

# Focus Model for Metric Depth Estimation in Standard Plenoptic Cameras

Said Pertuz<sup>a,c,\*</sup>, Edith Pulido-Herrera<sup>b</sup>, Joni-Kristian Kamarainen<sup>c</sup>

<sup>a</sup>Universidad Industrial de Santander, 680002 Bucaramanga, Colombia.

<sup>b</sup>Universidad Antonio Nariño, 111321 Bogotá, Colombia

<sup>c</sup>Tampere University of Technology, 33720 Tampere, Finland.

---

## Abstract

In recent years, a lot of efforts have been devoted to the problem of depth estimation from *lightfield* images captured by standard plenoptic cameras. However, most of the metric depth estimation methods in the state-of-the-art leverage pixel disparity only. In this paper, we tackle the problem of focus-based metric depth estimation in standard plenoptic cameras. For this purpose we propose a closed-form model that relates the refocusing parameter with the focus distance of a plenoptic camera in order to allow for metric depth estimation. Based on the proposed model, we develop a calibration procedure that allows finding the parameters of the model. Using measurements of a time-of-flight sensor as ground-truth, experimental validation in a distance range of 0.2-1.6 m shows that focus-based depth estimation is feasible with a root-mean-squared error of less than 5 cm.

**Keywords:** Plenoptic camera, Lightfield, Focus, Calibration, Depth estimation

---

## 1. Introduction

One of the main advantages of plenoptic cameras is to allow a fast capture of the *lightfield* with a single snapshot in a compact device [1]. In turn, depth estimation based on the captured lightfield has many potential applications, such as passive 3D video recording, 3D modeling, augmented reality, and depth-guided scene segmentation and tracking [2]. However, these applications may often require 2.5D images with *metric* information of the scene. Unfortunately, state-of-the-art metric depth estimation approaches based on standard plenoptic cameras can deliver metric depth estimates based on disparity only, whereas the focus cue is not exploited.

At this point, it is important to distinguish between three main architectures of plenoptic cameras: coded-aperture plenoptic cameras [3, 4], focused plenoptic cameras [1], and standard plenoptic cameras [5]<sup>1</sup>. Although the feasibility of coded-aperture plenoptic imaging has been demonstrated [6], its main practical limitation is the high computational cost and processing time of the algorithms used to reconstruct the lightfield using compressed sensing theory. In contrast, both standard and focused plenoptic cameras place a microlens array (MLA) in front of the sensor in order to allow for a fast, efficient capture of the lightfield. In the standard plenoptic camera, the sensor is located at

the focal length of the MLA. In contrast, in the focused plenoptic camera, the MLA is focused at the focal plane of the main lens. Commercial versions of the standard and focused plenoptic cameras are manufactured by Lytro and Raytrix, respectively.

The main advantage of the focused plenoptic camera is the improved spatial resolution on the sampled light-field. Alternatively, the standard plenoptic camera provides a more compact and flexible design, at a significantly reduced cost. Since the public release of the first commercial version of the standard plenoptic camera, namely the Lytro camera, these features have attracted the interest of the community and fostered research in applications of plenoptic imaging. However, in the literature, most efforts in camera modeling and calibration have been devoted to the disparity cue [7, 8, 2, 9]. In contrast, current calibration models for the focus of standard plenoptic cameras are not compatible with commercial devices since they require direct control and knowledge of the internal parameters of the camera [10, 11].

This paper tackles the problem of modeling and calibrating the focus for metric depth estimation in standard plenoptic cameras. As previous researchers have shown, the main depth cues in plenoptic images are two: focus and disparity [12]. Therefore, absolute depth  $z$  can be encoded either by means of the refocusing parameter, say  $\rho$ , and disparity. Unfortunately, to the best of our knowledge, to date there are not validated models for focus-based metric depth estimation, which poses important limitations for its application in different computer vision problems. For illustration purposes, Fig. 1 shows a synthetically refocused lightfield at two different distances. Without a

---

\*Corresponding author

Email address: [spertuz@uis.edu.co](mailto:spertuz@uis.edu.co) (Said Pertuz)

<sup>1</sup>In the literature, standard plenoptic cameras have been also referred to as unfocused plenoptic cameras, multi-focus plenoptic cameras or plenoptic camera 1.0. In this work, we have adopted the term *standard plenoptic camera*



Figure 1: Digital refocusing of standard plenoptic camera. Left: refocusing with  $\rho = -1.0$ . Right: refocusing with  $\rho = 1.0$ . Without a proper model, the real metric in-focus distance corresponding to each image is unknown.

proper model, the real focus distance would remain unknown as a function of the refocusing parameter  $\rho^2$

In this work, we propose a closed-form analytical model that relates the absolute depth  $z$  of an imaged point with the refocusing parameter  $\rho$ . Based on this model, we propose a simple, reproducible calibration method for absolute depth estimation.

The contribution of this paper is two-fold:

- We propose a closed-form model that relates absolute depth  $z$  with the focusing parameter  $\rho$  in standard plenoptic cameras.
- We propose and validate an efficient method for the calibration of the focus cue in standard plenoptic cameras.

Although depth estimation is out of the scope of this work, in addition to the aforementioned contributions, we provide a benchmark with ground-truth data suitable for quantitative assessment of depth estimation in standard plenoptic cameras. The contributions of this work are of interest to the community for future work in the development of metric depth estimation methods, the fusion of multiple cues for metric depth estimation, the comparison of different depth cues, and the objective comparison of depth estimation methods in standard plenoptic cameras.

## 2. Related work

The most closely related work to our approach can be found in [11]. In that work, they derived a closed-form solution for the real depth  $z$  and a refocusing parameter defined by the authors. With that model, they predicted the depth  $z$  of the refocused image for a distance range between 43 and 890 cm approximately. The model proposed by [11] has three limitations: first, it requires accurate knowledge of the internal parameters of the camera, such as the focal length of the microlenses, the distance between

the MLA and the main lens, and the locations of the principal planes of the main lens system. Secondly, it requires a careful calibration setting with perfectly aligned optics, which can be very difficult to meet in practice. Thirdly, the model in [11] has only been validated with simulated images. In contrast, our method does not depend explicitly on internal parameters since they can be found by means of a simple calibration process. In addition, our model is validated using real images.

Calibration methods that do not require explicit knowledge of all the internal parameters of the camera have been studied for focused plenoptic cameras [8, 7, 2, 9]. In that scope, calibration of intrinsic camera parameters is also performed similarly as with standard photography cameras [14]. In [8] a laser range finder is used to provide metric reference depth measurements for the calibration. In [7, 2] and [9] they use planar calibration targets for estimating internal camera parameters that allow mapping the *virtual depth* provided by Raytrix cameras to real depth values. Very recently, a method for the geometric calibration of standard plenoptic cameras has been developed in [15]. All these methods allow for absolute depth estimation based on pixel disparity only. That is, focus is not modeled as depth cue. In contrast to the aforementioned approaches, we propose a focus model tailored for standard plenoptic cameras (Section 3). In addition, based on the proposed model, we derive a calibration method to allow for focus-based metric depth estimation in standard plenoptic cameras (Section 4).

One of the most attracting features of plenoptic cameras is the possibility of exploiting the rich angular and spatial information contained in the lightfield. This has motivated a lot of research efforts in depth estimation based on different concepts, such as EPI representations [16, 17], focus stacks [18, 19, 20], multiview stereo [21, 22], and their combinations [12, 23, 24, 25]. However, all these works provide depth estimates in method-specific depth units, which are not linearly related to metric real-world distances. The method presented here tackles this problem and provides an analytical model with a calibration procedure for focus-based metric depth estimation.

