METRIC CALIBRATION OF A FOCUSED PLENOPTIC CAMERA BASED ON A 3D CALIBRATION TARGET

N. Zeller[a, c], C.A. Noury[b, d], F. Quint[a], C. Teulière[b, d], U. Stilla[c], M. Dhome[b, d]

[a] Karlsruhe University of Applied Sciences, 76133 Karlsruhe, Germany - niclas.zeller@hs-karlsruhe.de; franz.quint@hs-karlsruhe.de
[b] Université Clermont Auvergne, Université Blaise Pascal, Institut Pascal, BP 10448, F-63000 Clermont-Ferrand, France - charles_antoine.noury@univ-bpclermont.fr; celine.teuliere@univ-bpclermont.fr; michel.dhome@univ-bpclermont.fr
[c] Technische Universität München, Germany, 80290 Munich, Germany - stilla@tum.de
[d] CNRS, UMR 6602, IP, F-63178 Aubière, France

ICWG III/I

KEY WORDS: Bundle adjustment, depth accuracy, depth distortion model, focused plenoptic camera, metric camera calibration

ABSTRACT:

In this paper we present a new calibration approach for focused plenoptic cameras. We derive a new mathematical projection model of a focused plenoptic camera which considers lateral as well as depth distortion. Therefore, we derive a new depth distortion model directly from the theory of depth estimation in a focused plenoptic camera. In total the model consists of five intrinsic parameters, the parameters for radial and tangential distortion in the image plane and two new depth distortion parameters. In the proposed calibration we perform a complete bundle adjustment based on a 3D calibration target. The residual of our optimization approach is three dimensional, where the depth residual is defined by a scaled version of the inverse virtual depth difference and thus conforms well to the measured data. Our method is evaluated based on different camera setups and shows good accuracy. For a better characterization of our approach we evaluate the accuracy of virtual image points projected back to 3D space.

1. INTRODUCTION

In the last years light-field (or plenoptic) cameras became more and more popular. One reason therefor is the availability of such camera systems on the market.

Plenoptic cameras capture, different from regular cameras, not only a 2D image of a scene, but a complete 4D light-field representation (Adelson and Wang, 1992, Gortler et al., 1996). Due to this additional information plenoptic cameras find usage for a variety of applications in photogrammetry as well as computer vision. Here, plenoptic cameras replace for example other depth sensors like stereo camera systems.

To use such a depth sensor for instance in photogrammetric applications a precise metric calibration is mandatory. While the calibration of monocular cameras and stereo camera systems has been studied over the last decades, there is no overall accepted mathematical model available for plenoptic cameras yet. Therefore, this paper introduces a new model for focused plenoptic cameras (Lumsdaine and Georgiev, 2009, Perwaß and Wietzke, 2012) and presents how this model can be precisely determined in a metric calibration process using a 3D calibration target.

1.1 Related Work

During the last years different methods where published to calibrate a plenoptic camera. This section lists calibration methods for unfocused plenoptic cameras (Adelson and Wang, 1992, Ng et al., 2005) and focused plenoptic cameras (Lumsdaine and Georgiev, 2009, Perwaß and Wietzke, 2012) separately.

Until today the unfocused plenoptic camera is mostly used in consumer or image processing applications where a precise metric relation between object and image space is not mandatorily needed. Ng and Hanrahan for instance presented a method for correcting aberrations of the main lens on the recorded 4D light-field inside the camera (Ng and Hanrahan, 2006).

In 2013 Dansereau et al. presented the first complete mathematical model of an unfocused plenoptic camera (Dansereau et al., 2013). Their model consists of 15 parameters which include, besides the projection from object space to the sensor, the micro lens array (MLA) orientation as well as a distortion model for the main lens.

Bok et al. proposed a method which does not use point features to solve for the calibration model but line features extracted from the micro images (Bok et al., 2014).

A first calibration method for focused plenoptic cameras was proposed by Johannsen et al. (Johannsen et al., 2013). They proposed a camera model tailored especially for Raytrix cameras which consists of 15 parameters. This model considers lateral distortion (image distortion) as well as depth distortion and shows reasonable results for the evaluated distances from 36 cm to 50 cm.

The method of Johannsen et al. was further developed by Heinze et al. and resulted in an automated metric calibration method (Heinze et al., 2015). In their method calibration is performed based on a planar object which is moved freely in front of the camera.

In a previous work, Zeller et al. focused on developing efficient calibration methods for larger object distances ($> 50\,\text{cm}$) (Zeller et al., 2014). They present three different models to define the relationship between object distance and the virtual depth which is estimated based on the light-field. Besides, the calibration process is split into two parts, where the optical path and the depth are handled separately.

1.2 Contribution of this Work

This paper proposes a new model of a focused plenoptic camera which considers lateral distortion of the intensity image as

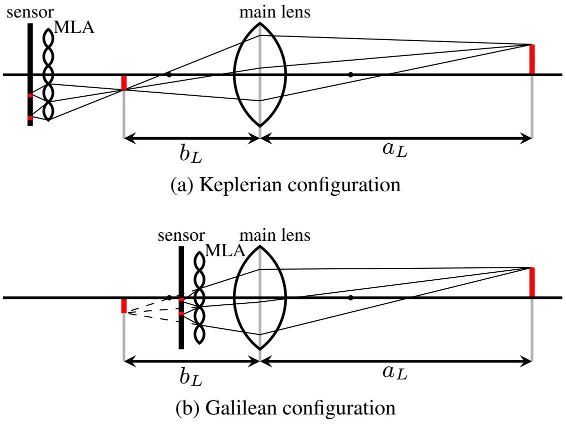


Figure 1. Cross view of the two possible configurations (Keplerian and Galilean) of a focused plenoptic camera.

well as virtual depth distortion. While the model shows similarity with respect to the one in (Heinze et al., 2015) some significant changes were made which adapt the model better to the physical camera.

Firstly, we consider lateral distortion after projecting all virtual image points, which are the image points created by the main lens, back to a mutual image plane, along their central ray. In contrast, Heinze et al. apply lateral distortion directly on the virtual image. Thereby, we receive a model for lateral distortion which conforms better to the reality since the virtual image is not formed on a plane but on an undulating surface defined by the image distance (distance to the main lens) of each point. Besides, we show that the depth distortion can be derived from the lateral distortion model since in a plenoptic camera the depth map is based on disparity estimation in the raw image.

