^T$, one can rede ne the projection (9) as $q \sim H^{ij}\,m$ where

$$H^{ij} = K^{ij}\,[\,r_1 \;\; r_2 \;\; {}^ct_w + t^{ij}\,] \quad (10)$$

(a) VIs array model, field of view, focus plane and 10m depths

(b) Detail of A in (a) for $z = 2.0 \times 10^{-3}$ m

(c) Detail of B in (a)

(d) Detail of C in (a)

Fig. 4: Field of view of a Lytro Illum camera analyzed from the VIs array model. (a) back-projection pyramids of the four corner VIs, $(i,j) = \{(1,1),(1,15),(15,1),(15,15)\}$, where A represents the array of projection centers, B is at the focus plane at depth z, and C is at depth $z = 10$m. (b) zoom of A in (a), other VIs projection centers shown by red lines and blue dots. (c) zoom of black rectangle B in (a) showing the region observed at z is the same for all VIs. (d) zoom of black rectangle C in (a) shows slight differences of regions observed by the different viewpoint cameras.

is the parametric homography matrix for the viewpoint camera (i,j), and ${}^{c}R_w = [r_1; r_2; r_3]$. This matrix can be estimated from the point correspondences $(\bar{m}, \bar{q})$ using a direct linear transformation (DLT) [23]. Each point correspondence originates 2 linearly independent equations. The homography matrix has 9 entries to estimate but is defined only up to scale. Thus, H^{ij} has 8 degrees of freedom needing at least 4 point correspondences to estimate its entries [24]. Assuming a plenoptic camera with N pixels within each microlens and considering an independent estimation of each of the viewpoint cameras' homography matrices, one has $8N$ unknowns to estimate.

A plenoptic camera introduces restrictions on the viewpoint camera array that allows to decrease the number of unknowns to estimate. Namely, the homography matrix H^{ij} change among viewpoints as a result of the principal point shift and baseline in (6). Let us consider that H^{ij} can be defined from the the homography matrix H^0 associated with the viewpoint coordinates $(i,j) = (0,0)$ and the homography viewpoint change matrix A^{ij} by

$$H^{ij} = \underbrace{\begin{bmatrix} h^0_{11} & h^0_{12} & h^0_{13} \\ h^0_{21} & h^0_{22} & h^0_{23} \\ h^0_{31} & h^0_{32} & h^0_{33} \end{bmatrix}}_{H^0} + \begin{bmatrix} i & 0 & 0 \\ 0 & j & 0 \\ 0 & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 \end{bmatrix}}_{A^{ij}}. \tag{11}$$

Considering the homography projection of a calibration grid corner $\bar{m} = [x; y; 1]^T$ in the object space to the image point $\bar{q}$ for the viewpoint camera (i,j), applying the cross product by $\bar{q}$ on each side of the projection equation leads to $[\bar{q}]_\times H^{ij} \bar{m} = 0_{3\times 1}$, where $[\,]_\times$ is a skew-symmetric matrix that applies the cross product. Using the properties of the Kronecker product [25] and solving for each of the unknown parameters, one obtains

$$\bar{m}^T \otimes [\bar{q}]_\times \; T \begin{bmatrix} h^0 \\ a^{ij} \end{bmatrix} = 0_{3\times 1} \tag{12}$$

where

$$T = \begin{bmatrix} \begin{matrix} i & 0 & 0 & 0 & 0 & 0 \\ 0 & j & 0 & 0 & 0 & 0 \\ & & 0_{1\times 6} & & & \\ 0 & 0 & i & 0 & 0 & 0 \\ 0 & 0 & 0 & j & 0 & 0 \\ & & 0_{1\times 6} & & & \\ 0 & 0 & 0 & 0 & i & 0 \\ 0 & 0 & 0 & 0 & 0 & j \\ & & 0_{1\times 6} & & & \end{matrix} \\ I_{9\times 9} \end{bmatrix}; \tag{13}$$

and h^0 and a^{ij} correspond to vectorizations of the matrix H^0 and A^{ij} by stacking their columns and removing the zero entries, respectively. The solution $[h^0; a^{ij}]^T$ for the parametric homography matrix can be estimated using singular value decomposition (SVD) (Supp. Material D).