## 3. Proposed focus model

One of the main features of standard plenoptic cameras is to allow digitally refocusing captured images [5]. That is, the captured image can be synthetically refocused to a different focus distance as a function of a refocusing parameter, say  $\rho$ . In general,  $\rho$  can take any real value. However, in practice, the amount of refocusing is limited by the resolution of the sampled lightfield. In order to derive a closed-form model that relates the refocusing parameter  $\rho$  with the metric focus distance  $z$ , we first describe the decoding process of the lightfield in section 3.1. The refocusing model that relates the metric focus distance  $z$  with the refocusing parameter  $\rho$  is presented in Section 3.2.

<sup>2</sup>In this work, the refocusing parameter  $\rho$  must not be confused with the blur radius or blurring circle of conventional digital cameras [13]

### 3.1. Lightfield decoding

The light flowing through space can be described in terms of the plenoptic function. The plenoptic function, as described by Adelson and Bergen [26], is a 7-dimensional field  $P(x, y, z, \gamma, \theta, \lambda, t)$  that depends on the spatial coordinates  $(x, y, z)$ , the direction of light rays  $(\gamma, \theta)$ , the wavelength  $\lambda$  and time  $t$ . In order to understand how refocusing is performed in a standard plenoptic camera, it is important to describe the sampling of the plenoptic function that takes place in the imaging process.

Without loss of generality, using the RGB representation of color and considering static scenes, the wavelength  $\lambda$  and temporal dimension  $t$  can be dropped in order to consider a 5-dimensional plenoptic function  $P(x, y, z, \gamma, \theta)$ . In order to further reduce the dimensionality of the plenoptic function, Levoy and Hanrahan proposed a representation with only four parameters [27], namely the lightfield  $L(u_1, u_2, s_1, s_2)$ . This representation allows to parametrize the plenoptic function within a finite space delimited by two planes in terms of angular coordinates  $(u_1, u_2)$  and spatial coordinates  $(s_1, s_2)$ . For further details on the two-plane representation of the lightfield, we refer the reader to [27] and [28].

In order to sample the lightfield, standard plenoptic cameras place a microlens array (MLA) in front of the sensor of a conventional camera (see Fig. 2). The 2D sensor is located at a distance  $\beta$  behind the MLA, where  $\beta$  is the focal length of the microlenses. The aim is to demultiplex the angular information of the lightfield on the pixels behind each microlens. In a post-capture process, *lightfield decoding* is used for rearranging the 2D raw data on the sensor,  $R(\mathbf{r})$ , into the 4D representation of the lightfield  $L(\mathbf{u}, \mathbf{s})$ , where  $\mathbf{r} = [r_1, r_2]^T$  is a vector with the horizontal and vertical continuous spatial coordinates on the sensor; and  $\mathbf{u} = [u_1, u_2]^T$  and  $\mathbf{s} = [s_1, s_2]^T$  are the discrete angular and spatial coordinates of the lightfield, respectively.

An alternative representation of the lightfield can be obtained in terms of *views*. Consider the plenoptic camera with an architecture of  $M \times N$  microlenses and with  $P \times P$  pixels behind each microlens. The distorted lightfield can also be interpreted as set of  $P \times P$  views  $I_{\mathbf{u}}(\mathbf{s})$  for every value of  $\mathbf{u}$ , such that  $u_1, u_2 \in \{[-P/2], \dots, 0, \dots, [P/2]\}$  with  $P$  an odd number<sup>3</sup>. In this case, each view corresponds to a projection of the imaged scene for a specific angle  $\mathbf{u}$ . The 2D-to-4D mapping can be interpreted as a re-sampling process. In this work, we utilize the decoding and re-sampling methods proposed by Dansereau *et al.* [29]. For additional details, please refer to the supplemental material.

### 3.2. Focus model

The lightfield captured by plenoptic cameras can be used to digitally reconstruct an image corresponding to

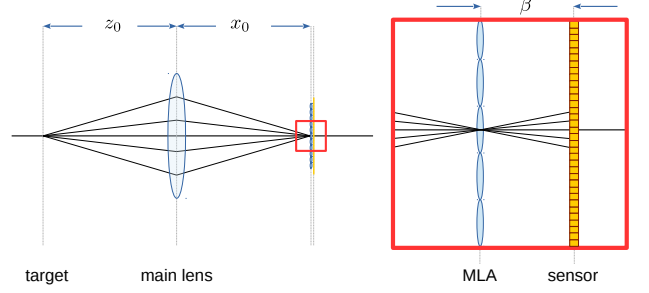


Figure 2: Architecture for lightfield capture using a standard plenoptic camera. The sensor is placed at a distance  $\beta$  behind the microlens array (MLA). The microlenses are used to demultiplex the light rays in each light cone. This architecture allows to decode as many directions as pixels there are behind each microlens.

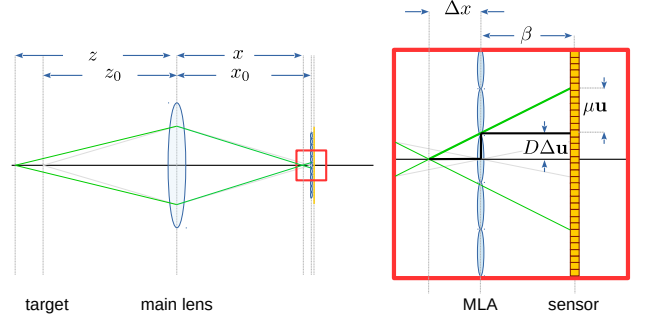


Figure 3: Lightfield refocusing. An image captured with the camera focusing to a distance  $z_0$  can be digitally refocused to a distance  $z$ . The shift  $\Delta \mathbf{s}$  is multiplied by the microlens diameter  $D$  (the same micro-lens spacing) since the units of  $\mathbf{s}$  are discrete samples. For displaying purposes, the focal length and internal lens distances are not represented up to scale.

different camera focus. Specifically, as illustrated in Fig. 3, when the MLA is located at a distance  $x_0$  from the main lens, only the points located at depth  $z_0$  will be in perfect focus. The relationship between  $x_0$  and  $z_0$  is given by the well known thin lens equation:

$$\frac{1}{f} = \frac{1}{x_0} + \frac{1}{z_0}, \quad (1)$$

where  $f$  is the focal length of the main lens<sup>4</sup>.

In this case, the image corresponding to the focal plane at  $x_0$  can be directly computed from the lightfield by accumulating all the views corresponding to different angles:

$$I(\mathbf{s}; x_0) = \sum_{\forall \mathbf{u}} I_{\mathbf{u}}(\mathbf{s}) \quad (2)$$

Taking advantage of the angular information of plenoptic images, the captured lightfield can be focused to an

<sup>3</sup>The angular views of the lightfield are often referred to as *sub-aperture images*. For the sake of simplicity we use the term *views* in this document

<sup>4</sup>The main lens focal length  $f$  (measured in meters) must not be confused with the focal lengths  $f_x$  and  $f_y$  (measured in pixels/meter) of the pin-hole model

arbitrary focal plane  $x$ . Specifically, this can be achieved by accumulating shifted versions of the views [30]:

$$I(\mathbf{s}; x) = \sum_{\forall \mathbf{u}} I_{\mathbf{u}}(\mathbf{s} + \Delta_{\mathbf{u}}) \quad (3)$$

From Fig. 3, using similar triangles it is possible to show that the shift in (3) depends on the pixel pitch  $\mu$ , the focal length of the microlenses  $\beta$ , the microlens diameter  $D$ , and the separation between the real and synthetic focal plane  $\Delta x$ :

$$\Delta_{\mathbf{u}} = \frac{\mu}{\beta D} \Delta x \mathbf{u} \quad (4)$$

For convenience, in this work we define the refocusing parameter  $\rho$  as:

$$\rho = \frac{\mu}{\beta D} \Delta x \quad (5)$$

By combining (5) and (4), and substituting in (3), we have the refocused image as a function of the refocusing parameter  $\rho$ :

$$I(\mathbf{s}; \rho) = \sum_{\forall \mathbf{u}} I_{\mathbf{u}}(\mathbf{s} + \rho \mathbf{u}) \quad (6)$$

In this paper, we are interested in establishing a closed form relationship between an arbitrary focusing distance  $z$  and the refocusing parameter  $\rho$ . From the thin lens equation in (1), it can be readily shown that  $z = xf/(x - f)$ . In addition, since  $x = x_0 - \Delta x$ , we obtain:

$$z = \frac{f(x_0 - \Delta x)}{x_0 - \Delta x - f} \quad (7)$$

Finally, by substituting  $\Delta x = \beta D \rho / \mu$  in (7), we find the sought relationship as:

$$z = z_0 \left( \frac{1 - a_0 \rho}{1 - a_1 \rho} \right), \quad (8)$$

where  $a_0 = \beta D / \mu x_0$ ,  $a_1 = z_0 a_0 / f$  and  $z_0 = x_0 f / (x_0 - f)$  is the current focus of the camera.