We are using a 3D target in order to obtain precise calibration parameters by performing a bundle adjustment on feature points.

In our optimization approach the depth residual is defined by a scaled version of the inverse virtual depth difference which conforms better to the measured data.

In contrast to previous methods we show that our calibration gives good results up to object distances of a few meters. In addition, we evaluate the accuracy of points projected back to object space in 3D, which has not been done in any of the previous publications.

The main part of this paper is structured as follows. Section 2 briefly presents the concept of a focused plenoptic camera. The camera model derived based on this concept is presented in Section 3. In Section 4 we formulate the non-linear problem which is optimized to obtain the plenoptic camera model. Section 5 presents the complete calibration workflow. Section 6 evaluates the calibration based on real data and Section 7 draws conclusion.

2. THE FOCUSED PLENOPTIC CAMERA

Even though there exist different concepts of MLA based plenoptic camera (Ng et al., 2005, Adelson and Wang, 1992) we will focus here only on the concept of a focused plenoptic camera (Lumsdaine and Georgiev, 2009).

As proposed by (Lumsdaine and Georgiev, 2008, Lumsdaine and Georgiev, 2009) a focused plenoptic camera can be set up in two different configurations. The Keplerian and the Galilean configuration (Fig. 1). While in the Keplerian configuration MLA and sensor are placed behind the focused image created by the main lens (Fig. 1a), in the Galilean configuration MLA and sensor are placed in front of the focused main lens image (Fig. 1b). Since in the Galilean configuration the main lens image exists only as a virtual image, we will call it as such in the following.

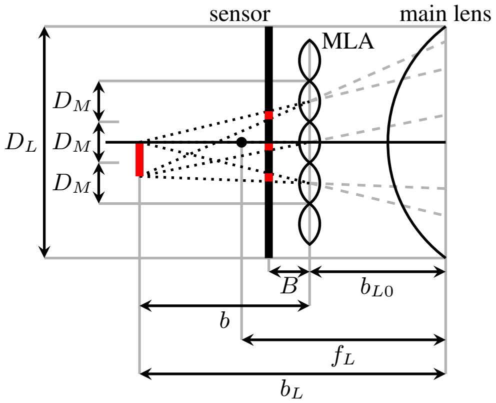


Figure 2. Optical path inside a focused plenoptic camera based on the Galilean configuration.

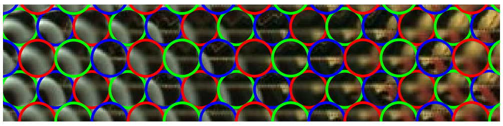


Figure 3. Section of the micro lens images (raw image) of a Raytrix camera. Different micro lens types are marked by different colors.

The camera model presented in this paper was developed based on a Raytrix camera, which is a plenoptic camera in the Galilean configuration. Nevertheless, by slightly adapting the model proposed in Section 3 a similar calibration can be applied to cameras in the Keplerian configuration.

An image point in a micro image is composed only by a sub-bundle of all rays tracing through the main lens aperture. Consequently the micro images have already a larger depth of field (DOF) than a regular camera with the same aperture. In a Raytrix camera the DOF is further increased by using an interlaced MLA in a hexagonal arrangement (see Fig. 3). This MLA consists of three different micro lens types, where each type has as different focal length and thus focuses a different virtual image distance (resp. object distance) on the sensor. The DOFs of the three micro lens types are chosen such that they are just adjacent to each other. Thus, the effective DOF of the camera is increased compared to an MLA with only one type of micro lenses. In the sensor image shown in Figure 3 the blue micro images are in focus while the green and red ones are blurred. Since each virtual image point is projected to multiple micro images it can be assured that each point is recorded in focus at least once and thus a complete focused image can be reconstructed. This also can be seen from Figure 3, where similar structures occur in adjacent micro images. For more details we refer to (Perwaß and Wietzke, 2012).

In the following part of this section we will discuss the projection process in a focused plenoptic camera based on the Galilean configuration. For the main lens a thin lens model is used, while each of the micro lenses is modeled by a pinhole.

It is well known that for an ideal thin lens the relation between object distance a_L and image distance b_L is defined dependent on the focal length f_L by the following equation:

$$\frac{1}{f_L} = \frac{1}{a_L} + \frac{1}{b_L} \tag{1}$$

Furthermore, for each point projected from object space to image space a so called virtual depth v can be estimated based on the disparity p_x of corresponding points in the micro images (Perwaß and Wietzke, 2012). Therefor various algorithms can be used (Zeller et al., 2015, Bishop and Favaro, 2009). The relationship between the virtual depth v and the estimated disparity p_x is given by the following equation:

$$v = \frac{b}{B} = \frac{d}{p_x} \tag{2}$$

As one can see, the virtual depth is just a scaled version of the distance b between a virtual image point p_V and the plane of the MLA. In this eq. (2) B is the distance between MLA and sensor and d defines the distance between the principal points of the respective micro lenses on the sensor. In general a virtual image point p_V is projected to more than two micro images (see Fig. 2) and thus multiple virtual depth observations with different baseline distances d are received. Here, d is a multiple of the micro lens aperture D_M ($d = k \cdot D_M$ with $k \geq 1$). Due to the 2D arrangement of the MLA the multiple k is not mandatory an integer.

3. CAMERA MODEL

This section presents the developed model for a focused plenoptic camera which will be used in the calibration process (Section 5).

The thin lens model and the pinhole model were used to represent the main lens and the micro lenses respectively. In reality of course the main lens does not satisfy the model of a thin lens. Nevertheless, this simplification does not effect the overall projection from object space to virtual image space.

To handle the imperfection of the main lens a distortion model for the virtual image space is defined. This model consists of lateral as well as depth distortion. One could argue that also the micro lenses add distortion to the projection. Nevertheless, since for the camera used in our research one micro lens has a very narrow field of view and only about 23 pixels in diameter, any distortion on the pixel coordinates generated by the micro lenses can be neglected.

3.1 Projection Model without Distortion

Unlike regular cameras, where a 3D object space is projected to a 2D image plane, in a focused plenoptic camera a 3D object space is projected to a 3D image space. This virtual image space is indirectly captured in the 4D light-field.

