

Generalized Camera Array Model for Standard Plenoptic Cameras

Nuno Barroso Monteiro^{1,2} and José António Gaspar¹

¹ Institute for Systems and Robotics, University of Lisbon, Portugal
² Institute for Systems and Robotics, University of Coimbra, Portugal
{nmonteiro,jag}@isr.tecnico.ulisboa.pt

Abstract. Plenoptic cameras discriminate the contribution of each ray emanating from a particular point by placing a microlens array between the main lens and the image sensor. The collection of rays captured by these cameras can represent the physical microlens camera array or can be rearranged to represent a virtual camera array with a very narrow baseline (viewpoint camera array). In this work, we extend the common camera arrays considered for standard plenoptic cameras (SPCs) and define the geometry associated with the different virtual camera arrays that can be obtained by shearing the lightfield (LF). This geometry is validated using a publicly available calibration dataset and calibration toolbox. The results show that the geometry proposed is capable of describing the multiple viewpoint and microlens camera arrays.

Keywords: Standard Plenoptic Camera, Viewpoint Camera Array, Microlens Camera Array, Shearing

1 Introduction

Plenoptic cameras discriminate the contribution of each light ray that emanate from a given point due to the positioning of a microlens array between the main lens and the image sensor. This allows to project a point in the scene onto several positions of the sensor (Figure 2.a). The collection of rays acquired by these cameras is called a lightfield (LF) [7, 11].

In this work, we will focus on the standard plenoptic camera (SPC) [15] whose geometry generates unfocused microlens images (MIs) (Figure 2.c). SPCs define several types of camera arrays by reorganizing the pixels captured by the camera on the 2D raw image (Figure 2.b) [15]. The raw image displays the pixels collected by each microlens in the microlens array (Figure 2.c) and represents the images captured by the physical microlens array placed in front of the sensor. There is another arrangement of pixels that is commonly used in SPCs, the viewpoint images (VIs). These images are obtained by selecting the same pixel position relatively to the microlens center for each microlens [15]. This rearrangement defines a virtual camera array with co-planar projection centers and with a very narrow baseline [1].

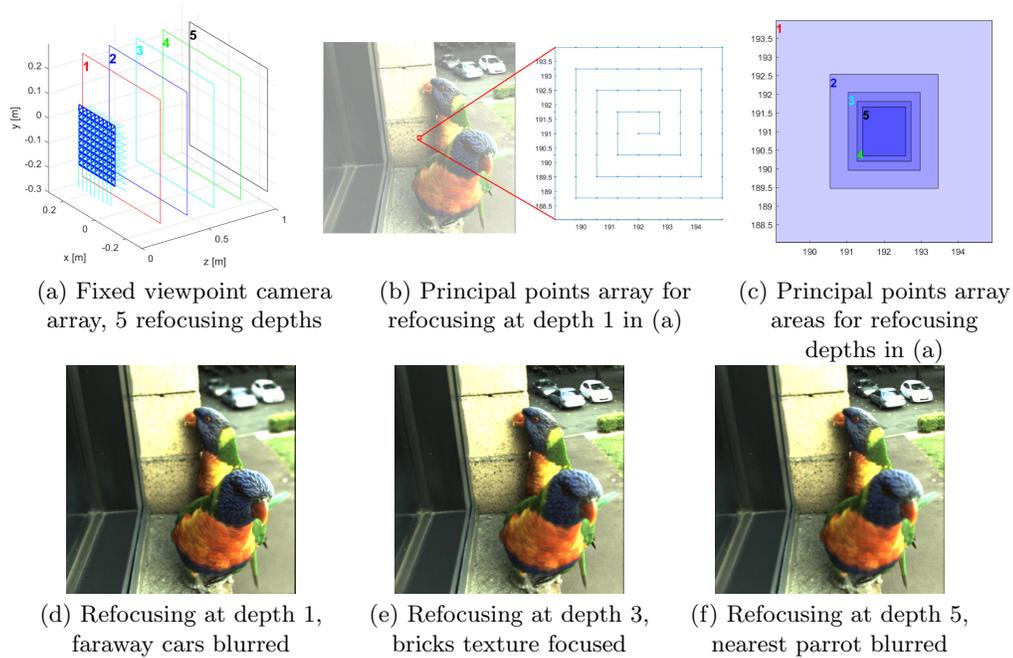


Fig. 1. Viewpoint camera arrays obtained considering shearing for refocusing at depths $z = 0.2, 0.4, \dots, 1.0$ m (a). The spacing among projection centers has been scaled 100 times to be perceptible on the 3D plot. The distribution of the principal points for the viewpoint camera arrays at different refocusing depths are depicted in (b) and (c). The corresponding refocused images are depicted in (d), (e) and (f).