The restrictions introduced by a plenoptic camera allow to represent the parametric homography matrix (11) using 15 parameters. According to equation (12), each point correspondence $(\bar{m}, \bar{q})$ originates three equations with only two being linearly independent. On the other hand, each point in the object space originates N image points, one for each viewpoint camera, assuming that the point is observed in all viewpoint cameras. These pairs provide $2N$ equations that, theoretically, are enough to estimate the parametric homography matrix, assuming that $N \geq 8$. Nonetheless, the restrictions on the viewpoint camera array also originate restrictions on the projections of a point in the object space. Namely, the ray in the image space $\bar{r}^{ij} = [i; j; k; l]^T$ associated with an arbitrary viewpoint (i,j) can be described from the ray coordinates $\bar{r}^0 = [0; 0; k_0; l_0]^T$ associated with the viewpoint $(i,j) = (0,0)$ by $\bar{r}^{ij} = \bar{r}^0 + [i; j; i\,\Delta; j\,\Delta]^T$, where Δ corresponds to the disparity of the point defined on the VIs. This reduces the number of linearly independent equations originated by a point in the object space to 4. Thus, one needs at least 4 non-collinear points to obtain the entries of the homography matrix H^{ij}.

Intrinsic and Extrinsic Estimation. The structure of the homography matrix (10) in conjunction with the orthogonality and identity of the column vectors of ${}^{c}R_w$ allow to define

constraints on the intrinsic parameters as $h_1^T B^{ij} h_2 = 0$ and $h_1^T B^{ij} h_1 - h_2^T B^{ij} h_2 = 0$ [26] where h_n refers to the n-th column vector of H^{ij}, and the symmetric matrix that describes the image of the absolute conic is defined as $B^{ij} = K^{ij^{-T}} K^{ij^{-1}}$ [26], [27]. These constraints can be used to obtain the intrinsic parameters independently for each of the viewpoint cameras [26]. Alternatively, one can use the knowledge of the intrinsic matrix defined in Section IV-A to perform the estimation of a parametric representation of the absolute conic B^{ij} for a viewpoint camera (i, j) using a minimal number of parameters.

The intrinsic matrix K^{ij} differs on the principal point for each viewpoint leading to different images of the absolute conic. The principal points change regularly between consecutive viewpoints by $\left[\frac{h_{ui}}{h_{uk}}; \frac{h_{vj}}{h_{vl}}\right]^T$ which can be used to constraint the parametric representation of B^{ij}. Namely, considering (6), B^{ij} can be defined as

$$B^{ij} = B^0 + i\, C^i + j\, D^j + i^2\, E^i + j^2\, F^j \qquad (14)$$

with

$$B^0 = \begin{bmatrix} h_{uk}^2 & 0 & h_u h_{uk} \\ 0 & h_{vl}^2 & h_v h_{vl} \\ h_u h_{uk} & h_v h_{vl} & 1 + h_u^2 + h_v^2 \end{bmatrix}; \qquad (15)$$

$$C^i = \begin{bmatrix} 0 & 0 & h_{ui} h_{uk} \\ 0 & 0 & 0 \\ h_{ui} h_{uk} & 0 & 2 h_u h_{ui} \end{bmatrix}; \; E^i = \begin{bmatrix} 0_{2\times 2} & 0 \\ 0 & h_{ui}^2 \end{bmatrix}; \qquad (16)$$

$$D^j = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & h_{vj} h_{vl} \\ 0 & h_{vj} h_{vl} & 2 h_v h_{vj} \end{bmatrix}; \text{ and } F^j = \begin{bmatrix} 0_{2\times 2} & 0 \\ 0 & h_{vj}^2 \end{bmatrix}. \qquad (17)$$

This allows to define a representation for B^{ij} using 11 distinct non-zero entries $b^{ij} = [b_{11}; b_{13}; b_{22}; b_{23}; b_{33}; c_{13}; c_{33}; d_{23}; d_{33}; e_{33}; f_{33}]^T$ where $(\,)_{mn}$ represents the entry in row m and column n of the matrix $(\,)$. Considering these parameters, the intrinsic parameters constraints can be redefined as

$$\begin{bmatrix} h_{11} h_{12} & h_{11}^2 - h_{12}^2 \\ h_{11} h_{32} + h_{12} h_{31} & 2(h_{11} h_{31} - h_{12} h_{32}) \\ h_{21} h_{22} & h_{21}^2 - h_{22}^2 \\ h_{21} h_{32} + h_{22} h_{31} & 2(h_{21} h_{31} - h_{22} h_{32}) \\ h_{31} h_{32} & h_{31}^2 - h_{32}^2 \\ i\,(h_{11} h_{32} + h_{12} h_{31}) & 2i\,(h_{11} h_{31} - h_{12} h_{32}) \\ i\,(h_{31} h_{32}) & i\,(h_{31}^2 - h_{32}^2) \\ j\,(h_{21} h_{32} + h_{22} h_{31}) & 2j\,(h_{21} h_{31} - h_{22} h_{32}) \\ j\,(h_{31} h_{32}) & j\,(h_{31}^2 - h_{32}^2) \\ i^2\,(h_{31} h_{32}) & i^2\,(h_{31}^2 - h_{32}^2) \\ j^2\,(h_{31} h_{32}) & j^2\,(h_{31}^2 - h_{32}^2) \end{bmatrix}^T b^{ij} = 0_{2\times 1}. \qquad (18)$$