In the following derivation we will consider the object space to be aligned to the camera frame. Thus, a point in object space p_O is defined by its homogeneous 3D camera coordinates $\boldsymbol{X}_C = (x_C, y_C, z_C, 1)^T$. The camera coordinate system has its origin in the optical center of the main lens, while the z-axis is pointing towards the object ($z_C > 0$), the y-axis is pointing downwards and the x-axis is pointing to the right.

For the virtual image space we define a mirrored coordinate system, which means that all unit vectors are pointing in the opposite direction as those of the camera coordinate system. Here a virtual image point p_V is defined by the homogeneous 3D coordinates $\boldsymbol{X}_V = (x_V, y_V, z_V, 1)^T$. The virtual image coordinates have their origin on the intersection of the optical axis with the plane of the MLA. Both camera coordinates $\boldsymbol{X}_C$ as well as virtual image coordinates $\boldsymbol{X}_V$ have metric dimensions.

Using the ideal thin lens model for the main lens a virtual image point, defined by the coordinates $\boldsymbol{X}_V$, can be calculated based on the corresponding object point coordinates $\boldsymbol{X}_C$, as given in eq. (3).

$$\begin{gathered}\lambda \cdot \boldsymbol{X}_V = K \cdot \boldsymbol{X}_C \\ \lambda \cdot \begin{pmatrix} x_V \\ y_V \\ z_V \\ 1 \end{pmatrix} = \begin{pmatrix} b_L & 0 & 0 & 0 \\ 0 & b_L & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_C \\ y_C \\ z_C \\ 1 \end{pmatrix}\end{gathered} \tag{3}$$

In eq. (3) the matrix coefficients b_L and b, which conform to the ones in Figure 2, are dependent on the object distance $a_L = z_C$ and are defined as follows:

$$b_L = \left(\frac{1}{f_L} - \frac{1}{z_C}\right)^{-1} = B \cdot v + b_{L0} \tag{4}$$

$$b = b_L - b_{L0} \tag{5}$$

The introduced scaling factor λ is just the object distance z_C:

$$\lambda = z_C \tag{6}$$

To receive the virtual image coordinates in the dimensions they are measured $\boldsymbol{X}'_V = (x'_V, y'_V, v, 1)^T$, a scaling of the axis as well as a translation along the x- and y-axis has to be performed (x'_V and y'_V are defined in pixels), as defined in eq. (7).

$$\begin{gathered}\boldsymbol{X}'_V = K_S \cdot \boldsymbol{X}_V \\ \begin{pmatrix} x'_V \\ y'_V \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} s_x^{-1} & 0 & 0 & c_x \\ 0 & s_y^{-1} & 0 & c_y \\ 0 & 0 & B^{-1} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_V \\ y_V \\ z_V \\ 1 \end{pmatrix}\end{gathered} \tag{7}$$

Here s_x and s_y define the size of a pixel, while B is the distance between MLA and sensor.

By combining the matrix K and K_S one can define the complete transform from camera coordinates $\boldsymbol{X}_C$ to virtual image coordinates $\boldsymbol{X}'_V$, as given in eq. (8).

$$\lambda \cdot \boldsymbol{X}'_V = K_S \cdot K \cdot \boldsymbol{X}_C \tag{8}$$

Since the pixel size (s_x, s_y) for a certain sensor is known in general, there are five parameters left which have to be estimated. These parameters are the principal point (c_x, c_y) expressed in pixels, the distance between MLA and sensor B, the distance from optical main lens center to MLA b_{L0} and the main lens focal length f_L.

3.2 Distortion Model

Section 3.1 defines the projection from an object point p_O to the corresponding virtual image point p_V without considering any imperfection in the lens or the setup. In this section we will define a lens distortion model which corrects the position of a virtual image point p_V in x, y and z direction.

To implement the distortion model another coordinate system with the homogeneous coordinates $\boldsymbol{X}_I = (x_I, y_I, z_I, 1)$ is defined. The coordinates $\boldsymbol{X}_I$ actually define the pinhole projection which would be performed in a conventional camera. Here an image point p_I is defined as the intersection of a central ray through the main lens with an image plane. In the presented model the image plane is chosen to be the same as the MLA plane. Thus, by splitting up the matrix K into K_I and K_V (eq. (9)) the image coordinates $\boldsymbol{X}_I$ result as defined in eq. (10).

$$\begin{aligned} K &= K_V \cdot K_I \\ &= \begin{pmatrix} \frac{b_L}{b_{L0}} & 0 & 0 & 0 \\ 0 & \frac{b_L}{b_{L0}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} b_{L0} & 0 & 0 & 0 \\ 0 & b_{L0} & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{aligned} \tag{9}$$

$$\lambda \cdot \boldsymbol{X}_I = K_I \cdot \boldsymbol{X}_C \tag{10}$$

By the introduction of image coordinates $\boldsymbol{X}_I$ we are able to introduce a lateral distortion model on this image plane, similar as for a regular camera. From eq. (9) and (10) one can see that the z-coordinates in the spaces defined by $\boldsymbol{X}_I$ and $\boldsymbol{X}_V$ are equivalent ($z_I = z_V$).

In the following we define our distortion model, which is split into lateral and depth distortion.

3.2.1 Lateral Distortion

For lateral distortion a model which consists of radial symmetric as well as tangential distortion is defined. This is one of the most commonly used models in photogrammetry.

The radial symmetric distortion is defined by a polynomial of the variable r, as defined in eq. (11).

$$\Delta r_{rad} = A_0 r^3 + A_1 r^5 + A_2 r^7 + \cdots \tag{11}$$

Here r is the distance between the principal point and the coordinates on the image plane:

$$r = \sqrt{x_I^2 + y_I^2} \tag{12}$$

This results in the correction terms Δx_{rad} and Δy_{rad} as given in eq. (13) and (14).

$$\begin{aligned} \Delta x_{rad} &= x_I \frac{\Delta r_{rad}}{r} \\ &= x_I \cdot \left(A_0 r^2 + A_1 r^4 + A_2 r^6 + \cdots\right) \end{aligned} \tag{13}$$

$$\begin{aligned} \Delta y_{rad} &= y_I \frac{\Delta r_{rad}}{r} \\ &= y_I \cdot \left(A_0 r^2 + A_1 r^4 + A_2 r^6 + \cdots\right) \end{aligned} \tag{14}$$

In our implementation we use a radial distortion model consisting of three coefficients (A_0 to A_2). Besides, for tangential distortion we used the model defined in (Brown, 1966) using only the two first parameters B_0 and B_1. The terms given in eq. (15) and (16) are defined.

