Computer Vision and Image Understanding 145 (2016) 140–147

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Stepwise calibration of focused plenoptic cameras

Klaus H. Strobl*, Martin Lingenauber

Robotics and Mechatronics Center, German Aerospace Center (DLR), D-82230 Wessling, Germany

ARTICLE INFO

Article history:
Received 15 December 2014
Accepted 22 December 2015

Keywords:
Focused plenoptic camera
Light-field
Plenoptic 2.0
Metric calibration

ABSTRACT

Monocular plenoptic cameras are slightly modified, off-the-shelf cameras that have novel capabilities as they allow for truly passive, high-resolution range sensing through a single camera lens. Commercial plenoptic cameras, however, are presently delivering range data in non-metric units, which is a barrier to novel applications e.g. in the realm of robotics. In this work we revisit the calibration of focused plenoptic cameras and bring forward a novel approach that leverages traditional methods for camera calibration in order to deskill the calibration procedure and to increase accuracy. First, we detach the estimation of parameters related to either brightness images or depth data. Second, we present novel initialization methods for the parameters of the thin lens camera model—the only information required for calibration is now the size of the pixel element and the geometry of the calibration plate. The accuracy of the calibration results corroborates our belief that monocular plenoptic imaging is a disruptive technology that is capable of conquering new markets as well as traditional imaging domains.

1. Introduction

A considerable number of machine vision users think that multi-view triangulation is required in order to retrieve accurate 3-D information using cameras without a priori information about the scene. This is predominantly accomplished either by passively taking images from different vantage points (stereo vision) or by actively projecting a known pattern from a separate location (Kinect). In other words, the notion that 3-D information is lost when light rays traverse the front lens of the camera is widespread. Experts know, however, that this is not the case as light rays are differently diffracted by a lens depending on the distance to the emitting object [1]. In fact, one of the potential outputs of monocular plenoptic cameras is range sensing.

1.1. Plenoptic imaging

Plenoptic (also light-field) imaging is about measuring light in a higher dimensionality than in standard 2-D imaging. In fact, light transmission can be contemplated in a higher-dimensional space, the so-called plenoptic function [2]. Current plenoptic imaging samples the plenoptic function in 4-D , viz. 2-D projection position on the chip together with the direction of incoming light rays. The quest for this extra information is anything but new, but advances in parallel computation and modern workmanship of microlens arrays (MLAs) have recently made commercial products possible [3–5]. In a nutshell, monocular plenoptic cameras capture the 3-D image produced by the main lens within the camera by using an MLA in front of the sensor chip. By capturing the whole 3-D image, classical habits when using planar sensors like keeping the aperture size small in order to increase depth of field are lifted and more light can be gathered from the scene. Different camera designs open up new possibilities to trade off lateral precision against angular resolution of the reprojected ray directions. The original monocular plenoptic cameras focus the image on the MLA, achieving a limited spatial resolution at that particular depth equal to the number of valid microlenses. These microlenses produce defocused images that sample the ray direction at the position of the microlens. In 2009 Lumsdaine and Georgiev introduced the focused plenoptic camera (or plenoptic camera 2.0), which makes it possible to adapt this rather rigid trade-off between angular and spatial resolutions towards more spatial resolution [6]. This is performed by a modification of the focus distance to the main lens with the result that microlenses produce focused images that, on the other hand, more loosely sample ray direction.

Many characteristics of plenoptic cameras are in conformity with the standard reference on disruptive technologies in Ref. [7]. For instance, plenoptic cameras initially produce a deficient standard output (fair images) at a higher cost, which makes them of no interest to the average consumer. They, however, clearly have the potential to improve and open up new markets while sharply reducing costs. Their current applications are offline refocusing and

* Corresponding author.
E-mail address: klaus.strobl@dlr.de (K.H. Strobl).

http://dx.doi.org/10.1016/j.cviu.2015.12.010
1077-3142/

total focusing (*i.e.*, increased depth of field) of 2-D images. More relevant potential applications are passive, 3-D video recording, 3-D modeling, range-based segmentation and tracking, industrial inspection (e.g. in narrow cavities), and imaging in challenging, low-light environments (e.g. underwater or in space). Most of these applications rely on the capability of plenoptic cameras to provide *metric* information of the scene in the form of 2.5-D depth images. This is, however, not yet commercially available as depth is currently being delivered in internal units related to image processing (disparities). We address the metric calibration of focused, monocular plenoptic cameras in order to transform their depth output into metric space.