Normally, each homography generates 2 equations for determining the matrix of the absolute conic image [26]. The parametric representation (11), representing an arbitrary viewpoint (i, j), generates 6 equations. Nonetheless, only 2 equations are independent regarding the entries of B^0, so one needs to acquire at least 3 calibration grid poses to estimate b^{ij} defined up to a scale factor.

The intrinsic matrix parameters can be recovered from B^{ij}. More specifically, rewriting the intrinsic matrix K^{ij} (6) as

$$K^{ij} = \underbrace{\begin{bmatrix} \frac{1}{h_{uk}} & 0 & \frac{h_u}{h_{uk}} \\ 0 & \frac{1}{h_{vl}} & \frac{h_v}{h_{vl}} \\ 0 & 0 & 1 \end{bmatrix}}_{K^0} \begin{bmatrix} i & 0 & 0 \\ 0 & j & 0 \\ 0 & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 0_{3\times 2} & \frac{h_{ui}}{h_{uk}} \\ & & \frac{h_{vj}}{h_{vl}} \\ & & 0 \end{bmatrix}}_{G^{ij}}; \qquad (19)$$

one can define $B^0 = K^{0^{-T}} K^{0^{-1}}$. This allows to estimate the entries of K^0 using the Cholesky decomposition of B^0 and correcting the scale factor considering $k^0_{33} = 1$. The principal point shift can be estimated considering $\frac{h_{ui}}{h_{uk}} = \frac{e^i_{33}}{c_{13}}$ and $\frac{h_{vj}}{h_{vl}} = \frac{f^j_{33}}{d_{23}}$.

The extrinsic parameters can be estimated once the intrinsic matrix K^{ij} is known. From (10), the rotation matrix $R_w = [r_1; r_2; r_3]$ is recovered considering

$$r_1 = \lambda K^{ij^{-1}} h_1; \; r_2 = \lambda K^{ij^{-1}} h_2; \text{ and } r_3 = r_1 \times r_2: \qquad (20)$$

with $\lambda = 1/\|K^{ij^{-1}} h_1\| = 1/\|K^{ij^{-1}} h_2\|$. The translation and projection center t^{ij} are recovered solving the following system of equations

$$h_3 = K^{ij} \begin{bmatrix} i k_1 & j k_2 & \begin{bmatrix} c_{t_w} \\ h_{si} \\ h_{tj} \end{bmatrix} \end{bmatrix}. \qquad (21)$$

where k_n corresponds to the n-th column of the parametric intrinsic matrix K^{ij}.

B. Nonlinear Optimization

In this section, the linear solution is refined and radial distortion is considered on the coordinates (u, v). Namely, the undistorted rays in the object space $\mathbf{u} = [s; t; u^u; v^u]^T$ are defined from distorted rays in the object space $\mathbf{d} = [s; t; u; v]^T$ by

$$\begin{bmatrix} u^u \\ v^u \end{bmatrix} = \left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix} + \begin{bmatrix} b_u \\ b_v \end{bmatrix} \qquad (22)$$

where $\underline{u} = u^u - b_u$, $\underline{v} = v^u - b_v$, $r^2 = u^2 + v^2$, and $d = (k_1; k_2; k_3; b_u; b_v)$ defines the distortion vector. In the distortion vector, k_1, k_2 and k_3 are the radial distortion correction coefficients while the vector $[b_u; b_v]^T$ defines the distortion center. In the nonlinear optimization, we minimize the ray re-reprojection error. This optimization refines the intrinsic parameters H, the extrinsic parameters R_m (parameterized by Rodrigues formula [28]) and t_m, $m = 1; \ldots; M$ where M is the number of poses, and the distortion vector d

$$\arg\min_{H; R_m; t_m; d} \sum_{m=1}^{M} \sum_{n=1}^{N_m} \Delta_n(H; d); R_m \mathbf{m}_n + t_m \qquad (23)$$

where N_m corresponds to the number of corners detected on a pose m, $\Delta(\,)$ defines the point-to-ray distance [12], defines the direction coordinates $(u^u; v^u)$ after mapping the ray in the image space $\mathbf{m}_n$ associated with the corner n to the ray in object space (equation (1)) and followed by distortion

rectification (equation (22)) m_n defines the 3D corner point in the world coordinate system. The nonlinear optimization is solved using the trust-region-reflective algorithm [29], where a sparsity pattern for the Jacobian matrix is provided. The number of parameters over which we optimize is 4 for the intrinsic parameters, 5 for the lens distortion parameters, and $6M$ for the extrinsic parameters.