A SPC allows to define additional cameras that collect rays that intersect at an arbitrary point in the scene [13] either by applying a shearing operation or creating surface camera images (SCams). Although these strategies are commonly used for disparity estimation [5, 16], the geometry associated with the corresponding cameras has not been defined. In this work, we derive the mappings between a SPC [6] and the multiple viewpoint and microlens camera arrays that can be obtained from this camera.

In terms of structure, we present in Section 2 a review of the camera array mappings for SPCs. In Section 3, we introduce the SPC model and the viewpoint and microlens camera arrays mappings considered in the literature from this model. The generalized mappings proposed for the viewpoint and microlens camera arrays are presented in Section 4. In Section 5, these mappings are validated experimentally and the major conclusions are presented in Section 6.

Notation: 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 (\cdot) .

2 Related Work

Dansereau *et al.* [6] proposed a mapping between the LF in the image space defined in pixels and microlenses indices and the LF in the object space defined by a position and a direction in metric units (Figure 2.a). Nonetheless, there is not provided a connection between this mapping and the projection matrix for either the microlens or viewpoint cameras. The definition of the projection matrices for the microlens and the viewpoint cameras appeared in the work of Bok *et al.* [4]. The geometry of the camera arrays is described using the parameters of the optical setup and the knowledge of the corresponding microlenses centers in the raw image but no relationship with the originally proposed model for SPCs [6] is provided. Additionally, the geometry proposed for the viewpoint cameras does not explain the zero disparity for points in the world focal plane of the main lens.

Marto *et al.* [12] represented a camera array composed of identical co-planar cameras using a mapping similar to the one proposed by Dansereau *et al.* [6]. However, the mapping proposed does not explain the zero disparity. The mapping between the SPC model [6] and the viewpoint camera array that is consistent with the zero disparity for points in the world focal plane of the main lens is described in [1]. This model for the viewpoint cameras consider co-planar cameras with a shifted principal point among the different viewpoint cameras. The corresponding mapping between the SPC model [6] and the microlens camera array is described in [2].

In this work, we extend the characterization of the microlens and viewpoint camera arrays found in the literature [1, 2, 4] and define the geometry of the several cameras that can be defined by collecting the rays captured by a SPC that intersect in an arbitrary point in the object space.

3 Standard Plenoptic Camera

A SPC can be represented by a 5×5 matrix \mathbf{H} [6] which maps rays $\tilde{\Phi} = [i, j, k, l, 1]^T$ in the image space to rays $\tilde{\Psi} = [s, t, u, v, 1]^T$ in the object (metric) space:

$$\tilde{\Psi} = \mathbf{H} \tilde{\Phi} \quad (1)$$

where rays Φ are parameterized using pixels (i, j) and microlenses (k, l) indices and rays Ψ are parameterized using a position (s, t) and a direction (u, v) defined on a plane Γ in metric units [14] (Figure 2.a). The mapping \mathbf{H} [6] has 12 non-zero entries

$$\mathbf{H} = \begin{bmatrix} h_{si} & 0 & h_{sk} & 0 & h_s \\ 0 & h_{tj} & 0 & h_{tl} & h_t \\ h_{ui} & 0 & h_{uk} & 0 & h_u \\ 0 & h_{vj} & 0 & h_{vl} & h_v \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \quad (2)$$

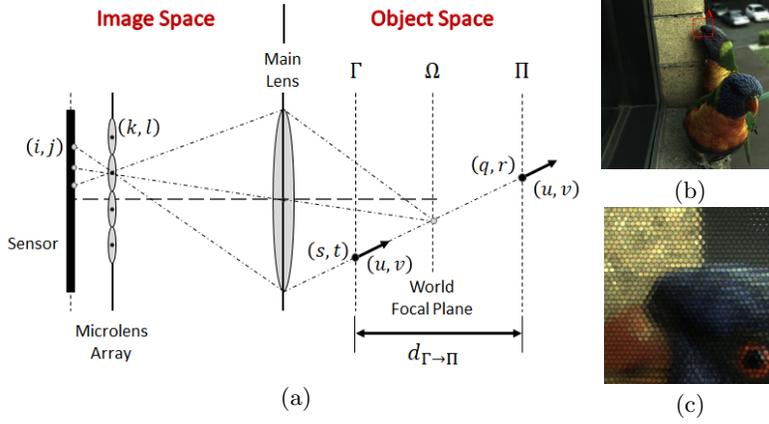


Fig. 2. Geometry of a SPC whose main lens focal plane corresponds to plane Ω (a). The LF in the image space is parameterized using pixels (i, j) and microlenses (k, l) indices while the LF in the object space is parameterized using a point (s, t) and a direction (u, v) defined on the parameterization plane Γ . (b) shows a raw image acquired with a SPC with the main lens world focal plane placed near the wall and (c) exhibits the details of the microlenses in red box A.

In the following, we denominate \mathbf{H} as the lightfield intrinsics matrix (LFIM)³.