$$\Delta x_{tan} = B_0 \cdot \left(r^2 + 2x_I^2\right) + 2B_1 x_I y_I \tag{15}$$

$$\Delta y_{tan} = B_1 \cdot \left(r^2 + 2y_I^2\right) + 2B_0 x_I y_I \tag{16}$$

Based on the correction terms the distorted image coordinates x_{Id} and y_{Id} are calculated from the ideal projection as follows:

$$x_{Id} = x_I + \Delta x_{rad} + \Delta x_{tan} \tag{17}$$

$$y_{Id} = y_I + \Delta y_{rad} + \Delta y_{tan} \tag{18}$$

3.2.2 Depth Distortion

To describe the depth distortion a new model is defined. While other calibration methods define the depth distortion by just adding a polynomial based correction term (Johannsen et al., 2013, Heinze et al., 2015), our goal was to find an analytical expression describing the depth distortion as a function of the lateral distortion and thus reflecting the physical reality. Therefore, in the following the depth correction term Δv is derived from the relation between the virtual depth v and the corresponding estimated disparity p_x. In the following equations all parameters with subscript d refer to distorted parameters which are actually measured, while the parameters without a subscript are the undistorted ones resulting from the ideal projection.

Equation (19) defines, similar to eq. (2), the estimated virtual depth based on corresponding points $\boldsymbol{x}_{di}$ in the micro images and its micro image centers $\boldsymbol{c}_{di}$. Here the pixel coordinates as well as the micro lens centers are affected by the distortion.

$$v_d = \frac{d_d}{p_{xd}} = \frac{\|\boldsymbol{c}_{d2} - \boldsymbol{c}_{d1}\|}{\|\boldsymbol{x}_{d2} - \boldsymbol{x}_{d1} - (\boldsymbol{c}_{d2} - \boldsymbol{c}_{d1})\|} \tag{19}$$

In general both, $\boldsymbol{x}_{di}$ and $\boldsymbol{c}_{di}$ can be defined as pixel coordinates on the sensor (in the raw image).

Due to epipolar geometry and rectified micro images, the difference vectors $(\boldsymbol{c}_{d2} - \boldsymbol{c}_{d1})$ and $(\boldsymbol{x}_{d2} - \boldsymbol{x}_{d1})$ point always in the same direction. In addition, since eq. (19) is only defined for positive virtual depths, $\|\boldsymbol{c}_{d2} - \boldsymbol{c}_{d1}\| \geq \|\boldsymbol{x}_{d2} - \boldsymbol{x}_{d1}\|$ always holds. Therefore, eq. (19) can be simplified as follows:

$$v_d = \frac{\|\boldsymbol{c}_{d2} - \boldsymbol{c}_{d1}\|}{\|\boldsymbol{c}_{d2} - \boldsymbol{c}_{d1}\| - \|\boldsymbol{x}_{d2} - \boldsymbol{x}_{d1}\|} \tag{20}$$

Replacing the distorted coordinates by the sum of undistorted coordinates and their correction terms results in eq. (21). Here one can assume that a micro lens center $\boldsymbol{c}_i$ undergoes the same distortion $\Delta \boldsymbol{x}_i$ as the underlying image point $\boldsymbol{x}_i$ since both have similar coordinates.

$$v_d = \frac{\|\boldsymbol{c}_2 + \Delta\boldsymbol{x}_2 - \boldsymbol{c}_1 - \Delta\boldsymbol{x}_1\|}{\begin{aligned}\big(\|\boldsymbol{c}_2 + \Delta\boldsymbol{x}_2 - \boldsymbol{c}_1 - \Delta\boldsymbol{x}_1\| \\ - \|\boldsymbol{x}_2 + \Delta\boldsymbol{x}_2 - \boldsymbol{x}_1 - \Delta\boldsymbol{x}_1\|\big)\end{aligned}} \tag{21}$$

Under the assumption that only radial symmetric distortion is present and that $\frac{\|\Delta\boldsymbol{x}_2\|}{\|\boldsymbol{x}_2\|} \approx \frac{\|\Delta\boldsymbol{x}_1\|}{\|\boldsymbol{x}_1\|}$, as well as $\frac{\|\Delta\boldsymbol{x}_2\|}{\|\boldsymbol{c}_2\|} \approx \frac{\|\Delta\boldsymbol{x}_1\|}{\|\boldsymbol{c}_1\|}$ holds, eq. (21) can be simplified as follows:

$$\begin{aligned} v_d &\approx \frac{\|\boldsymbol{c}_2 - \boldsymbol{c}_1\| \pm \|\Delta\boldsymbol{x}_2 - \Delta\boldsymbol{x}_1\|}{\|\boldsymbol{c}_2 - \boldsymbol{c}_1\| - \|\boldsymbol{x}_2 - \boldsymbol{x}_1\|} \\ &= \frac{d}{p_x} \pm \frac{\|\Delta\boldsymbol{x}_2 - \Delta\boldsymbol{x}_1\|}{p_x} \end{aligned} \tag{22}$$

Here the ± considers the two cases that the vector $\Delta\boldsymbol{x}_2 - \Delta\boldsymbol{x}_1$ is pointing in the same direction as the vector $\boldsymbol{c}_2 - \boldsymbol{c}_1$ or in the opposite direction.

The assumption which was made above holds well for micro lenses which are far from the principal point. For close points the distortion anyway is small and thus can be neglected. Besides, radial symmetric distortion in general is dominant over other distortion terms.

From eq. (22) one obtains that the distorted virtual depth v_d can be defined as the sum of the undistorted virtual depth v and a correction term Δv, as defined in eq. (23).

$$\Delta v = \pm \frac{\|\Delta\boldsymbol{x}_2 - \Delta\boldsymbol{x}_1\|}{p_x} \tag{23}$$

Presuming that the vectors $\Delta\boldsymbol{x}_1$ and $\Delta\boldsymbol{x}_2$ are pointing in more or less the same direction, eq. (23) can be approximated as follows:

$$\Delta v = \pm\frac{\|\Delta \boldsymbol{x}_2 - \Delta \boldsymbol{x}_1\|}{p_x} \approx \frac{\Delta r_{rad}(r_2) - \Delta r_{rad}(r_1)}{p_x} \tag{24}$$

with $r_1 \approx r - \frac{d}{2}$ and $r_2 \approx r + \frac{d}{2}$. The assumption that $\Delta \boldsymbol{x}_2$ and $\Delta \boldsymbol{x}_1$ point in the same direction again holds well for large image coordinates $\boldsymbol{x}_i$. The radius r defines the distance from the principal point to the orthogonal projection of the virtual image point p_V on the sensor plane.