The model in (8) has several important properties. Firstly, it is clear that for  $\rho = 0$ , the refocusing distance is  $z = z_0$ . This is the behavior expected since (6) is equivalent to (2) when no refocusing is performed. A careful analysis of (8) also reveals that the refocusing distance  $z$  depends on the sign of  $\rho$ . Specifically,  $z > z_0$  when  $\rho > 0$  and  $z < z_0$  when  $\rho < 0$ . Finally, it is possible to verify that there is a singularity at  $\rho = 1/a_1$ . It is possible to demonstrate that this appears when the refocusing plane is located at  $x = f$ . This is, when the lightfield is refocused at the focal length of the main lens. This singularity is, therefore, expected since this corresponds to a focus at infinity  $z \rightarrow \infty$ . These facts are summarized in Fig. 4. A formal demonstration of these properties can be found in Appendix 8. The experimental validation of this model is presented in Section 6.1.

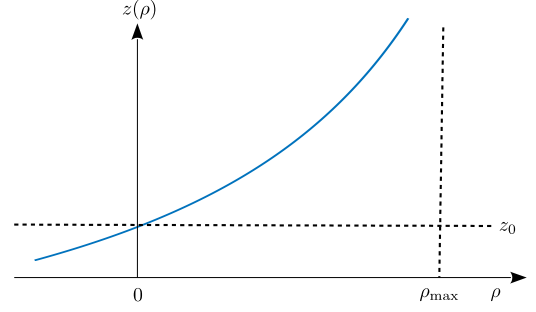


Figure 4: Proposed focus model.  $z_0$  is the focusing distance with no refocus (i.e., the actual focus of the camera).  $\rho_{\max}$  is the value of the refocusing parameter for target at infinity.

## 4. Focus calibration

Based on the focus model developed in previous section, we propose a calibration method that allows finding the parameters  $a_0$ ,  $a_1$  and  $z_0$  of (8). The overall procedure is summarized in Fig. 5.

### 4.1. Estimation of depth references

Since the un-rectified central view  $I_0(\mathbf{s})$  is generated from the actual projection centers of the micro-lenses, it can be interpreted as a conventional 2D photograph captured with a camera with a reduced aperture [5]. Since the change in aperture does not affect the parameters of the pin-hole projection model, it is possible to utilize standard camera calibration algorithms in order to find the intrinsic matrix and distortion parameters of the main lens of the plenoptic camera. Specifically, in this paper we use the calibration method presented in [14, 31]. This method allows obtaining not only the pin-hole parameters of the main lens but also a set of reference depth values for subsequent calibration of the refocus parameter. This procedure is summarized below.

Consider a 3D point in real world metric units  $\mathbf{P} = [x, y, z]^T$  that is projected to the *undistorted* image coordinates  $\tilde{\mathbf{s}} = [\tilde{s}_1, \tilde{s}_2]^T$  by means of the pinhole projection model  $\omega[\tilde{\mathbf{s}}^T, 1]^T = \mathbf{K}\mathbf{M}[\mathbf{P}^T, 1]^T$ . In this projection,  $\omega$  is a scale factor, and  $\mathbf{K}$  and  $\mathbf{M}$  are the intrinsic and extrinsic matrices of the pin-hole model:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{M} = [\mathbf{R} \quad \mathbf{t}] \quad (9)$$

where  $f_x$  and  $f_y$  are the horizontal and vertical focal lengths of the main lens;  $c_x$  and  $c_y$  are the coordinates of the principal point;  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$  is a rotation matrix, and  $\mathbf{t} \in \mathbb{R}^{3 \times 1}$  is a translation vector.

The distorted image coordinates  $\mathbf{s}$  are related to the undistorted coordinates  $\tilde{\mathbf{s}}$  by  $s_1 = p_1(\tilde{\mathbf{s}}; \kappa)$  and  $s_2 = p_2(\tilde{\mathbf{s}}; \kappa)$ , where  $p_1(\cdot)$  and  $p_2(\cdot)$  are two polynomials that model the



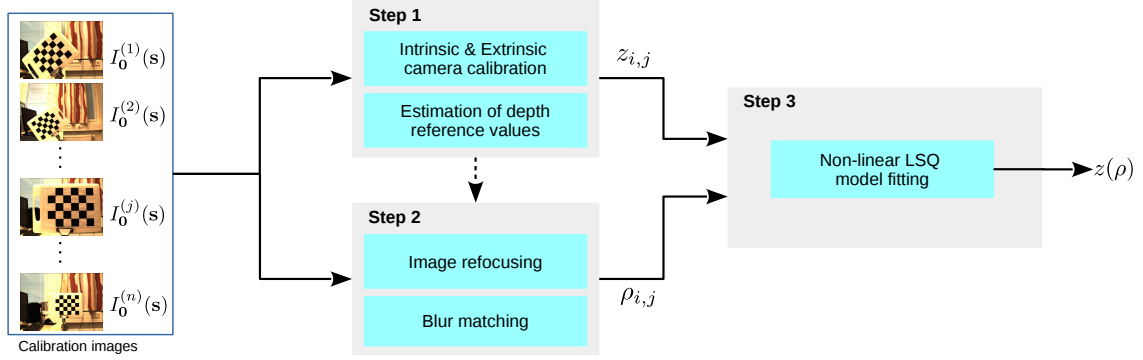


Figure 5: Proposed focus calibration method. The input is the set of central views corresponding to  $n$  different poses of a checkerboard calibration target:  $I_0^{(j)}(s)$  for  $j = 1, \dots, n$ . In the first step, standard camera calibration is used in order to compute a set of reference depth values  $z_{i,j}$  for  $j = 1, \dots, n$  and  $i = 1, \dots, m$ , where  $m$  is the number of corners of the calibration target. In the second step, a value of the refocusing parameter is associated to each reference point by means of blur-matching  $z_{i,j} \leftarrow \rho_{i,j}$  (sec. 4.2). Finally, the parameters of the model are found by non-linear LSQ fit to all  $z_{i,j} - \rho_{i,j}$  pairs (sec. 4.3).

radial and tangential distortion of the lens, and  $\kappa = \{k_1, k_2, \dots, k_n\}$  is the set of five distortion coefficients of the model [14].

Now consider a planar calibration target with a checkerboard pattern of known dimensions. As shown in Fig. 5, the checkerboard pattern is placed at  $n$  different poses with respect to the camera and contributes with a set of corner points  $c = \{\mathbf{P}_{1,1}, \mathbf{P}_{1,2}, \dots, \mathbf{P}_{m,n}\}$ , where  $m$  is the number of corners of the calibration pattern, and  $\mathbf{P}_{i,j} = [x_{i,j}, y_{i,j}, z_{i,j}]^T$  are the 3D coordinates of the  $i$ -th corner point at the  $j$ -th plane pose. The output of the calibration process is the set of intrinsic camera parameters, and the set of 3D coordinates (in camera frame) of each corner point  $\mathbf{P}_{i,j}$  for all  $(i, j)$ -pairs. Notice that each calibration point is associated to a specific depth value  $z_{i,j}$ . In the sequel, we refer to each  $z_{i,j}$  as a *depth reference value*.