One ray $\Psi = [s, t, u, v]^T$ can be represented as one parametric 3D line [8], namely $[x, y, z]^T = [s, t, 0]^T + \lambda[u, v, 1]^T$ for $\lambda \in \mathbb{R}$. 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 Φ in the image space [14] as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{H}_{ij}^{st} \begin{bmatrix} i \\ j \end{bmatrix} + \mathbf{H}_{kl}^{st} \begin{bmatrix} k \\ l \end{bmatrix} + \mathbf{h}_{st} + z \left(\mathbf{H}_{ij}^{uv} \begin{bmatrix} i \\ j \end{bmatrix} + \mathbf{H}_{kl}^{uv} \begin{bmatrix} k \\ l \end{bmatrix} + \mathbf{h}_{uv} \right) \quad (3)$$

where the LFIM is partitioned in four 2×2 sub-matrices and two 2×1 vectors $\mathbf{h}_{st} = [h_s, h_t]^T$ and $\mathbf{h}_{uv} = [h_u, h_v]^T$. The sub-matrices follow the notation \mathbf{H}_{ij}^{kl} where the subscript selects the columns and the superscript selects the lines, *i.e.* for example, \mathbf{H}_{ij}^{st} selects the first two columns, denoted by ij , and the first two lines, denoted by st .

Viewpoint Camera Array. The SPC can be represented by a camera array of viewpoints [1]. Let us represent the viewpoint camera array by a parametric projection matrix \mathbf{P}^{ij} varying with the coordinates (i, j)

$$\mathbf{P}^{ij} = \mathbf{K}^{ij} \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{t}^{ij} \end{bmatrix} {}^c\mathbf{T}_w \quad (4)$$

³ We note that LFIM is a simplified term, as \mathbf{H} effectively contains intrinsic parameters information, however, it also contains baseline information, as detailed in Section 3. Conventional extrinsic parameters, as found in pinhole camera models, defining a world coordinate system, are in fact not contained in \mathbf{H} .

where \mathbf{K}^{ij} denotes the intrinsic matrix, $\mathbf{I}_{3 \times 3}$ is a 3×3 identity matrix, \mathbf{t}^{ij} is the projection center and ${}^c\mathbf{T}_w = \begin{bmatrix} {}^c\mathbf{R}_w & {}^c\mathbf{t}_w \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}$ defines the rigid body transformation between the world and camera coordinate systems with rotation ${}^c\mathbf{R}_w \in SO(3)$ and translation ${}^c\mathbf{t}_w \in \mathbb{R}^3$, and $\mathbf{0}_{1 \times 3}$ is the 1×3 null matrix. Note that the intrinsic matrix and the projection center are different for each viewpoint camera (i, j) . More in detail, the intrinsic camera model takes into account that the principal point is different for each viewpoint while the scale factor remains the same. The intrinsic matrix and projection center are

$$\mathbf{K}^{ij} = \begin{bmatrix} \frac{1}{h_{uk}} & 0 & -\frac{h_u}{h_{uk}} & -i \frac{h_{ui}}{h_{uk}} \\ 0 & \frac{1}{h_{vl}} & -\frac{h_v}{h_{vl}} & -j \frac{h_{vj}}{h_{vl}} \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{t}^{ij} = - \begin{bmatrix} h_s - \frac{h_{sk}}{h_{uk}} h_u + i \left(h_{si} - \frac{h_{sk}}{h_{uk}} h_{ui} \right) \\ h_t - \frac{h_{tl}}{h_{vl}} h_v + j \left(h_{tj} - \frac{h_{tl}}{h_{vl}} h_{vj} \right) \\ -\frac{h_{sk}}{h_{uk}} \end{bmatrix}. \quad (5)$$

MicroLens Camera Array. The SPCs can also be represented by a microlens camera array [2]. Let us represent the microlens camera array by a parametric projection matrix \mathbf{P}^{kl} varying with the coordinates (k, l)

$$\mathbf{P}^{kl} = \mathbf{K}^{kl} \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{t}^{kl} \end{bmatrix} {}^c\mathbf{T}_w \quad (6)$$

where \mathbf{K}^{kl} denotes the intrinsic matrix and \mathbf{t}^{kl} is the projection center. As for the viewpoint camera array, the intrinsic matrix and the projection center are different for each microlens camera (k, l) . Namely, the intrinsic camera model takes into account that the principal point is different for each microlens while the scale factor remains the same. The intrinsic matrix and projection center are

$$\mathbf{K}^{kl} = \begin{bmatrix} \frac{1}{h_{ui}} & 0 & -\frac{h_u}{h_{ui}} & -k \frac{h_{uk}}{h_{ui}} \\ 0 & \frac{1}{h_{vj}} & -\frac{h_v}{h_{vj}} & -l \frac{h_{vl}}{h_{vj}} \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{t}^{kl} = - \begin{bmatrix} h_s - \frac{h_{si}}{h_{ui}} h_u + k \left(h_{sk} - \frac{h_{si}}{h_{ui}} h_{uk} \right) \\ h_t - \frac{h_{tj}}{h_{vj}} h_v + l \left(h_{tl} - \frac{h_{tj}}{h_{vj}} h_{vl} \right) \\ -\frac{h_{si}}{h_{ui}} \end{bmatrix}. \quad (7)$$

4 Generalized Camera Arrays

In Section 3, one defined a parametric projection matrix to define either a viewpoint or a microlens array. In this section, one shows that a plenoptic camera can define multiple camera arrays by collecting rays with different combinations of pixel and microlens coordinates.