1.2. Related work

There is less research on the metric calibration of focused plenoptic cameras and the works on the calibration of the original, unfocused ones are only of partial use [8]. Next we review the only available approaches in Refs. [9–12]. They all have in common that they start out from synthetic images generated by the RxLive software of Raytrix GmbH (viz. the total focus image and the depth image), not from the raw images of the camera. The conformity of the generated synthetic images with the camera models used for calibration is of course critical. It is a judicious decision to rely on the manufacturers, however, since (i) they are most qualified to do that job, (ii) they still keep individual design details in secret, and (iii) in order to avoid mismatching between our potential reconstruction attempts and the eventual operation on GPGPUs. In addition, the calibration process is simplified by using the synthetic images because we leverage established methods for pinhole camera calibration [13,14]. It is worth noting that the geometry of the MLA is not included in the calibration process as it can be estimated in a separate procedure using the Raytrix software. The best-known work in Ref. [9] details the modeling and calibration of the focused plenoptic camera, failing to obtain absolute range accuracy. Further the automatic initialization of calibration parameters is not addressed; the authors make use of privileged information from the manufacturer. The recent Master's thesis in Ref. [12], however, does achieve superior results by largely implementing the above approach. Still, considerations on the initialization of calibration parameters are not being addressed. The author makes strong use of filtering approaches to wipe out peripheral artifacts in depth estimation, which might constrain the general applicability of the approach. Zeller et al. in Ref. [11] perform calibration by minimizing the reprojection residuals with respect to (w.r.t.) a set of measured calibration points for which the object distance is known—at least for the initialization of their method. In addition, the method requires assumed intrinsic values. In the same spirit, Luhmann et al. in Ref. [10] opt for measuring ranges of a planar calibration object, which is error-prone and inconvenient [15,16]. Incidentally, the current internal approach for metric calibration at Raytrix GmbH also relies on a linear actuator in order to produce a polynomial that directly converts virtual depths into metric distances [12]. This type of empirical models, however, is only applicable within the scope of the calibration data.

1.3. Contributions

In this work we revisit the type of calibration approaches that are based on the standard camera calibration method described in Refs. [13,14]. We suggest modifications to particular modeling details and present justifications. Special care has been taken to keep the approach in the spirit of the standard method, *i.e.*, to take images of a known planar calibration pattern in unknown pose and to facilitate automatic bootstrapping of the parameters prior to nonlinear optimization. This keeps the amount of required prior knowledge (e.g. specifications by the manufacturer) to a minimum, making the whole calibration process more generic and easier. More importantly, we introduce a novel approach for stepwise calibration by alternate use of total focus and depth (synthetic) images. Our motivation is to avoid the impact of higher levels of noise in the depth images on a large part of the intrinsic parameters like the focal length and the radial lens distortion. These parameters can be estimated in advance by exclusively using total focus images, as in traditional camera calibration. After that, the optimization of the remaining parameters using the depth images and the results of the first optimization takes place, see Fig. 1. By doing so, calibration accuracy is increased and the optimization robustness is promoted as the formulations of both optimizations become better conditioned compared with joint optimization methods [9]. In addition, the optimization of the lens distortion model will not get entangled with the optimization of the (potentially very similar) depth distortion model.

Overall, we produce an easy-to-use, automatic method for metric plenoptic camera calibration. The only required data are synthetic total focus and virtual depth images of a planar calibration plate of known geometry and the metric size of their virtual sensor elements.

2. Proposed method

2.1. The thin lens camera model and the focused plenoptic camera

The pinhole camera model is a valid approximation for most cameras and applications. It relates projection directions in the camera reference frame S_C with projection positions in the sensor reference frame S_S. This projection is independent of the actual range to the scene, which makes it unsuitable for modeling plenoptic cameras aimed at inferring the depth of the scene out of inner camera projections. The pinhole camera model is derived from the thin lens camera model in the case of smaller aperture sizes. The thin lens camera model embraces the thin lens approximation of light rays passing through a thin lens, which states that ray directions are still projected following the pinhole camera