#### 4.2. Blur matching

In the proposed model, the relationship between the depth  $z$  and the refocusing parameter  $\rho$  depends on the focal length of the main lens,  $f$ , the focus at which the image was captured,  $z_0$ , and the constant  $a_0$ . In general, these parameters are unknown or, in the best of cases, are only known approximately. Therefore, we propose to estimate these parameters by fitting the curve  $z$  vs.  $\rho$  as follows:

Let  $\mathbf{P}_{i,j} = [x_{i,j}, y_{i,j}, z_{i,j}]^T$  be the real coordinates of the  $(i, j)$ -th corner obtained in the calibration of main lens, as described in previous section. The aim is to find the refocusing parameter  $\rho_{i,j}$  that corresponds to that specific depth reference value  $z_{i,j}$ . For this purpose, we define  $\phi_{i,j}(\rho)$ , the focus degree of the corner at  $\mathbf{P}_{i,j}$ , as a function of the refocusing parameter  $\rho$ :

$$\phi_{i,j}(\rho) = \sum_{\mathbf{s} \in \Omega(\mathbf{P}_{i,j})} |\nabla_x^2 I(\mathbf{s}; \rho)| + |\nabla_y^2 I(\mathbf{s}; \rho)| \quad (10)$$

where  $\nabla_x^2$  and  $\nabla_y^2$  are the horizontal and vertical Laplacian operators; and  $\Omega(\mathbf{P}_{i,j})$  are the set of image coordinates

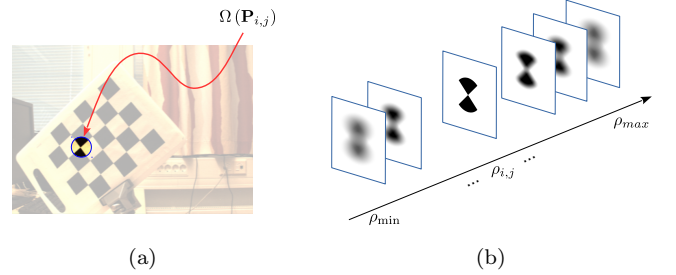


Figure 6: Blur matching. In order to find the refocus parameter corresponding to the depth reference  $z_{i,j}$ , we seek to maximize the degree of focusing of  $\Omega(\mathbf{P}_{i,j})$ , namely  $\phi_{i,j}(\rho)$ .

$\mathbf{s}$  such that  $\|\mathbf{s} - \mathbf{s}_{i,j}\| < r$ , where  $r$  is a given radius in pixels, and  $\mathbf{s}_{i,j}$  is the projection of  $\mathbf{P}_{i,j}$  on the image plane (see Fig. 6a). Intuitively, blur matching for calibration purposes is equivalent to finding the sharpest image of the edges of the calibration pattern. However, the focus degree of (10) yields a more robust result for general targets (see supplemental material).

The sought refocusing parameter  $\rho_{i,j}$  is found by means of a greedy search that maximizes the focus degree of the corresponding corner:

$$\rho_{i,j} = \arg \max_{\rho} \phi_{i,j}(\rho) \quad (11)$$

By solving (11), each depth reference  $z_{i,j}$  is matched to a specific refocusing parameter  $\rho_{i,j}$ .

#### 4.3. Non-linear LSQ fitting

The problem in (10) and (11) is analogous to the shape-from-focus problem [19]. In Fig. 6b, we show the refocused corner in  $\Omega(\mathbf{P}_{i,j})$  for different values of  $\rho$ . The problem in (11) finds the refocusing parameter that corresponds to the maximal focus degree as measured by (10). Since the corner at  $\mathbf{P}_{i,j}$  is associated with a unique depth  $z_{i,j}$ , this solution establishes a one-to-one correspondence between

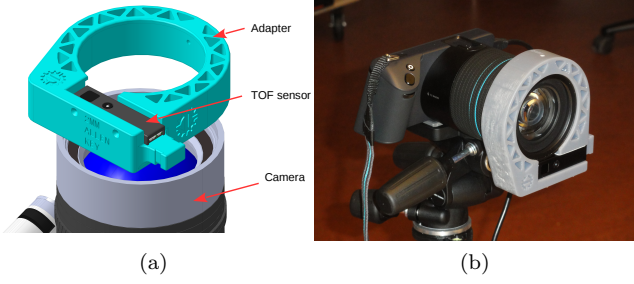


Figure 7: TOF sensor-panoptic camera setup for experimental validation. (a) Adapter with TOF sensor. (b) Assembling with panoptic camera.

a reference depth value and the corresponding refocusing parameter  $z_{i,j} \leftarrow \rho_{i,j}$ . Finally, the parameters in (8) can be found by non-linear least squares fit to all the  $z_{i,j} - \rho_{i,j}$  pairs:

$$\arg \min_{z_0, a_0, a_1} \sum_{\forall(i,j)} |z(\rho_{i,j}) - z_{i,j}|^2 \quad (12)$$

## 5. Validation methodology

In order to validate the proposed model using independent, objective measurements, we used a TOF sensor. The aim is to compare depth estimates as predicted by the proposed model with ground-truth measures obtained from the TOF sensor. For this purpose, we designed and 3D-printed an adapter for assembling a standard panoptic camera Lytro Illum<sup>5</sup> with a Camboard Picoflex TOF sensor<sup>6</sup> (Fig. 7).

The TOF sensor and the panoptic camera are pre-calibrated using standard camera calibration and 3D-to-2D pose estimation [14, 32, 33]. The pre-calibration step allows finding the matrix of intrinsic parameters and distortion coefficients of the camera, namely  $\mathbf{K}$  and  $\kappa$ , as well as the matrix of extrinsic parameters of the camera with respect to the TOF sensor  $\mathbf{M}$ . In the validation process, we use the corners of the checkerboard pattern corresponding to a *test target*. It is important to remark that it is not strictly necessary to use corner points in the validation step since the test targets are not part of the calibration process. However, since the corners of the checkerboard can be detected reliably they can be used in order to reduce the effects due to registration errors between the sensor and camera that could bias the evaluation of the model. The proposed model is validated using measurements from the TOF sensor as ground-truth as described below.

Firstly, a corner point is detected based on the intensity image captured by the TOF sensor (Fig. 8a). Let  $\mathbf{P}_s = [x_s, y_s, z_s]^T$  be the 3D world coordinates provided by the TOF sensor corresponding to the detected corner. The

---

### Algorithm 1 Validation using TOF sensor

---

**Require:** a set of 3D coordinates of  $n$  corners in the test target:  $\mathbf{P} = \{\mathbf{P}_s^{(j)} \text{ for } j = 1, 2, \dots, n\}$   
**for all**  $\mathbf{P}_s^{(j)} \in \mathbf{P}$  **do**  
 $\tilde{\mathbf{s}} \leftarrow \text{project}(\mathbf{P}_s)$  using (13)  
 $\mathbf{s} \leftarrow \text{distort}(\tilde{\mathbf{s}})$  using (14)  
 $z^* \leftarrow \text{depth}(\mathbf{s})$   
 $e_j = z^* - z_s$   
**end for**  
**return** RMSE( $e$ )

---

point is then projected to undistorted image coordinates  $\tilde{\mathbf{s}}$ , that corresponds to the central view in camera frame, using the pin-hole projection model:

$$\omega \begin{bmatrix} \tilde{s} \\ 1 \end{bmatrix} = \mathbf{KMP}_s, \quad (13)$$

where  $\omega$  is a scaling factor.

The coordinates of the corner point are then transformed into distorted image coordinates by (Fig. 8c):

$$\mathbf{s} = \begin{bmatrix} p_1(\tilde{\mathbf{s}}; \kappa) \\ p_2(\tilde{\mathbf{s}}; \kappa) \end{bmatrix}, \quad (14)$$

where  $p_1(\cdot)$  and  $p_2(\cdot)$  are the polynomials that describe radial and tangential distortion of the camera.