Consider the LF in the object space $L_\Gamma(s, t, u, v)$ acquired by a plenoptic camera with the plane Ω in focus (Figure 2.a). The LF captured by the plenoptic camera can be defined on another plane by shifting the parameterization plane along the normal to the plane Γ . Assuming that the plane Π is at a distance $d_{\Gamma \rightarrow \Pi}$ from the plane Γ , one can re-parameterize the LF captured by the

plenoptic camera relatively to the plane Π [3], $L_\Pi(q, r, u, v)$, by $\tilde{\Psi}_\Pi = \mathbf{D} \tilde{\Psi}$ where

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & d_{\Gamma \rightarrow \Pi} & 0 & 0 \\ 0 & 1 & 0 & d_{\Gamma \rightarrow \Pi} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (8)$$

and $\tilde{\Psi}_\Pi = [q, r, u, v, 1]^T$ correspond to rays parameterized by a point (q, r) and a direction (u, v) on plane Π . Mapping the LF in the object space $L_\Gamma(s, t, u, v)$ to the LF in the image space $L(i, j, k, l)$ by the intrinsic matrix \mathbf{H} (2), one obtains

$$\tilde{\Psi}_\Pi = \mathbf{D} \mathbf{H} \tilde{\Phi} \quad . \quad (9)$$

The new intrinsic matrix $\mathbf{H}_\Pi = \mathbf{D}\mathbf{H}$ allows to relate the LF in the object space $L_\Pi(q, r, u, v)$ and the LF in the image space $L(i, j, k, l)$. The re-parameterization (9) allows to define a constraint to identify the rays that intersect at an arbitrary point of the plane Π [13]. Let Φ_a and Φ_b be two rays in the image space with the same coordinates (q, r) on plane Π , by taking their difference one defines a constraint on the LF coordinates in the image space as

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{H}_{ij}^{qr} \begin{bmatrix} \Delta i \\ \Delta j \end{bmatrix} + \mathbf{H}_{kl}^{qr} \begin{bmatrix} \Delta k \\ \Delta l \end{bmatrix} \quad (10)$$

where $\Delta(\cdot) = (\cdot)_b - (\cdot)_a$, and $\mathbf{H}_{(\cdot)}^{qr}$ corresponds to 2×2 sub-matrices of \mathbf{H}_Π obtained from selecting the entries of the first two rows, denoted by qr , and selecting either the entries of the 1st and 2nd columns, denoted by ij , or the 3rd and 4th columns, denoted by kl . Using the constraint (10) and considering (i_r, j_r) as reference coordinates to enforce the constraint, one defines a sampling on the viewpoint coordinates (i, j) as

$$k_S = k + \beta_{ik} (i - i_r) \wedge l_S = l + \beta_{jl} (j - j_r) \quad (11)$$

where the parameters $\beta_{ik} = -\frac{h_{si} + d_{\Gamma \rightarrow \Pi} h_{ui}}{h_{sk} + d_{\Gamma \rightarrow \Pi} h_{uk}}$ and $\beta_{jl} = -\frac{h_{tj} + d_{\Gamma \rightarrow \Pi} h_{vj}}{h_{tl} + d_{\Gamma \rightarrow \Pi} h_{vl}}$ correspond to the disparities considered on the VIs for a point at depth $d_{\Gamma \rightarrow \Pi}$.

The sampling (11) corresponds to the sampling performed during the shearing operation defined by Tao *et al.* [16]. The shearing can be interpreted as a redefinition of the epipolar plane images (EPIs) [15, 16] of the acquired LF $L(i, j, k, l)$ according to a given slope that corresponds to disparity on the VIs. Assuming that $\beta_{ik} = \beta_{jl} = \beta$ and denoting the rays in the sheared LF as $\Phi_S = [i, j, k_S, l_S]^T$, the relationship between the rays of the acquired and the sheared LF (11) can be redefined as

$$\tilde{\Phi}_S = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \beta & 0 & 1 & 0 & -\beta i_r \\ 0 & \beta & 0 & 1 & -\beta j_r \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{U}} \tilde{\Phi} . \quad (12)$$

The matrix \mathbf{U} allows to define a virtual microlens camera (k_S, l_S) that collects rays from the acquired LF that intersect at a common point in plane Π . Let us analyze the influence of shearing on the camera arrays presented in Section 3.