To simplify the depth distortion model only the first coefficient of the radial distortion A_0, defined in eq. (11), is considered to be significant. Thus, based on eq. (11) the following definition for Δv is received:

$$\begin{aligned}\Delta v &= \frac{A_0}{p_x}\left(\left(r+\frac{d}{2}\right)^3 - \left(r-\frac{d}{2}\right)^3\right)\\ &= \frac{A_0}{p_x}\left(3dr^2 + \frac{d^3}{4}\right)\end{aligned} \tag{25}$$

Replacing p_x by $\frac{d}{v}$ (see eq. (2)) results in the following equation:

$$\Delta v = \frac{A_0 \cdot v}{d}\left(3dr^2 + \frac{d^3}{4}\right) \quad = A_0 \cdot v\left(3r^2 + \frac{d^2}{4}\right) \tag{26}$$

For a virtual image point p_V with virtual depth v all micro lenses which see this point lie within a circle with diameter $D_M \cdot v$ around the orthogonal projection of the virtual image point p_V on the MLA plane. Thus, in first approximation the maximum baseline distance used for depth estimation is equivalent to this diameter. Therefore, we replace d by $D_M \cdot v$ which results in the final definition of our depth distortion model, as given in eq. (27).

$$\begin{aligned}\Delta v &= 3A_0 \cdot v \cdot r^2 + \frac{A_0 \cdot D_M^2}{4} \cdot v^3\\ &= D_0 \cdot v \cdot r^2 + D_1 \cdot v^3\end{aligned} \tag{27}$$

Since the distortion is just defined by additive terms they can be represented by a translation matrix K_D as defined in eq. (28).

$$K_D = \begin{pmatrix} 1 & 0 & 0 & \Delta x_{rad} + \Delta x_{tan} \\ 0 & 1 & 0 & \Delta y_{rad} + \Delta y_{tan} \\ 0 & 0 & 1 & B\Delta v \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{28}$$

This matrix consists of seven distortion parameters (A_0, A_1, A_2, B_0, B_1, D_0 and D_1) which have to be estimated in the calibration process.

3.3 Complete Camera Model

Under consideration of the projection as well as the distortion, the complete projection from an object point in camera coordinates $\boldsymbol{X}_C$ to the corresponding virtual image point in distorted coordinates $\boldsymbol{X}'_{Vd}$, which are actually measured, is defined as follows:

$$\lambda \cdot \boldsymbol{X}'_{Vd} = K_S \cdot K_V \cdot K_D \cdot K_I \cdot \boldsymbol{X}_C \tag{29}$$

The projection model defined in eq. (29) consists of 12 unknown intrinsic parameters which have to be estimated during calibration.

4. NON-LINEAR OPTIMIZATION PROBLEM

To estimate the camera model an optimization problem has to be define, which can be solved based on recorded reference points. We define this optimization problem as a bundle adjustment, where the intrinsic as well as extrinsic camera parameters are estimated based on multiple recordings of a 3D target.

The relation between an object point $p_O^{\{i\}}$ with homogeneous world coordinates $\boldsymbol{X}_W^{\{i\}}$ and the corresponding virtual image point in the j-th frame $p_V^{\{i,j\}}$ is defined based on the model presented in Section 3 as follows:

$$\boldsymbol{X}'^{\{i,j\}}_{Vd} = \frac{1}{\lambda_{ij}} \cdot K_S \cdot K_V \cdot K_D \cdot K_I \cdot G_j \cdot \boldsymbol{X}_W^{\{i\}} \tag{30}$$

Here $G_j \in SE(3)$ defines the rigid body transform (special Euclidean transform) from world coordinates $\boldsymbol{X}_W^{\{i\}}$ to the camera coordinates of the j-th frame $\boldsymbol{X}_C^{\{i,j\}}$. Besides, λ_{ij} is defined by $z_C^{\{i,j\}}$ (see eq. (6)).

In the following we will denote a virtual image point $\boldsymbol{X}'_{Vd}$ which was calculated based on the projection given in eq. (30) by the coordinates x_{proj}, y_{proj}, and v_{proj}, while we denote the corresponding reference point measured from the recorded image by the coordinates x_{meas}, y_{meas}, and v_{meas}.

Based on the projected as well as measured points the cost function given in eq. (31) can be defined.

$$C = \sum_{i=1}^{N}\sum_{j=1}^{M} \theta_{ij}\|\boldsymbol{\varepsilon}_{ij}\|^2 \tag{31}$$

Here N defines the number of 3D points in object space, while M is the number of frames used for calibration. The function θ_{ij} represents the visibility of an object point in a certain frame (see eq. (32)).

$$\theta_{ij} = \begin{cases} 1, & \text{if } p_O^{\{i\}} \text{ is visible in } j\text{-th image,} \\ 0, & \text{if not.} \end{cases} \tag{32}$$

The vector $\boldsymbol{\varepsilon}_{ij}$ is defined as given in the following equations:

$$\boldsymbol{\varepsilon}_{ij} = \begin{pmatrix} \varepsilon_x^{\{i,j\}} & \varepsilon_y^{\{i,j\}} & \varepsilon_v^{\{i,j\}} \end{pmatrix}^T \tag{33}$$

$$\varepsilon_x^{\{i,j\}} = x_{proj}^{\{i,j\}} - x_{meas}^{\{i,j\}} \tag{34}$$

$$\varepsilon_y^{\{i,j\}} = y_{proj}^{\{i,j\}} - y_{meas}^{\{i,j\}} \tag{35}$$

$$\varepsilon_v^{\{i,j\}} = \left(\left(v_{proj}^{\{i,j\}}\right)^{-1} - \left(v_{meas}^{\{i,j\}}\right)^{-1}\right) \cdot v_{proj}^{\{i,j\}} \cdot w \tag{36}$$

The residual of the virtual depth ε_v is defined as given since the virtual depth v is inverse proportional to the disparity p_x estimated from the micro images (see eq. (2)) and thus the inverse virtual depth can be considered to be Gaussian distributed (as analyzed in (Zeller et al., 2015)). As already stated in Section 3.2.2, the baseline distance d is on average proportional to the estimated virtual depth v. Thus, the difference in the inverse virtual depth is scaled by the virtual depth v itself. The parameter w is just a constant factor which defines the weight between ε_x, ε_y, and ε_v.