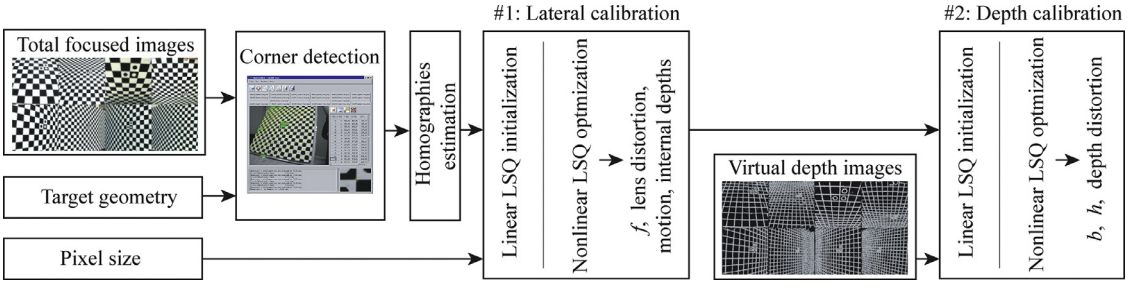


Fig. 1. Diagram representing the information flow in the presented method.

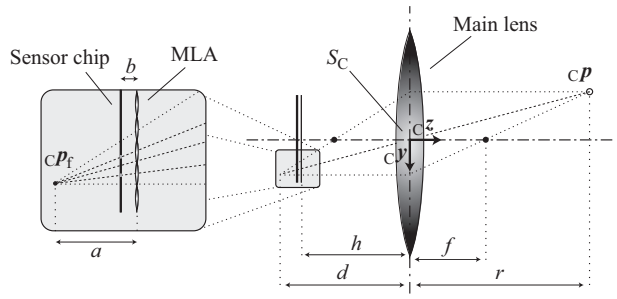


Fig. 2. The thin lens camera model and the focused plenoptic camera.

model *and* that projections are only focused at particular depths d that depend both, on the focal length f of the lens and on the object ranges r in the direction of the principal axis of the camera as follows:

$$\frac{1}{f} = \frac{1}{d} + \frac{1}{r}. \tag{1}$$

The thin lens camera model does consider ranges to the scene and therefore serves as a starting point for the camera model of a plenoptic camera. The use of the pinhole camera model has been convenient because of its linear projective formulation in homogeneous coordinates ($\bar{}$). Similarly, the thin lens camera model allows for such a formulation as follows:

$$ {}_C\bar{\boldsymbol{p}}_f = \begin{bmatrix} {}_Cx_f \\ {}_Cy_f \\ {}_Cz_f \\ 1 \end{bmatrix} = \begin{bmatrix} {}_Sx \cdot p \\ {}_Sy \cdot p \\ {}_Cz_f \\ 1 \end{bmatrix} \propto \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix} \cdot \begin{bmatrix} {}_Cx \\ {}_Cy \\ {}_Cz \\ 1 \end{bmatrix}, \tag{2}$$

introducing the focused projection $_C\boldsymbol{p}_f$ of the point in space $_C\boldsymbol{p}$, both vectors represented in 3-D in S_C. The metric 2-D orthogonal projection of $_C\boldsymbol{p}_f$ $\{_Cx_f, {}_Cy_f\}$ can be transformed to a virtual S_S using the length p of the virtual sensor element. The virtual sensor corresponds to the sensor that would produce the synthetic images delivered e.g. by the RxLive software out of raw sensor images. Note that we choose to shift pixel values to the center of projection at the principal axis of the camera since camera calibration is less sensitive to the actual position of the principal point (at the cost of a slightly different pose of S_C, see Refs. [17,18]) and because lens distortion will feature its own 2-D origin $\{_Cx_r, {}_Cy_r\}$ in the first place.

The internal depth d of projections $_Cz_f$ is the central value of this model. It depends on the actual range r to the scene $_Cz$ and on the focal length f of the main lens. In the case of the focused plenoptic camera it is possible to estimate $_Cz_f$ using raw camera projections, see Fig. 2. In fact, the depth estimation algorithm of Raytrix cameras delivers the distance a between the MLA and the internal depth $_Cz_f$ in multiples of the distance b between the MLA and the sensor chip, which is an unknown value related with camera production [19]. That relative distance is called virtual depth $v = a/b$. In detail, virtual depths are provided as normalized values P coded in 16 bits that are to be converted to real values and divided by their potential maximum 65535, resulting in values in the range of [0.5, 1.0). This normalized internal depth value is transformed to actual, relative virtual depths $v = (1 - P)^{-1}$, which are in turn related to the internal depth projections using b and the distance h between the MLA and S_C as follows:

$$ {}_Cz_f = v \cdot b + h. \tag{3}$$

Virtual depths are, however, of limited use to final applications because they are not metric and because they involve a nonlinear relationship with the actual depths in the scene $_Cz$. In order to be able to transform virtual depths into actual metric depths it is necessary to estimate b and h with high accuracy. It is worth noting that it is the knowledge of virtual depths v that enables plenoptic cameras to deliver synthetic, total focus 2-D brightness images composed of projections at their respective focused depths $_Cz_f$.

2.2. Calibration approach

Our method is based on the two types of synthetic images explained above: first, depth images featuring virtual depths v at their virtual sensor projections $\{_Sx, {}_Sy\}$, and second, total focus images featuring actual, focused brightness values captured at the same virtual sensor projections $\{_Sx, {}_Sy\}$. Traditional pinhole camera calibration approaches only use the latter brightness images, minimizing reprojection residuals in S_S for optimal estimation of intrinsic parameters following the maximum likelihood criterion, which holds because checkerboard corner detection by image processing is prone to errors that can be modeled by 2-D independent and identically distributed zero-mean Gaussian distributions [13,14]. In the case of the thin lens camera model, the information conveyed by brightness images, together with the model of the calibration target, does allow to parameterize Eq. (2) including $_Cz_f$. In order to estimate the unknown intrinsic parameters b and h in Eq. (3), however, the information v conveyed by the virtual depth images is also needed. The previous calibration approaches for focused plenoptic cameras included both types of data into the thin lens camera model in Eqs. (2) and (3) in order to obtain *all* intrinsic and extrinsic parameters by global minimization of Euclidean 3-D residual distances [9]. It is difficult to justify the optimality of that approach in the face of multi-modal data and an aleatory choice of residuals. Virtual depth images feature a much higher level of noise than brightness images after all. It is clear that by using multi-modal data for joint optimization, the accuracy of all results will be compromised by both noise sources—especially by the strongest source, *i.e.*, the virtual depth images. Even though information theory says that even the noisiest bit of information is able to increase the overall information budget, in reality this is here not the case as the stochastic error distribution of virtual depths has not yet been properly modeled.

In this work we note that most intrinsic parameters of the thin lens camera model can be estimated *without* having to make recourse to noisy virtual depth images. Virtual depths are a new type of information that is not yet accurately understood and modeled. We propose to recur to this type of data only when strictly necessary by the following procedure (cf. Fig. 2):

1. First, the focal length f and, in the case of radial lens distortion, the parameters ($_Cx_r$, $_Cy_r$, k_1 and perhaps k_2) are automatically estimated using brightness images. As a by-product, we obtain the optimal extrinsic parameters of the camera (*i.e.*, the camera motion) with highest accuracy as well as the optimal internal depths of projections $_Cz_f$. The intrinsic parameters are then fixed for future estimations. Camera calibration from brightness images is a trusted science after all.
2. Second, virtual depths images are used to estimate the inner lengths b and h in Eq. (3)—perhaps together with the parameters of a depth distortion model.

In addition, the detachment of these two types of information during the calibration process proves to be useful whenever it is required to robustify the calibration process against data outliers. For example, the whole set of brightness features can be used to perform lateral calibration (Stage #1) whereas the most noisy depth values can be readily removed during the estimation of the depth-related parameters (Stage #2). This is not possible when using the state-of-the-art calibration algorithms that mix both data

types during calibration. An additional advantage is that the lens distortion model and the depth distortion model will not become entangled within a sole optimization, which otherwise would be a problem since both models are potentially similar. Lastly, the unwanted correlations between the focal length f and the inner camera lengths b and h are suppressed.

Without loss of generality in this work we suggest checkerboard calibration patterns [20]. In Ref. [12] the author chooses circular features because Raytrix, historically, did so. They calculate their projected centroids and average over the depth values of the whole ellipse. It is worth noting, however, that in the case of circular features, the center of the ellipse is generally *not* the same as the projected circle center [21]. Similarly, averaging over depth values of the ellipse to find the depth of its centroid is also prone to errors. Both of these aspects are better managed when using checkerboard patterns. The correspondence problem is easily solved e.g. by using the calibration software DLR CalDe [22].