Viewpoint Camera Array. Following the strategy defined in [1] to compute the caustic surface, one can conclude that the caustic profile for the non-sheared and sheared viewpoint camera (i, j) is the same. Namely, the constraint to ensure a unique projection center does not change as well as the location of the projection center relatively to the non-sheared viewpoint camera. The projection matrix \mathbf{P}_S^{ij} for the sheared viewpoint camera is obtained considering the back-projection equation (3) redefined with the LFIM $\mathbf{H}_S = \mathbf{H} \mathbf{U}$ and solving relatively to (k, l) . This gives a projection matrix defined by

$$\mathbf{P}_S^{ij} = \mathbf{K}_S^{ij} [\mathbf{I}_{3 \times 3} \mathbf{t}^{ij}]^c \mathbf{T}_w \quad (13)$$

where \mathbf{K}_S^{ij} denotes the intrinsic matrix for the sheared viewpoint camera. More in detail, the camera model for the sheared viewpoint camera only differs on the principal point relatively to the non-sheared viewpoint camera counterpart (4) (Figure 1.b-c), which is consistent with the strategy to translate the VIs to perform shearing of the LF [10, 16]. The intrinsic matrix is given by

$$\mathbf{K}_S^{ij} = \begin{bmatrix} \frac{1}{h_{uk}} & 0 & -\frac{h_u}{h_{uk}} - i \frac{h_{ui}}{h_{uk}} - \beta(i - i_r) \\ 0 & \frac{1}{h_{vl}} & -\frac{h_v}{h_{vl}} - j \frac{h_{vj}}{h_{vl}} - \beta(j - j_r) \\ 0 & 0 & 1 \end{bmatrix} . \quad (14)$$

Setting $\beta = 0$, the intrinsic matrix \mathbf{K}_S^{ij} (14) reduces to the intrinsic matrix \mathbf{K}^{ij} (5).

Microlens Camera Array. Following the strategy defined in [2] to compute the caustic surface, one can conclude that the caustic profile for the sheared microlens camera (k, l) is different from the non-sheared counterpart. Namely, the constraint to ensure a unique projection center is given by

$$\frac{h_{si} + \beta h_{sk}}{h_{ui} + \beta h_{uk}} = \frac{h_{tj} + \beta h_{tl}}{h_{vj} + \beta h_{vl}} , \quad (15)$$

and the projection center is defined on a plane at a depth $z_\beta = -\frac{h_{si} + \beta h_{sk}}{h_{ui} + \beta h_{uk}}$ (Figure 3) by

$$\mathbf{t}_S^{kl} = - \begin{bmatrix} h_s + z_\beta h_u + (h_{sk} + z_\beta h_{uk})(k - \beta i_r) \\ h_t + z_\beta h_v + (h_{tl} + z_\beta h_{vl})(l - \beta j_r) \\ z_\beta \end{bmatrix} . \quad (16)$$

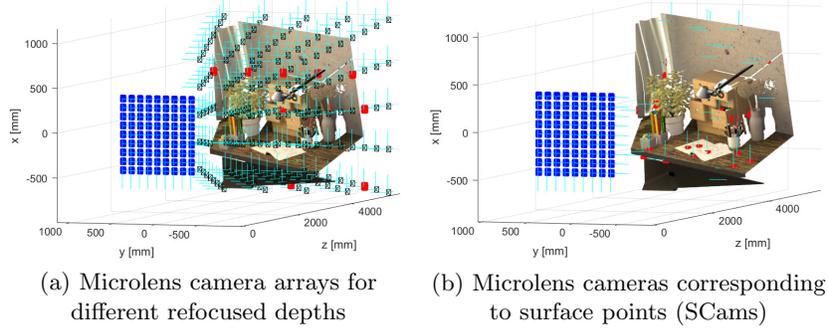


Fig. 3. Microlens camera arrays obtained considering shearing for refocusing at different depths for the synthetic Table dataset [9]. Shearing allows to obtain microlens cameras with projection centers at different depths (a). These cameras obtain relevant information for depth estimation [5] when the projection center corresponds to a surface point, *i.e.* a SCam is defined (b). The viewpoint camera array is represented in blue with the spacing among projection centers scaled by 4 times.

The projection matrix \mathbf{P}_S^{kl} for the sheared microlens camera is obtained considering the back-projection equation (3) redefined with the LFIM \mathbf{H}_S and solving relatively to (i, j) . This gives a projection matrix defined by

$$\mathbf{P}_S^{kl} = \mathbf{K}_S^{kl} \left[\mathbf{I}_{3 \times 3} \mathbf{t}_S^{kl} \right]^c \mathbf{T}_w \quad (17)$$

where \mathbf{K}_S^{kl} denotes the intrinsic matrix for the sheared microlens camera and is given by

$$\mathbf{K}_S^{kl} = \begin{bmatrix} \frac{1}{h_{ui} + \beta h_{uk}} & 0 & -\frac{h_u - \beta h_{uk} i_r}{h_{ui} + \beta h_{uk}} - k \frac{h_{uk}}{h_{ui} + \beta h_{uk}} \\ 0 & \frac{1}{h_{vj} + \beta h_{vl}} & -\frac{h_v - \beta h_{vl} j_r}{h_{vj} + \beta h_{vl}} - l \frac{h_{vl}}{h_{vj} + \beta h_{vl}} \\ 0 & 0 & 1 \end{bmatrix}. \quad (18)$$