The estimated depth at  $\mathbf{s}$ , namely  $z^*$ , is found by means of the focus-based depth estimation method presented in the next section. Finally, the error  $e$  between the estimated depth  $z^*$  and the depth provided by the sensor  $z_s$  is computed. This process is repeated for each corner of the test target, as summarized in Algorithm 1.

#### 5.1. Performance measures

In order to assess the accuracy of the proposed model, as well as the suitability for its use in depth estimation, two different quantitative performance measures have been considered: root mean squared error (RMSE) and Pearson's correlation coefficient ( $r$ ).

The RMSE is used as an objective metric error measure and is simply computed as:

$$\text{RMSE} = \left( \frac{1}{n} \sum_{i=1}^N (z_i - z_i^*)^2 \right)^{0.5} \quad (15)$$

where  $n$  is the total number of measurements in the experiment,  $z^*$  is the depth estimate obtained using the proposed method and  $z$  is the depth value for comparison. Depending on the experiment,  $z$  can be a reference depth value, as described in section 4.2, or a direct measurement from the TOF sensor.

## 6. Experiments and results

For all the experiments, a Lytro Illum camera was used and raw lightfield images were decoded into  $P \times P$  views

<sup>5</sup><https://www.lytro.com/imaging>

<sup>6</sup><http://pmdtec.com/picofamily/>

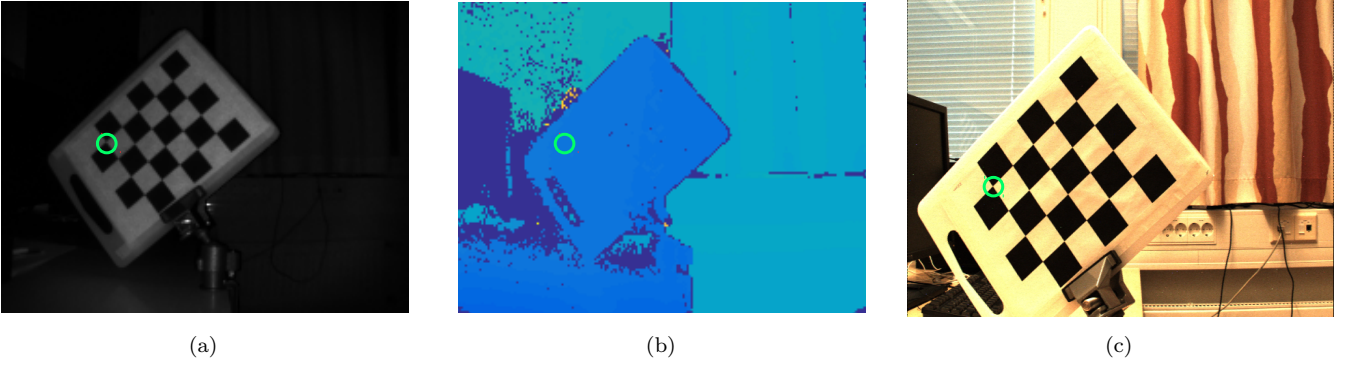


Figure 8: Validation methodology. (a) A corner is detected in the sensor frame. (b) 3D coordinates measured by the sensor are given by  $\mathbf{P}_s = [x_s, y_s, z_s]$ . (c) This point is projected in order to find the 2D coordinates  $\mathbf{s}$  in camera frame. The estimated depth  $z^*$  corresponding to  $\mathbf{s}$  is found using (11) and is compared with  $z_s$  in order to compute the error. Test targets are excluded from calibration of the model using leave-one-out validation.

(with  $P = 15$ ), as described in Section 3.1. We conducted experiments in order to assess the accuracy of the proposed focus model and calibration method in Section 6.1. We also analyze the performance of the proposed model in the presence of noise and distortion in Sections 6.2 and 6.3. Notice that the development of methods for depth estimation is out of the scope of this work. However, experiments on depth estimation based on the focus cue and its comparison with methods in the state-of-the-art are provided in the supplemental material.

### 6.1. Model accuracy

In order to assess the accuracy of the model in (8), a calibration dataset was captured using a planar target at 40 different orientations. Two images were excluded due to inaccurate detection of the corners and therefore only 38 images were considered in the subsequent experiments. The calibration target consisted of a checkerboard pattern with 20 corners and a square size of 38 mm. Due to limitations on the distance range of the TOF device used for validation, the calibration range was restricted between 0.2 to 1.6 m. The calibration of the camera was performed as described in section 4.1, thus obtaining  $760 = 20 \times 38$  reference depth values: one for every corner of each pose of the calibration target. In order to facilitate subsequent experiments, the calibration images were simultaneously captured by the TOF sensor in order to register the range data of the sensor with the viewpoint of the plenoptic camera using 3D-to-2D pose registration, as described in section 5.

We found the parameters of the model ( $a_0$ ,  $a_1$  and  $z_0$ ) using leave-one-out validation with the proposed calibration method. In each fold of the validation, the depth reference values corresponding to 37 calibration images are used for fitting the model. The reference values corresponding to the remaining image are then used for testing by comparing with depth values predicted by the model. This process is then repeated 37 more times (one for each

calibration image available). This validation strategy has been preferred for several reasons: 1) it allows a fair test in which the fitted model is used to predict the depth of “unseen” reference values, 2) it allows for an exhaustive validation throughout the whole calibration range, and 3) it allows performing a statistical analysis of the results.

Fig. 9a illustrates the fitting process of the model in one of the validation folds. This figure shows that the proposed model accurately describes the real behavior of the focus distance  $z$  as a function of the refocusing parameter  $\rho$ . The average absolute error between the adjusted model and the reference depth values was 3.3 ( $\pm 2.5$ ) cm. Fig. 9b shows the scatter plot between test reference values and predicted depth values across all the folds of the validation process. Predicted and reference depth values are highly consistent with a Pearson’s correlation coefficient of  $r = 0.99$  (95% CI: 0.98-0.99). A highly linear relationship between the model and metric depth values is of great importance since this demonstrates how well the model conforms to the real physical behavior of the camera and allows for removing systematic errors in subsequent depth estimation tasks.

A key step in the proposed method is the estimation of the reference depth values  $z_{i,j}$ . This is in fact what allows to perform the whole calibration process without the need of external sensors -recall that the TOF sensor is only used for validation purposes. In our experiments, the RMSE between  $z_{i,j}$  and  $z_{TOF}$  was 18.4 mm, which suggests that the estimation of depth reference values is accurate enough for the application.

The experiments described previously validate the proposed model under the assumption that standard camera calibration can be used for finding the intrinsic parameters of the central view of a plenoptic camera. In order to relax this assumption, the experiments described previously are repeated but, instead of comparing depth estimates against depth reference values found by standard camera calibration, depth measurements of the TOF sensor are

used as described in Section 5. As shown in Fig. 9c, comparisons with depth measurements obtained from the TOF sensor show an average absolute error of 4.7 ( $\pm 2.9$ ) cm and correlation coefficient of  $r = 0.99$  (95% CI: 0.98-0.99).

### 6.2. Robustness analysis

There are two main potential error sources in the calibration process: the estimation of the reference depth values  $z_{i,j}$  used for calibration, and the estimation of the refocusing parameter  $\rho_{i,j}$  by blur matching. It is clear that an erroneous estimation of either of these variables will affect the non-linear LSQ fitting of the model in (12).

The stability of the proposed calibration method to the aforementioned error sources is assessed empirically by adding noise to the depth reference and refocusing parameters. Specifically, the noisy values are computed as:

$$\hat{\rho}_{i,j} = \rho_{i,j} + \eta e(\rho_r) \quad (16)$$

and

$$\hat{z}_{i,j} = z_{i,j} + \eta e(z_r), \quad (17)$$

where  $\hat{z}_{i,j}$  and  $\hat{\rho}_{i,j}$  are the noisy depth reference and refocus parameter, respectively;  $\eta \in [0, 50]\%$  is the amount of noise;  $e(x)$  is a random variable with a uniform distribution with values in the range  $[-0.5x, 0.5x]$ ; and  $z_r$  and  $\rho_r$  are the range of the depth reference values, and the refocusing parameter, respectively.