5. CALIBRATION PROCESS

To perform the calibration, we use a 3D calibration target, as shown in Figure 5a. The complete calibration process is visualized in the flow chart shown in Figure 4.

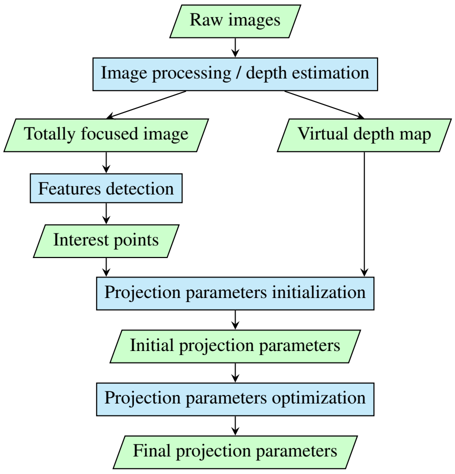
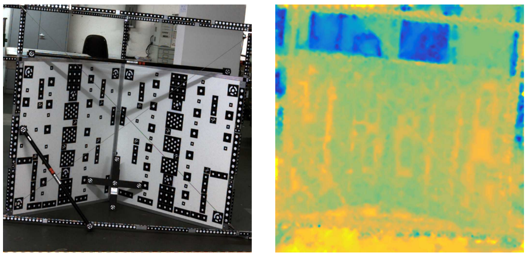

Raw images
Image processing / depth estimation
Totally focused image
Virtual depth map
Features detection
Interest points
Projection parameters initialization
Initial projection parameters
Projection parameters optimization
Final projection parameters

Figure 4. Flow chart of the calibration process.



(a) Totally focused image (b) Color coded virtual depth

Figure 5. Sample image of the 3D calibration target.

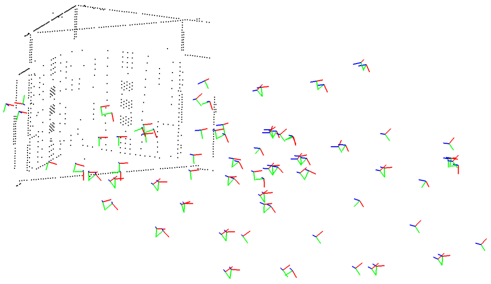


Figure 6. Calibration points and camera poses in 3D object space. Camera poses are represented by the three orthogonal unit vectors of the camera coordinates.

5.1 Preprocessing

In a preprocessing step from the recorded raw images virtual depth maps are estimated as well as totally focused intensity images are synthesized. For the experiments presented in Section 6.1 the algorithm implemented in the Raytrix API was used for depth estimation. Figure 5 shows an example of the input images (totally focused intensity image and virtual depth map) for the calibration process. Based on the totally focused intensity image feature points of the calibration target are extracted.

5.2 Intrinsic and Extrinsic Parameters Initialization

After the preprocessing initial intrinsic and extrinsic parameters are estimated based on the feature points. Hence, we perform a bundle adjustment based on a proprietary calibration software (Aicon 3D Systems, 1990) only using the totally focused images. In the bundle adjustment a regular pinhole camera model is used. Of course this does not correspond to reality but good initial parameters are received.

In that way, beside the initial extrinsic orientation for each frame (ω, ϕ, κ, t_x, t_y, t_z) and the 3D points in world coordinates (Fig. 6), the intrinsic camera parameters of a pinhole camera are received. These intrinsic parameters are the principal distance f, the principal point (c_x, c_y), as well as radial and tangential distortion parameters. In a first approximation the principal distance is assumed to be equal to the main lens focal length $f_L = f$. This assumption is sufficient to receive an initial parameter for the main lens focal length.

There are still two missing parameters to initialize the plenoptic camera model. Those parameters are b_{L0}, the distance between the main lens and the MLA, and B, the distance between the MLA and the sensor (Fig. 2). To estimate the initial values of those two parameters, we used the physical model of the camera as explained in (Zeller et al., 2014). Here the parameters b_{L0} and B are received by solving the linear equation given in expression (37) which is obtained from eq. (4).

$$b_L = b + b_{L0} = v \cdot B + b_{L0} \tag{37}$$

This is done by using all feature points in all recorded frames. Based on the initial value for the focal length of the main lens f_L and the z-component of the camera coordinates z_C the corresponding image distance b_L is calculated using the thin lens equation (eq. (4)). Finally, based on the image distances b_L, calculated for all feature points, and the corresponding virtual depth v, received from the depth maps, the parameters B and b_L0 are estimated.

5.3 Optimization based on Plenoptic Camera Model

After the initialization the plenoptic camera model is solved in a non-linear optimization process. In this optimization process all intrinsic as well as extrinsic parameters are adjusted based on the optimization problem defined in Section 4. Evaluations showed, that the 3D points received from the software have sub-millimeter accuracy and therefore will not to be adjusted in the optimization.

The cost function C, defined in eq. (31), is optimized with respect to the intrinsic and extrinsic parameters by the Levenberg-Marquardt algorithm implemented in the Ceres-Solver library (Agarwal et al., 2010).

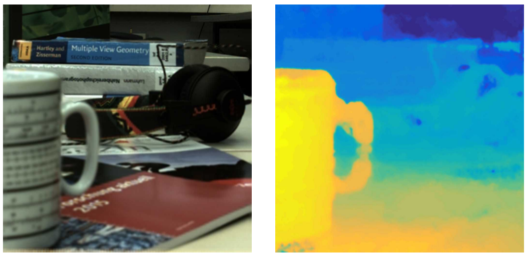
6. EVALUATION

In this section we evaluate our proposed calibration method. Here, we want to evaluate on one side the validity of the proposed model and on the other side to accuracy of the plenoptic camera itself as a depth sensor.

All experiments were performed based on a Raytrix R5 camera. Three setups with different main lens focal lengths ($f_L = 35$ mm, $f_L = 16$ mm, $f_L = 12.5$ mm) were used.