Setting $\beta = 0$, the intrinsic matrix \mathbf{K}_S^{kl} (18) and the projection center \mathbf{t}_S^{kl} (16) reduce to the intrinsic matrix \mathbf{K}^{kl} and projection center \mathbf{t}^{kl} defined in (7). Additionally, if we replace $\beta = -\frac{h_{si} + d_{\Gamma \rightarrow \Pi} h_{ui}}{h_{sk} + d_{\Gamma \rightarrow \Pi} h_{uk}}$ in $z\beta$, one can see that the depth of the projection center corresponds to the plane Π at $d_{\Gamma \rightarrow \Pi}$.

Generalized EPI Geometry. Considering equation (3) and the sheared viewpoint cameras (13), one can obtain the EPI geometry that relates the depth of a point with the disparity on the VIs $\left[\frac{\Delta k}{\Delta i}, \frac{\Delta l}{\Delta j} \right]^T$ for the sheared LF

$$\frac{\Delta k}{\Delta i} = -\frac{h_{si} + zh_{ui}}{h_{sk} + zh_{uk}} - \beta \quad \text{and} \quad \frac{\Delta l}{\Delta j} = -\frac{h_{tj} + zh_{vj}}{h_{vl} + zh_{vl}} - \beta. \quad (19)$$

The EPI geometry shows that the zero disparity plane, also known as the optical focal plane [15] of the SPC main lens is affected by the shearing operation. This is in accordance with the creation of a virtual focal plane during the refocus

operation that implicitly requires a shearing of the LF [15] (Figure 1.d-f). These equations reduce to the ones presented in [12] for $\beta = 0$.

5 Experimental Results

In this section, the mappings proposed in Section 4 are validated experimentally using the publicly available calibration dataset [6] (Dataset A) acquired with a 1st generation Lytro camera. Namely, the viewpoint and microlens cameras obtained after calibration of the sheared versions of the calibration dataset LFs are compared with the cameras obtained using the mappings proposed in Section 4 with the LFIM obtained from the calibration of the non-sheared calibration dataset.

Let us start by calibrating the non-sheared calibration dataset using the calibration procedure [6]. The estimated LFIM \mathbf{H} (2) and the corresponding viewpoint (4) and microlens (6) cameras are given in Tables 1 and 2, respectively, where $k_{nm}^{(\cdot)}$ denotes the entry (n, m) of the intrinsic matrix and $t_n^{(\cdot)}$ denotes the entry n of the projection center associated with the viewpoint (i, j) or microlens (k, l) . Using the values in Table 1 and the mappings (13) and (17), one obtains the characterization of the camera arrays for different values of disparity β .

Table 1. LFIM obtained after calibration of Dataset A [6] with $h_{sk} = h_{tl} = 0$.

h_{si}	h_s	h_{tj}	h_t	h_{ui}	h_{uk}	h_u	h_{vj}	h_{vl}	h_v
0.0003	-0.0013	0.0003	-0.0013	-0.0011	0.0019	-0.3508	-0.0011	0.0019	-0.3515

Table 2. Intrinsic matrices and projection centers for viewpoint and microlens cameras. These values are obtained after applying the mappings (5) and (7) with $\Delta i = \Delta j = 1$ and $\Delta k = \Delta l = 1$, respectively.

k_{11}^{ij}	k_{22}^{ij}	k_{13}^{ij}	k_{23}^{ij}	t_1^{ij}	t_2^{ij}	t_3^{ij}	k_{11}^{kl}	k_{22}^{kl}	k_{13}^{kl}	k_{23}^{kl}	t_1^{kl}	t_2^{kl}	t_3^{kl}
538.6	534.9	189.6	188.6	0.001	0.001	0	881.0	892.0	-307.5	-311.9	-0.081	-0.081	-0.227

The characterization of the viewpoint and microlens cameras obtained using the mappings proposed in Section 4 is compared with the characterization obtained by applying (4) and (6) to the LFIM obtained from the calibration of the sheared versions of the calibration dataset LFs. The sheared LFs are obtained considering different disparities β for the re-parameterization of the EPIs (shearing). The disparities considered range from 0.1 to 2.0 pixels. Figure 4 depicts the entries of the viewpoint intrinsic matrix and projection center with the disparity β used for shearing considering a unitary displacement from the reference viewpoint (i_r, j_r) , *i.e.* $\Delta i = \Delta j = 1$. Similarly, Figure 5 depicts the entries of the microlens intrinsic matrix and projection center considering $\Delta k = \Delta l = 1$. Tables 3 and 4 represent the mean and Standard Deviation (STD) of the errors $\epsilon_{(\cdot)} = \left| \frac{(\cdot)^M - (\cdot)^E}{(\cdot)^M} \right|$, in percentage, for each entry of the intrinsic matrix and projection center for the viewpoint and microlens camera, respectively. In the error $\epsilon_{(\cdot)}$, $(\cdot)^M$ corresponds to the entries obtained from the mappings (13)

and (17), and $(\cdot)^E$ corresponds to the entries obtained from the mappings (4) and (6).