The effect of errors on the reference depth values and the refocusing parameter are assessed independently. Specifically, the effect of errors on the reference depth values are assessed by fitting the curve  $\hat{z}_{i,j}$  vs.  $z(\rho_{i,j})$  in each fold of the validation process. In turn, the effect of errors in the estimated refocusing parameter are assessed by fitting the curve  $z_{i,j}$  vs.  $z(\hat{\rho}_{i,j})$ . The experiment is repeated 30 times and the average RMSE error is computed. Obtained results are summarized in Fig. 10. As shown in Fig. 10a, the RMSE of the predicted depth values increases for increasing noise levels in the refocus parameter. In contrast, a more robust result is observed for errors in the reference depth values (Fig. 10b). Notice that, strictly, the depth reference values do not correspond to actual measurements but they are estimated as part of the calibration process. This experiment help to better understand the robustness of the proposed approach. Remarkably, the calibration methods seems to perform robustly in the presence of errors with the RMSE reaching a maximum value of 14 cm for a 50% error in the refocusing parameter.

### 6.3. Effect of image distortion

The decoding strategy presented in Section 3.1 assumes that distortion parameters (i.e., tangential and radial distortion) of the central view can be used for decoding the rest of angular views. As a result, the non-linear effect of image distortion across the field of view of the camera can be a source of error in the calibration process. In order to

assess this effect, we computed the error between depth estimates using the proposed method and depth values measured by the TOF sensor as a function of the projection of the 760 3D corner points on the image field. In Fig. 11a, each dot represents the projection of one 3D point on the image plane of the plenoptic camera. The red circles represent the magnitude of the error between depth estimates and the measured depth values. From this figure, the error magnitude does not seem to follow a specific trend since large and small errors are distributed across the field of view without following a specific pattern. Fig. 11b shows box plots of the distribution of the error magnitude for different distances from the image center. An increase in the error could be expected for larger distances to the center. However, we could not verify this experimentally. Two-sample t-test showed no statistically significant difference between any of the distance ranges considered, thus confirming the observations made on Fig. 11a.

## 7. Discussion

Previous work has consistently shown that the focus cue can be successfully exploited for depth estimation in conventional digital cameras [19, 34, 18]. In plenoptic cameras, some researchers have used the focus cue, jointly with stereopsis or other cues, in order to obtain non-metric depth estimation [12, 25, 20]. In the literature, several researchers have tackled the problem of disparity-based calibration and depth estimation in focused plenoptic cameras [7, 9, 2, 8]. Notwithstanding, results on metric depth estimation in standard plenoptic cameras are scarce. Recently, a method for the geometric calibration of standard plenoptic cameras with disparity-based depth estimation has been presented in [15].

In this work, we proposed a model that relates metric depth units with the refocusing parameter of a standard plenoptic camera. One of the main advantages of the proposed method is that it can be readily calibrated by capturing a set of planar checkerboard patterns at arbitrary orientations. Although the calibration process has been specifically developed for standard plenoptic cameras, from the user perspective, this procedure is very similar to the calibration of classic digital cameras which will hopefully facilitate reproducibility.

### Calibration accuracy

The proposed calibration model was tested using a commercially available Lytro camera. Experiments in a distance range between 0.2 m and 1.6 m yielded an average depth estimation error of 3.3 cm. When using an external TOF sensor for comparison, the average depth estimation error rises to 4.7 cm in the same distance range. Note, however, that this error estimate includes the uncertainties due to the TOF sensor in the measurements. Specifically, in this case, there are two sources of uncertainty when using depth measures from the TOF sensor as ground-truth:

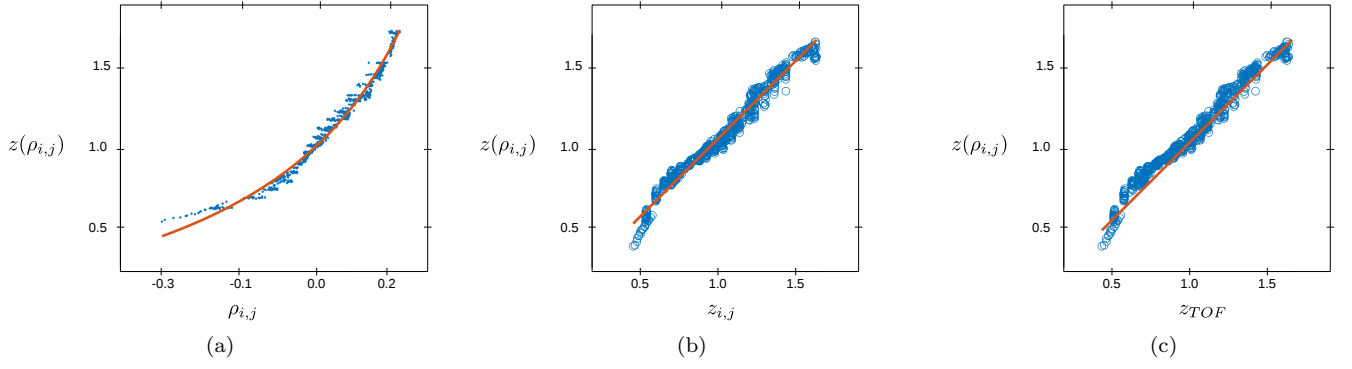


Figure 9: Model accuracy. (a) Fitting the focus model (continuous line) to measurements (dots). (b) Scatter plot between measured depth reference values  $z_{i,j}$  and estimated depth  $z(\rho_{i,j})$ . (c) Scatter plot between TOF depth values  $z_{TOF}$  and estimated depth  $z(\rho_{i,j})$ . All units are in meters.