6.1 Experiments

6.1.1 Estimating the Camera Parameters In a first experiment for all three different main lens focal lengths the plenoptic camera model was estimated based on the proposed calibration method. Here for the $f_L = 35$ mm focal length 99 frames, for the $f_L = 16$ mm focal length 63 frames, and for the $f_L = 12.5$ mm focal length 50 frames were recorded and used for calibration.



(a) Totally focused image (b) Color coded virtual depth

Figure 7. Sample scene recorded with focal length $f_L = 35\,\text{mm}$ used for qualitative evaluation.

6.1.2 Measuring Metric 3D Accuracy In a second experiment we evaluate the measuring accuracy of the plenoptic camera itself, using our camera model. During the calibration we were not able to record frames for an object distance closer than approximately 1 m. For close distances the camera does not capture enough feature points to calculate meaningful statistics. Because of that reason, and since the depth range strongly decays with decreasing main lens focal length f_L we were only able to record meaningful statistics for the $f_L = 35\,\text{mm}$ focal length.

Thus, in this second experiment we evaluate the deviation of a virtual image point, measured by the camera and projected back to 3D space, from the actual ground truth which we received from our 3D calibration object.

Therefore, based on the recorded data we calculate the 3D projection error of the point with index i in the j-th frame $\Delta\boldsymbol{X}_C^{\{i,j\}}$, as given in eq. (39).

$$\boldsymbol{X}_C^{\{i,j\}} = (K_S \cdot K_V \cdot K_D \cdot K_I \cdot)^{-1} \cdot \lambda_{ij} \cdot \boldsymbol{X}_{Vd}'^{\{i,j\}} \tag{38}$$

$$\Delta\boldsymbol{X}_C^{\{i,j\}} = \boldsymbol{X}_C^{\{i,j\}} - G_j \cdot \boldsymbol{X}_W^{\{i\}} \tag{39}$$

Here, $\boldsymbol{X}_C^{\{i,j\}}$ is the measured virtual image point $\boldsymbol{X}_{Vd}'^{\{i,j\}}$ projected back to 3D space as defined in eq. (38), while $\boldsymbol{X}_W^{\{i\}}$ is the corresponding 3D reference point in world coordinates (ground truth).

The 3D projection error $\Delta\boldsymbol{X}_C^{\{i,j\}}$ is evaluated for object distances z_C from 1.4 m up to 5 m.

6.1.3 Calculating the Point Cloud of Real Scene In a last experiment we recorded a real scene and calculate the 3D metric point cloud out of it for a qualitative evaluation of the model. Therefore, the 3D metric point cloud was calculated for the scene shown in Figure 5 which was recorded with the $f_L = 35\,\text{mm}$ setup.

6.2 Results

6.2.1 Estimating the Camera Parameters Table 1 shows the resulting parameters of the estimated plenoptic camera model for all three setups ($f_L = 35\,\text{mm}$, $f_L = 16\,\text{mm}$, and $f_L = 12.5\,\text{mm}$). As one can see, for all three focal lengths the estimated parameter f_L is quite similar to the nominal focal length of the respective lens. The parameter B, which actually should not be dependent on the setup, stays quite constant over all setups and thus confirms the plausibility of the estimated parameters. The slight deviation in the parameter B for $f_L = 12.5\,\text{mm}$ probably can be explained by a too short range in which the virtual depth values of the feature points are distributed. Here, recording feature points for closer object distances could be beneficial.

lens	35 mm	16 mm	12.5 mm
f_L (in mm)	34.85	16.26	12.61
b_{L0} (in mm)	34.01	15.30	11.55
B (in mm)	0.3600	0.3695	0.4654
c_x (in pixel)	508.4	510.2	502.5
c_y (in pixel)	514.9	523.6	520.8

Table 1. Estimated intrinsic parameters of the proposed focused plenoptic camera model for different main lens focal length.

lens	35 mm	16 mm	12.5 mm
pinhole cam. (in pixel)	1.365	5.346	8.118
our model (in pixel)	0.069	0.068	0.078

Table 2. Comparison of the reprojection error using a pinhole camera model as well as our proposed focused plenoptic camera model for different main lens focal length.

Another indication for plausible parameters is, that to focus an image up to an object distance of infinity the condition $f_L \leq 2 \cdot B + b_{L0}$ has to be fulfilled (see (Perwaß and Wietzke, 2012)), which is the case for all three setups.

Furthermore, the reprojection error calculated from the feature points confirms the validity of our model. In Table 2 we compare the error for a regular pinhole camera model and for the proposed focused plenoptic camera model. As one can see, the error is improved by some orders of magnitude by introducing the plenoptic camera model. One can see that for a shorter main lens focal length f_L the reprojection error increases. This is the case, since here the projection in the plenoptic camera stronger deviates from the pinhole camera projection.

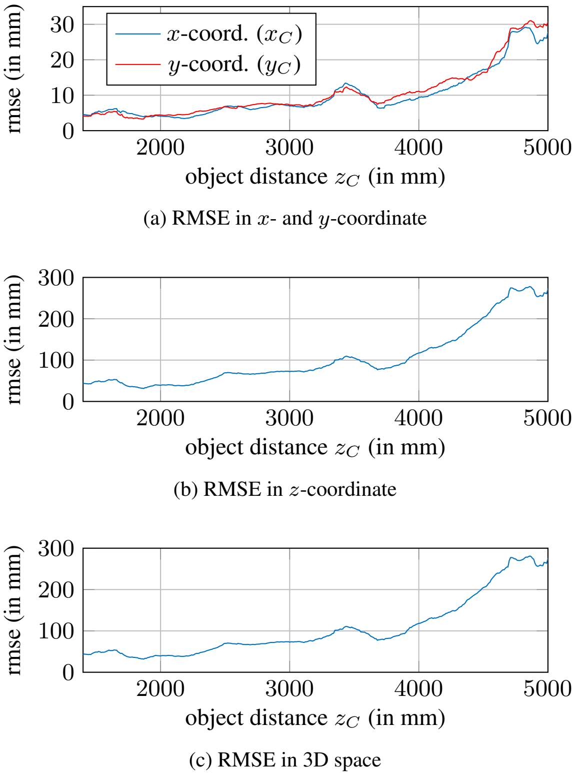
6.2.2 Measuring Metric 3D Accuracy Figure 8 shows the RMSE between the reference points and the back projected points in 3D object space for object distances from $z_C = 1.4\,\text{m}$ up to $z_C = 5\,\text{m}$, using the $f_L = 35\,\text{mm}$ focal length. We calculated the RMSE for each coordinate separately as well as for the complete 3D error vector $\Delta\boldsymbol{X}_C$.