The viewpoint mapping (13) models the changes with the disparity β very accurately (Figure 4). In Table 3, one can see that the mean error is below 0.2% which shows that the estimate values are in accordance with the mapping (13). The difference on the estimated values appears to be the result of the interpolation and discretization that occurs in the shearing operation. This also affects the position of the detected corners that are used in the calibration.

The microlens mapping (17) also models the changes with the disparity β very accurately except for $\beta = 0.6$ (Figure 5). This disparity value is close to the singularity that occurs for $\beta = -h_{ui}/h_{uk} = 0.611$ which causes some numerical instability in the mapping. Indeed, in Table 4, one can see that the mean error considering all disparity values is below 4.5%. Nonetheless, removing the disparity $\beta = 0.6$, one obtains a mean error below 0.5% which shows that the estimate values are in accordance with the mapping (17). Notice that the viewpoint mapping obtains a lower error than the microlens mapping. This can be justified by the strategy of the calibration procedure [6] that calibrates a SPC using detected corners on VIs.

Table 3. Mean and STD error, in percentage, for each entry of the viewpoint intrinsic matrix and projection center.

k_{11}^{ij}	k_{22}^{ij}	k_{13}^{ij}	k_{23}^{ij}	t_1^{ij}	t_2^{ij}
0.022 ± 0.018	0.022 ± 0.017	0.004 ± 0.003	0.002 ± 0.002	0.174 ± 0.067	0.100 ± 0.075

Table 4. Mean and STD error, in percentage, for each entry of the microlens intrinsic matrix and projection center. First line considers all disparity values while the second line excludes the disparity $\beta = 0.6$.

k_{11}^{kl}	k_{22}^{kl}	k_{13}^{kl}	k_{23}^{kl}	t_1^{kl}	t_2^{kl}	t_3^{kl}
1.85 ± 6.45	4.45 ± 19.07	1.84 ± 6.45	4.44 ± 19.07	1.89 ± 6.38	1.89 ± 6.38	1.90 ± 6.38
0.41 ± 0.44	0.19 ± 0.18	0.40 ± 0.44	0.18 ± 0.18	0.47 ± 0.42	0.47 ± 0.42	0.48 ± 0.43

6 Conclusions

In this work, one defined the geometry of the multiple camera arrays that can be obtained from the rays captured by a SPC. This geometry extends the characterization of the microlens and viewpoint camera arrays found in the literature [1, 2, 4] and that are associated with the images that can be obtained directly from the LF acquired by a SPC. The mappings proposed for the different microlens and viewpoint cameras obtained after shearing were validated using a publicly available dataset and calibration toolbox [6]. The results show that the mappings proposed are in accordance with the calibration estimates obtained.

In terms of future work, we want to characterize the camera arrays originated considering a sampling on the microlens coordinates (k, l) instead of the sampling on the viewpoint coordinates (i, j) considered in this work.

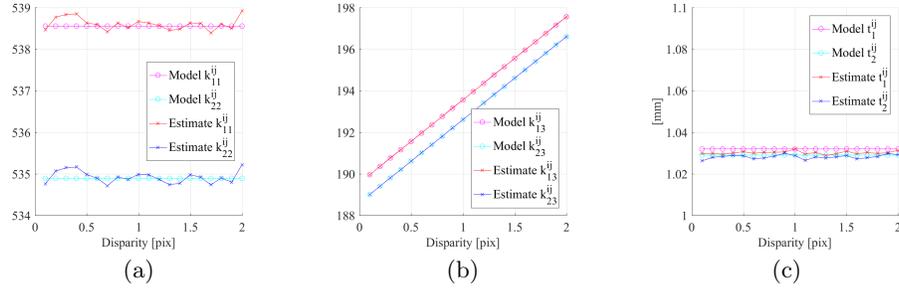


Fig. 4. Variation of viewpoint camera intrinsic matrix and projection center with disparity β for shearing. These entries are estimated considering that $\Delta i = \Delta j = 1$. The scale factors of the intrinsic matrix k_{11}^{ij} and k_{22}^{ij} are represented in (a). The principal point $[k_{13}^{ij}, k_{23}^{ij}]^T$ is depicted in (b). In (c), the x - and y - components of the projection are presented. The z -component of the projection center is not represented since it is always zero regardless of the disparity β considered for shearing.