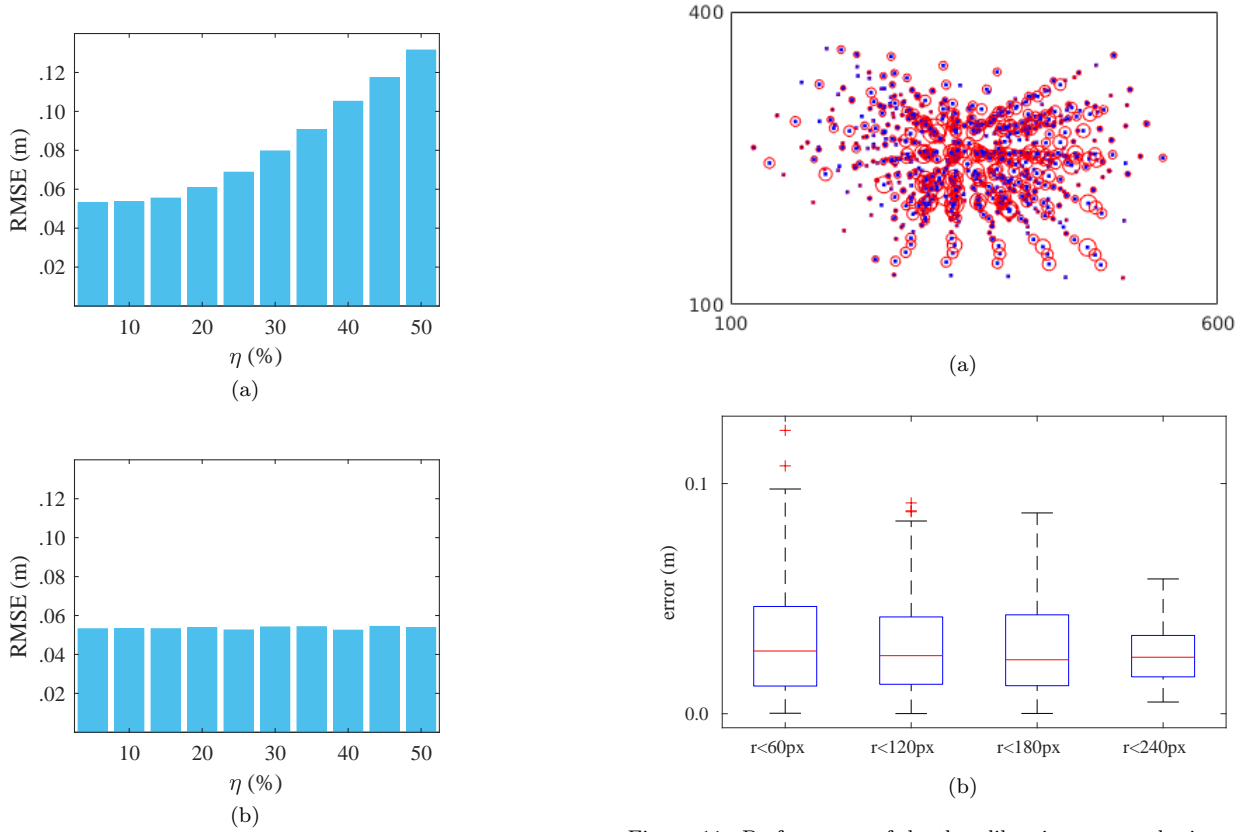


Figure 10: Empirical stability analysis. (a) RMSE as a function of the noise level in the refocus parameter. (b) RMSE as a function of the noise level in the depth reference values

Figure 11: Performance of depth calibration across the image field of view. (a) Depth estimation error for different locations in the image field. For visualization, the radius of the circles representing the error magnitude have been scaled by a factor of 100. (b) Error box plots for different distances from the image center



the inherent uncertainty of the depth sensor, and the error between measurements by the TOF sensor and reference depth values found by multi-view stereo calibration (section 4.1). Determining the accuracy of classical multi-view stereo calibration methods, as well as estimating the precision of the TOF sensor is beyond the scope of this work, but results should be interpreted accordingly.

Experiments conducted to assess the robustness of the proposed calibration method showed that the system is stable. Specifically, as shown in Fig. 10, the depth error rises smoothly with increasing levels of noise in the estimated refocus parameter. The experimental evidence provided here suggest that a simple low-order polynomial models could be used to empirically adjust the physical behavior of the real system.

### Limitations

In the validation, the precision of error estimates is highly dependent on the accuracy of the ground truth. For the future work, more accurate ground truths could be considered in order to better characterize the limitations of the proposed approach. Although the results in Fig. 11 suggest that the accuracy of the model is constant through the whole field of view, potential errors due to non-linear distortion effects of the main lens of the camera and the deviations on the projection centers of micro-lenses cannot be discarded. Recent advances in the geometrical calibration of standard plenoptic cameras could be utilized to study this issue [15].

Due to the limitations in the performance of focus-based depth estimation methods in the lack of texture, and the limited depth range of the TOF sensor used for validation, the scenes considered here were captured in controlled scenarios with limited depth range. Currently, the test dataset is being expanded by including more challenging scenarios. Specifically, real scenes with large distance ranges, specular surfaces, and low-textured scenes are being used in order to identify limitations of depth estimation methods for plenoptic cameras. The results and data used in this work will hopefully serve as a baseline for future developments in metric depth estimation using standard plenoptic cameras. In order to foster research on metric depth estimation in standard plenoptic cameras, source codes and all images with depth ground-truth will be publicly available as the TOF-LF dataset<sup>7</sup>.

## 8. Conclusions

In this paper, we tackled the problem of metric, focus-based depth estimation in plenoptic cameras. For this purpose, we have proposed a closed-form model for the refocusing distance of a standard plenoptic camera,  $z$ , as a function of the refocusing parameter,  $\rho$ . The model was derived based on a first-order optical analysis of the image

formation process of the camera. Based on the proposed model, we introduce a calibration method for finding the parameters of the model. This calibration method has the advantage that does not require knowledge of the internal parameters of the camera. In addition, it can be conveniently performed without need of external sensors by only using a simple planar checkerboard pattern.

The proposed model and calibration methods were tested in metric depth estimation in real datasets in distance range of 0.2-1.6 m. Obtained results show an average RMSE error of approximately 5 cm. Experimental results show for the first time that metric depth estimation based on focus in standard plenoptic cameras is feasible. We hope that the proposed model and calibration methods can foster research on metric depth estimation in this type of cameras.

## Appendix

### Properties of proposed focus model

The proposed focus model that relates the refocusing distance  $z$  with the refocusing parameter  $\rho$  is given by:

$$z(\rho) = z_0 \left( \frac{1 - a_0 \rho}{1 - a_1 \rho} \right), \quad (18)$$

where  $a_0 = \beta D / \mu x_0$ ,  $a_1 = z_0 a_0 / f$  and  $z_0$  is the focus at which the image was captured.

Based on (18) the following properties are satisfied:

- **Property 1:** The refocusing distance  $z$  is equal to the focus of the camera  $z_0$  if  $\rho = 0$ :

$$z = z_0 \Leftrightarrow \rho = 0$$

**Demonstration:** By direct substitution of  $\rho = 0$  in (18).

- **Property 2:** The refocusing distance  $z$  is greater than  $z_0$  for  $\rho > 0$ .

$$z > z_0 \Leftrightarrow \rho > 0$$

**Demonstration:** From (18) it is clear that  $z > z_0$  if  $1 - a_0 \rho > 1 - a_1 \rho$  for  $\rho > 0$ . Therefore,  $z > z_0$  if:

$$a_1 > a_0 \quad (19)$$

Since  $a_1 = z_0 a_0 / f$ , then (19) is satisfied if  $z_0 > f$  is assumed. In practice, this assumption is valid since, due to mechanical limitations in the lens-sensor systems, the minimum focus distance of a camera increases for larger focal lengths. Therefore, property 2 holds for  $z_0 > f$ .

- **Property 3:** The refocusing distance  $z$  is lower than  $z_0$  for  $\rho < 0$ .

$$z < z_0 \Leftrightarrow \rho < 0$$

<sup>7</sup><https://sites.google.com/view/cvia/plenoptic-imaging>

**Demonstration:** From (18) it is clear that  $z < z_0$  if  $1 - a_0\rho < 1 - a_1\rho$ . However, by substituting  $\rho$  by  $-\rho$ , the same result as in (19) is reached. Similarly as in property 2, property 3 is valid for focusing distances greater than the focal length of the camera.

- **Property 4:** There exists a finite value  $\rho_{\max}$  for which  $z(\rho) \rightarrow \infty$  that corresponds to a focal plane located at  $x = f$ :

$$\lim_{\rho \rightarrow \rho_{\max}} z(\rho) = \infty \Leftrightarrow x = f$$

**Demonstration:** From equation (18), it is straightforward to show that there exist a singular point at  $1 - a_1\rho = 0$ . Therefore,  $z(\rho_{\max}) = \infty$  for:

$$\rho_{\max} = \frac{1}{a_1} = \frac{f\mu x_0}{\beta D z_0} \quad (20)$$

By definition,  $\rho = \mu\Delta x/\beta D$ . Therefore:

$$\Delta x = \frac{\beta D}{\mu} \rho \quad (21)$$

By replacing (20) in (21), we get:

$$\Delta x = f \frac{x_0}{z_0} \quad (22)$$

From the thin-lens equation, by replacing  $z_0 = x_0 f/(x_0 - f)$  in (22), we get:

$$\Delta x = x_0 - f$$

Since  $x = x_0 + \Delta x$ , we have therefore demonstrated that  $x = f$  and  $z = \infty$  for  $\rho = \rho_{\max}$ .