VI. Experimental Results

The calibration methodology proposed in Section V is assessed in this section using calibration datasets acquired with commercially available SPCs: the 1st generation Lytro camera and the most recent Lytro Illum.

Plenoptic cameras acquire images that have higher storage requirements than conventional cameras. Namely, the 1st generation Lytro has a raw image with 3280×3280 pixels (Figure 5.a-c) that allows to define 9×9 VIs with a resolution of 383×381 pixels after the decoding process described in [12], [22]. On the other hand, the Lytro Illum has a higher spatial and angular resolution in consequence of the higher number of microlenses in the sensor and the higher number of pixels within each microlens. More specifically, the raw image has 7728×5368 pixels (Figure 5.d-e) that allows to define 15×15 VIs with a resolution of 625×433 pixels after the decoding process described in [12], [22].

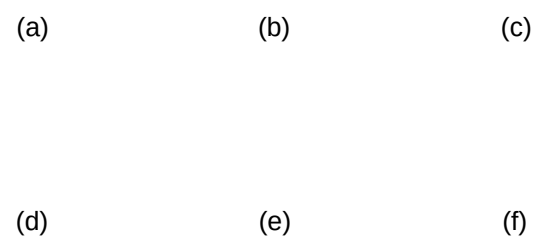


Fig. 5: Calibration data (raw images) from 1st generation Lytro camera [12] (a)-(c) and Lytro Illum (d)-(e).

A. 1st Generation Lytro State of the Art Comparison

In this section, we compare the results of the calibration procedure proposed in Section V with the calibrations proposed by Dansereau et al. [12] (denoted as Dans13) and Bok et al. [13] (denoted as Bok17). The calibration procedures are applied to publicly available calibration datasets [12] that were obtained using a 1st generation Lytro camera. For this comparison, we considered the root mean square (RMS) of the re-projection error, the ray re-projection error (i.e. distance between ray and 3D point as defined in [12]), and the reconstruction error, for three stages of the calibration process: the initial linear solution (Section V-A), the nonlinear refinement, with and without distortion estimation (Section V-B). Three errors are used in this comparison since the re-projection error is the usual error while evaluating pinhole camera calibration procedures but, in plenoptic cameras, the error normally used is the ray re-projection error [12], [13]. In addition, the reconstruction error is used to assess the quality of the reconstruction at the different stages of the calibration. The errors are summarized in Tables I, II, and III. Notice that the values from Bok et al. [13] are retrieved directly from their paper.

Comparing the results of the calibration proposed with the ones obtained using Dans13[12], one can see that the major

RMS Re-Projection Error (pix)		Dataset A	Dataset B	Dataset C	Dataset D	Dataset E
Initial	Dans13[12]	(10) 1.678 (5) 1.673	(18) 1.687 (5) 1.695	(12) 1.687 (5) 1.671	(10) 1.748 (5) 1.714	(17) 4.290 (5) 4.700
	Bok17* [13]	-	-	-	-	-
	Ours	(10) 0.838 (5) 0.797	(18) 0.856 (5) 1.035	(12) 0.950 (5) 0.953	(10) 0.965 (5) 0.790	(17) 0.840 (5) 0.627
Optimized	Dans13[12]	(10) 0.435 (5) 0.372	(18) 0.406 (5) 0.429	(12) 0.402 (5) 0.392	(10) 0.404 (5) 0.461	(17) 0.218 (5) 0.185
	Bok17* [13]	-	-	-	-	-
	Ours	(10) 0.427 (5) 0.366	(18) 0.405 (5) 0.435	(12) 0.420 (5) 0.392	(10) 0.389 (5) 0.489	(17) 0.219 (5) 0.177
Optimized (with Distortion)	Dans13[12]	(10) 0.226 (5) 0.221	(18) 0.191 (5) 0.240	(12) 0.161 (5) 0.164	(10) 0.150 (5) 0.163	(17) 0.190 (5) 0.153
	Bok17* [13]	(5) 0.374	(9) 0.259	-	-	(14) 0.274
	Ours	(10) 0.226 (5) 0.211	(18) 0.179 (5) 0.194	(12) 0.156 (5) 0.159	(10) 0.145 (5) 0.163	(17) 0.134 (5) 0.127

TABLE I: RMS re-projection error in pixels for three stages of the calibration procedure: initial linear solution, and nonlinear refinement with and without distortion estimation. The number of poses M considered for the calibration is denoted as (M). The symbol * indicates that the values reported are retrieved directly from the corresponding paper.