As could be expected, the error along the x- and y-coordinate is much smaller than the one in z-direction. Nevertheless, all three errors increase approximately proportional to each other with increasing object distance z_C. This is quite obvious, since the coordinates x_C and y_C are linearly dependent on z_C and thus also effected by the depth error.

The overall error in 3D space ranges from 50 mm at an object distance of 1.4 m up to 300 mm at 5 m distance. This conforms quite well to the evaluations made in other publications as well as to the specifications given by Raytrix (see (Zeller et al., 2014)).

6.2.3 Calculating the Point Cloud of Real Scene Finally, Figure 9 shows the calculated metric point cloud for the sample scene given in Figure 7. Here one can see, that the cup which occurs quite large in the perspective image is scaled down to its metric size. Besides, the compressed virtual depth range is stretched to real metric depths.

In the figure are some artifacts visible on the right side of the scene. This artifacts result from wrong depth values which occur on one hand due to textureless regions in the scene and on the other hand due to very far object distances. For textureless regions no depth can be estimated and therefore the depth map is filled by interpolation based on neighboring depth pixels. The artifacts on the ”Multiple View Geometry” book result from the headphone cable in the front of the book (see Fig.7).



(a) RMSE in x- and y-coordinate

(b) RMSE in z-coordinate

(c) RMSE in 3D space

Figure 8. RMSE of feature points projected back to 3D object space.

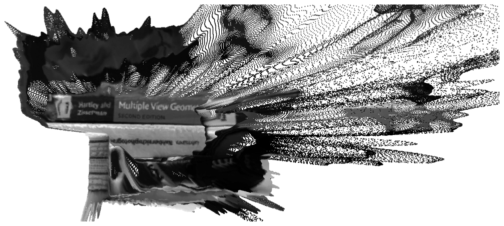


Figure 9. 3D metric point cloud of a sample scene (Fig. 7) calculated based on our focused plenoptic camera model.

7. CONCLUSION

In this paper we presented a new model for a focused plenoptic camera as well as a metrical calibration approach based on a 3D target. Different to priorly published models we consider lens distortion to be constant along a central ray through the main lens. Besides, we derived a depth distortion model directly from the theory of depth estimation in a focused plenoptic camera.

We applied our calibration to three different setups which all gave good results. In addition we measured the accuracy of the focused plenoptic camera for different object distances. The measured accuracy conforms quite well to what is theoretically expected.

In future work we want to further improve our calibration approach. On one hand our calibration target has to be adapted to be able to get feature points for object distances closer than 1 m. Besides, a denser feature point pattern is needed to reliably estimated depth distortion.

ACKNOWLEDGEMENTS

This research is funded partly by the Federal Ministry of Education and Research of Germany in its program "IKT 2020 Research for Innovation" and partly by the French government research program "Investissements d'avenir" through the IMobS^3 Laboratory of Excellence (ANR-10-LABX-16-01) and the FUI n° 17 CLEAN Robot.

REFERENCES

Adelson, E. H. and Wang, J. Y. A., 1992. Single lens stereo with a plenoptic camera. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14(2), pp. 99–106.

Agarwal, S., Mierle, K. and Others, 2010. Ceres solver. http://ceres-solver.org.

Aicon 3D Systems, 1990. Aicon 3d studio. http://aicon3D.com.

Bishop, T. E. and Favaro, P., 2009. Plenoptic depth estimation from multiple aliased views. In: *IEEE 12th Int. Conf. on Computer Vision Workshops (ICCV Workshops)*, pp. 1622–1629.

Bok, Y., Jeon, H.-G. and Kweon, I., 2014. Geometric calibration of micro-lens-based light-field cameras using line features. In: *Computer Vision - ECCV 2014*, Lecture Notes in Computer Science, Vol. 8694, pp. 47–61.

Brown, D. C., 1966. Decentering distortion of lenses. *Photogrammetric Engineering* 32(3), pp. 444–462.

Dansereau, D., Pizarro, O. and Williams, S., 2013. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1027–1034.

Gortler, S. J., Grzeszczuk, R., Szeliski, R. and Cohen, M. F., 1996. The lumigraph. In: *Proc. 23rd Annual Conf. on Computer Graphics and Interactive Techniques, SIGGRAPH*, ACM, New York, NY, USA, pp. 43–54.

Heinze, C., Spyropoulos, S., Hussmann, S. and Perwass, C., 2015. Automated robust metric calibration of multi-focus plenoptic cameras. In: *Proc. IEEE Int. Instrumentation and Measurement Technology Conf. (I2MTC)*, pp. 2038–2043.

Johannsen, O., Heinze, C., Goldluecke, B. and Perwaß, C., 2013. On the calibration of focused plenoptic cameras. In: *GCPR Workshop on Imaging New Modalities*.

Lumsdaine, A. and Georgiev, T., 2008. Full resolution lightfield rendering. Technical report, Adobe Systems, Inc.

Lumsdaine, A. and Georgiev, T., 2009. The focused plenoptic camera. In: *Proc. IEEE Int. Conf. on Computational Photography (ICCP)*, San Francisco, CA, pp. 1–8.

Ng, R. and Hanrahan, P., 2006. Digital correction of lens aberrations in light field photography. In: *Proc. SPIE 6342, International Optical Design Conference*, Vol. 6342.

Ng, R., Levoy, M., Brédif, M., Guval, G., Horowitz, M. and Hanrahan, P., 2005. Light field photography with a hand-held plenoptic camera. Technical report, Stanford University, Computer Sciences, CSTR.

Perwaß, C. and Wietzke, L., 2012. Single lens 3d-camera with extended depth-of-field. In: *Proc. SPIE 8291, Human Vision and Electronic Imaging XVII*, Burlingame, California, USA.

Zeller, N., Quint, F. and Stilla, U., 2014. Calibration and accuracy analysis of a focused plenoptic camera. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3, pp. 205–212.

Zeller, N., Quint, F. and Stilla, U., 2015. Establishing a probabilistic depth map from focused plenoptic cameras. In: *Proc. Int. Conf. on 3D Vision (3DV)*, pp. 91–99.

doi:10.5194/isprsannals-III-3-449-2016