## Acknowledgements

The authors would like to thank Ionut Schiopu and Olli Suominen of the Centre for Immersive Visual Technologies (CIVIT) at Tampere University of Technology for their advice and help in the selection and use of the most suitable TOF sensor for this work, as well as for granting us access to their facilities for performing some experiments under controlled scenarios. The authors gratefully acknowledge the Vicerrectoría de Investigación y Extensión de Universidad Industrial de Santander for supporting this research registered under the project title Movement Analysis in Compressive Multi-Spectral Video Imaging, (VIE code 2347). E. Pulido-Herrera would like to thank Antonio Nariño University for the financial support (project number 2017224).

## References

- [1] A. Lumsdaine, T. Georgiev, The focused plenoptic camera, in: IEEE International Conference on Computational Photography, 2009, pp. 1–8. doi:10.1109/ICCPHOT.2009.5559008.
- [2] K. H. Strobl, M. Lingenauber, Stepwise calibration of focused plenoptic cameras, Computer Vision and Image Understanding 145 (2016) 140 – 147, light Field for Computer Vision. doi:10.1016/j.cviu.2015.12.010.
- [3] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, J. Tumblin, Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing, ACM Transactions on Graphics 26 (3). doi:10.1145/1276377.1276463.
- [4] K. Marwah, G. Wetzstein, Y. Bando, R. Raskar, Compressive light field photography using overcomplete dictionaries and optimized projections, ACM Transactions on Graphics 32 (4) (2013) 46:1–46:12. doi:10.1145/2461912.2461914.
- [5] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, P. Hanrahan, Light field photography with a hand-held plenoptic camera, Computer Science Tech Report 02, Stanford University (2005).
- [6] G. Wetzstein, I. Ihrke, D. Lanman, W. Heidrich, Computational plenoptic imaging, Computer Graphics Forum 30 (8) (2011) 2397–2426. doi:10.1111/j.1467-8659.2011.02073.x.
- [7] O. Johannsen, C. Heinze, B. Goldluecke, C. Perwaß, On the Calibration of Focused Plenoptic Cameras, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 302–317. doi:10.1007/978-3-642-44964-2\_15.
- [8] N. Zeller, F. Quint, U. Stilla, Depth estimation and camera calibration of a focused plenoptic camera for visual odometry, ISPRS Journal of Photogrammetry and Remote Sensing 118 (2016) 83 – 100. doi:10.1016/j.isprsjprs.2016.04.010.
- [9] C. Heinze, S. Spyropoulos, S. Hussmann, C. Perwass, Automated robust metric calibration algorithm for multifocus plenoptic cameras, IEEE Transactions on Instrumentation and Measurement 65 (5) (2016) 1197–1205. doi:10.1109/TIM.2015.2507412.
- [10] C. Hahne, A. Aggoun, S. Haxha, V. Velisavljevic, J. C. J. Fernández, Light field geometry of a standard plenoptic camera, Optics Express 22 (22) (2014) 26659–26673. doi:10.1364/OE.22.026659.
- [11] C. Hahne, A. Aggoun, V. Velisavljevic, The refocusing distance of a standard plenoptic photograph, in: 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2015, pp. 1–4. doi:10.1109/3DTV.2015.7169363.
- [12] M. W. Tao, S. Hadap, J. Malik, R. Ramamoorthi, Depth from combining defocus and correspondence using light-field cameras, in: IEEE International Conference on Computer Vision, 2013, pp. 673–680. doi:10.1109/ICCV.2013.89.
- [13] S. Pertuz, M. A. Garcia, D. Puig, Efficient focus sampling through depth-of-field calibration, International Journal of Computer Vision 112 (3) (2015) 342–353. doi:10.1007/s11263-014-0770-0.
- [14] Z. Zhang, Flexible camera calibration by viewing a plane from unknown orientations, in: IEEE International Conference on Computer Vision, Vol. 1, 1999, pp. 666–673 vol.1. doi:10.1109/ICCV.1999.791289.
- [15] Y. Bok, H. G. Jeon, I. S. Kweon, Geometric calibration of micro-lens-based light field cameras using line features, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2) (2017) 287–300. doi:10.1109/TPAMI.2016.2541145.
- [16] A. Criminisi, S. B. Kang, R. Srinivasan, R. Szeliski, P. Anandan, Extracting layers and analyzing their specular properties using epipolar-plane-image analysis, Computer Vision and Image Understanding 97 (1) (2005) 51–85. doi:10.1016/j.cviu.2004.06.001.
- [17] S. Wanner, B. Goldluecke, Globally consistent depth labeling of 4d light fields, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 41–48. doi:10.1109/CVPR.2012.6247656.
- [18] M. Subbarao, G. Surya, Depth from defocus: A spatial do-

- main approach, *International Journal of Computer Vision* 13 (3) (1994) 271–294. doi:10.1007/BF02028349.
- [19] S. K. Nayar, Y. Nakagawa, Shape from focus, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (8) (1994) 824–831. doi:10.1109/34.308479.
  - [20] H. Lin, C. Chen, S. B. Kang, J. Yu, Depth recovery from light field using focal stack symmetry, in: *IEEE International Conference on Computer Vision*, 2015, pp. 3451–3459. doi:10.1109/ICCV.2015.394.
  - [21] T. E. Bishop, P. Favaro, The light field camera: Extended depth of field, aliasing, and superresolution, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (5) (2012) 972–986. doi:10.1109/TPAMI.2011.168.
  - [22] N. Sabater, M. Seifi, V. Drazic, G. Sandri, P. Pérez, Accurate Disparity Estimation for Plenoptic Images, Springer International Publishing, Cham, 2015, pp. 548–560. doi:10.1007/978-3-319-16181-5\_42.
  - [23] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, I. S. Kweon, Accurate depth map estimation from a lenslet light field camera, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1547–1555. doi:10.1109/CVPR.2015.7298762.
  - [24] T. C. Wang, A. A. Efros, R. Ramamoorthi, Depth estimation with occlusion modeling using light-field cameras, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (11) (2016) 2170–2181. doi:10.1109/TPAMI.2016.2515615.
  - [25] M. W. Tao, P. P. Srinivasan, S. Hadap, S. Rusinkiewicz, J. Malik, R. Ramamoorthi, Shape estimation from shading, defocus, and correspondence using light-field angular coherence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (3) (2017) 546–560. doi:10.1109/TPAMI.2016.2554121.
  - [26] E. H. Adelson, J. R. Bergen, The plenoptic function and the elements of early vision, in: *Computational Models of Visual Processing*, MIT Press, 1991, pp. 3–20.
  - [27] M. Levoy, P. Hanrahan, Light field rendering, in: *Annual Conference on Computer Graphics and Interactive Techniques*, ACM, 1996, pp. 31–42. doi:10.1145/237170.237199.
  - [28] D. G. Dansereau, Plenoptic signal processing for robust vision in field robotics, Ph.D. thesis, University of Sydney (2014).
  - [29] D. G. Dansereau, O. Pizarro, S. B. Williams, Decoding, calibration and rectification for lenselet-based plenoptic cameras, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1027–1034. doi:10.1109/CVPR.2013.137.
  - [30] R. Ng, Digital lightfield photography, Ph.D. thesis, Stanford University (2005).
  - [31] J. Y. Bouguet, Camera Calibration Toolbox for Matlab, online (10 2015). URL [http://www.vision.caltech.edu/bouguetj/calib\\_doc/index.html](http://www.vision.caltech.edu/bouguetj/calib_doc/index.html)
  - [32] P. Torr, A. Zisserman, Mlesac: A new robust estimator with application to estimating image geometry, *Computer Vision and Image Understanding* 78 (1) (2000) 138 – 156. doi:http://dx.doi.org/10.1006/cviu.1999.0832.
  - [33] X.-S. Gao, X.-R. Hou, J. Tang, H.-F. Cheng, Complete solution classification for the perspective-three-point problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (8) (2003) 930–943. doi:10.1109/TPAMI.2003.1217599.
  - [34] A. P. Pentland, A new sense for depth of field, *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9* (4) (1987) 523–531. doi:10.1109/TPAMI.1987.4767940.