RMS Ray Re-Projection Error (mm)		Dataset A	Dataset B	Dataset C	Dataset D	Dataset E
Initial	Dans13* [12]	(10) 3.200	(18) 5.060	(12) 8.630	(10) 5.920	(17) 13.800
	Dans13[12]	(10) 0.577 (5) 0.627	(18) 0.603 (5) 0.570	(12) 1.036 (5) 0.974	(10) 1.231 (5) 1.081	(17) 8.900 (5) 11.970
	Bok17* [13]	-	-	-	-	-
	Ours	(10) 0.307 (5) 0.314	(18) 0.341 (5) 0.353	(12) 0.609 (5) 0.593	(10) 0.640 (5) 0.478	(17) 1.657 (5) 1.709
Optimized	Dans13* [12]	(10) 0.146	(18) 0.148	(12) 0.255	(10) 0.247	(17) 0.471
	Dans13[12]	(10) 0.154 (5) 0.145	(18) 0.147 (5) 0.139	(12) 0.260 (5) 0.245	(10) 0.260 (5) 0.268	(17) 0.485 (5) 0.546
	Bok17* [13]	-	-	-	-	-
	Ours	(10) 0.151 (5) 0.143	(18) 0.143 (5) 0.139	(12) 0.271 (5) 0.247	(10) 0.251 (5) 0.277	(17) 0.489 (5) 0.532
Optimized (with Distortion)	Dans13* [12]	(10) 0.084	(18) 0.063	(12) 0.106	(10) 0.105	(17) 0.363
	Dans13[12]	(10) 0.085 (5) 0.086	(18) 0.066 (5) 0.069	(12) 0.104 (5) 0.102	(10) 0.116 (5) 0.117	(17) 0.390 (5) 0.456
	Bok17* [13]	(5) 0.108	(9) 0.071 (5) 0.072	-	-	(14) 0.492 (5) 0.454
	Ours	(10) 0.085 (5) 0.085	(18) 0.066 (5) 0.066	(12) 0.103 (5) 0.103	(10) 0.114 (5) 0.116	(17) 0.393 (5) 0.457

TABLE II: RMS ray re-projection error in mm for three stages of the calibration procedure: initial linear solution, and nonlinear refinement with and without distortion estimation. As in Table I, (M) denotes M poses, and * indicates values retrieved from related work.

RMS Reconstruction Error (mm)		Dataset A	Dataset B	Dataset C	Dataset D	Dataset E
Initial	Dans13[12]	(10) 2100.536 (5) 3139.904	(18) 325.215 (5) 203.736	(12) 1293.985 (5) 1397.874	(10) 844.038 (5) 517.783	(17) 2702.292 (5) 3370.725
	Ours	(10) 3.039 (5) 3.904	(18) 6.212 (5) 8.023	(12) 14.899 (5) 12.558	(10) 20.751 (5) 25.316	(17) 79.681 (5) 102.281
Optimized	Dans13[12]	(10) 3.370 (5) 3.627	(18) 4.367 (5) 3.112	(12) 10.174 (5) 10.607	(10) 15.050 (5) 12.401	(17) 123.728 (5) 253.959
	Ours	(10) 3.747 (5) 3.682	(18) 4.516 (5) 3.927	(12) 10.229 (5) 8.277	(10) 15.168 (5) 12.216	(17) 142.231 (5) 187.750
Optimized (with Distortion)	Dans13[12]	(10) 4.408 (5) 4.283	(18) 4.652 (5) 4.415	(12) 9.995 (5) 8.007	(10) 15.425 (5) 13.051	(17) 135.851 (5) 179.697
	Ours	(10) 4.443 (5) 4.256	(18) 4.706 (5) 4.415	(12) 9.976 (5) 7.932	(10) 15.553 (5) 12.700	(17) 138.968 (5) 183.037

TABLE III: RMS reconstruction error in mm for three stages of the calibration procedure: initial linear solution, and nonlinear refinement with and without distortion estimation. As in Tables I and II, (M) denotes M poses.