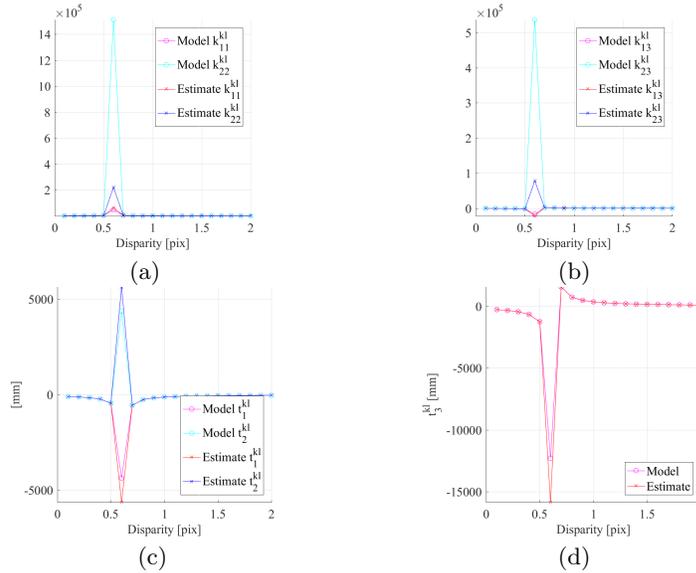


Fig. 5. Variation of microlens camera intrinsic matrix and projection center with disparity β for shearing. These entries are estimated considering that $\Delta k = \Delta l = 1$. The scale factors of the intrinsic matrix k_{11}^{kl} and k_{22}^{kl} are represented in (a). The principal point $[k_{13}^{kl}, k_{23}^{kl}]^T$ is depicted in (b). The x - and y - components of the projection are presented in (c) while the z -component is presented in (d).

Funding: This work was supported by the Portuguese Foundation for Science and Technology projects [UID/EEA/50009/2019] and [PD/BD/105778/2014], the RBCog-Lab [PINFRA/22084/2016] and FIREFRONT [PCIF/SSI/0096/2017].

References

1. Anonymous: Standard plenoptic cameras mapping to camera arrays and calibration based on dlt. TCSVT 2018 Submission ID TCSVT-02137-2018 (2018)
2. Anonymous: Geometric calibration of multi-focus plenoptic cameras. ICCV 2019 Submission ID 2256 (2019)
3. Birklbauer, C., Bimber, O.: Panorama light-field imaging. *Computer Graphics Forum* 33(2), 43–52 (2014)
4. Bok, Y., Jeon, H.G., Kweon, I.S.: Geometric calibration of micro-lens-based light field cameras using line features. *IEEE transactions on pattern analysis and machine intelligence* 39(2), 287–300 (2017)
5. Chen, C., Lin, H., Yu, Z., Bing Kang, S., Yu, J.: Light field stereo matching using bilateral statistics of surface cameras. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1518–1525 (2014)
6. Dansereau, D.G., Pizarro, O., Williams, S.B.: Decoding, calibration and rectification for lenselet-based plenoptic cameras. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1027–1034 (2013)
7. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: *Proceedings of the International conference on computer graphics and interactive techniques (SIGGRAPH)*. vol. 96, pp. 43–54. ACM (1996)
8. Grossberg, M.D., Nayar, S.K.: The raxel imaging model and ray-based calibration. *International Journal of Computer Vision* 61(2), 119–137 (2005)
9. Honauer, K., Johannsen, O., Kondermann, D., Goldluecke, B.: A dataset and evaluation methodology for depth estimation on 4d light fields. In: *Asian Conference on Computer Vision*. pp. 19–34. Springer (2016)
10. Jeon, H.G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y.W., So Kweon, I.: Accurate depth map estimation from a lenslet light field camera. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1547–1555 (2015)
11. Levoy, M., Hanrahan, P.: Light field rendering. In: *Proceedings of the International conference on computer graphics and interactive techniques (SIGGRAPH)*. vol. 96, pp. 31–42. ACM (1996)
12. Marto, S.G., Monteiro, N.B., Barreto, J.P., Gaspar, J.A.: Structure from plenoptic imaging. In: *Joint IEEE International conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. pp. 338–343. IEEE (2017)
13. Monteiro, N.B., Barreto, J.P., Gaspar, J.: Surface cameras from shearing for disparity estimation on a lightfield. In: *RECPAD - Portuguese Conference on Pattern Recognition* (2018)
14. Monteiro, N.B., Marto, S., Barreto, J.P., Gaspar, J.: Depth range accuracy for plenoptic cameras. *Computer Vision and Image Understanding* 168, 104–117 (2018)
15. Ng, R.: Digital light field photography. Ph.D. thesis, Stanford University (2006)
16. Tao, M.W., Hadap, S., Malik, J., Ramamoorthi, R.: Depth from combining defocus and correspondence using light-field cameras. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 673–680 (2013)