Lastly, we introduce automatic, sequential initialization schemes for all parameters involved in the staged calibration.

2.3. Stage #1: Lateral calibration

By removing the third, depth-related row in Eq. (2) we obtain:

$$\begin{bmatrix} {}_{\mathrm{S}}x \\ {}_{\mathrm{S}}y \\ 1 \end{bmatrix} \propto \begin{bmatrix} \frac{1}{p} & 0 & 0 & 0 \\ 0 & \frac{1}{p} & 0 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix} \cdot \begin{bmatrix} {}_{\mathrm{C}}x \\ {}_{\mathrm{C}}y \\ {}_{\mathrm{C}}z \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{p} & 0 & 0 & 0 \\ 0 & \frac{1}{p} & 0 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix} \cdot \begin{bmatrix} & {}_{\mathrm{C}}\boldsymbol{R}^{\mathrm{O}} & & {}_{\mathrm{C}}\boldsymbol{t}^{\mathrm{O}} \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} {}_{\mathrm{O}}x \\ {}_{\mathrm{O}}y \\ {}_{\mathrm{O}}z \\ 1 \end{bmatrix}, \tag{4}$$

which represents a projection model from 3-D coordinates in S_{C} or the object reference frame S_{O} to 2-D projections in S_{S}. This formulation is fully in the spirit of the thin lens camera model. Since the calibration object is planar, *i.e.*, ${}_{\mathrm{O}}z \triangleq 0$, this formulation allows for rapid initialization of the focal length f and the rigid body transformation $\left\{ {}_{\mathrm{C}}\boldsymbol{R}^{O} = [\boldsymbol{r}_1 \boldsymbol{r}_2 \boldsymbol{r}_3],\ {}_{\mathrm{C}}\boldsymbol{t}^{O} = \left[{}_{\mathrm{C}}x^{O}\ {}_{\mathrm{C}}y^{O}\ {}_{\mathrm{C}}z^{O} \right]^{\mathsf{T}} \right\}$ using planar homographies $\boldsymbol{H}_{(3\times3)}$ similar to the traditional pinhole approach in Refs. [13,14]. Homographies $\boldsymbol{H}$ can be easily estimated for every calibration image as the linear least squares solution of the homogeneous formulation including two equations for every measured ($\tilde{}$) feature ${}_{\mathrm{S}}\tilde{\boldsymbol{p}}$ [23]. The importance of normalizing data cannot be overestimated at this point.

First, a novel correspondence between the homography and the unknown transformations is established:

$$\boldsymbol{H}_{(3\times3)} = [\boldsymbol{h}_1\, \boldsymbol{h}_2\, \boldsymbol{h}_3] \propto \begin{bmatrix} \frac{1}{p} & 0 & 0 & 0 \\ 0 & \frac{1}{p} & 0 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & {}_{\mathrm{C}}\boldsymbol{t}^{O} \\ 0 & 0 & 1 \end{bmatrix} \approx \underbrace{\begin{bmatrix} \frac{1}{p} & 0 & 0 \\ 0 & \frac{1}{p} & 0 \\ 0 & 0 & -\frac{1}{f} \end{bmatrix}}_{\boldsymbol{A}_{(3\times3)}} \cdot \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & {}_{\mathrm{C}}\boldsymbol{t}^{O} \end{bmatrix} \Leftrightarrow f \ll {}_{\mathrm{C}}z. \tag{5}$$

The correspondence has been simplified because the focal length is generally much shorter than the range to the feature in front of the main lens ${}_{\mathrm{C}}z$, when represented in the same distance units. Now using the orthonormality constraints $\boldsymbol{r}_1 \cdot \boldsymbol{r}_2 = 0,\ \boldsymbol{r}_1 \cdot \boldsymbol{r}_1 = 1$, and $\boldsymbol{r}_2 \cdot \boldsymbol{r}_2 = 1$, *i.e.*, ${}_{\mathrm{C}}\boldsymbol{R}^{O} \in SO(3)$, we obtain:

$$\left.\begin{aligned} (\boldsymbol{A}^{-1}\boldsymbol{h}_1)^{\mathsf{T}} \cdot (\boldsymbol{A}^{-1}\boldsymbol{h}_2) &= 0 \\ (\boldsymbol{A}^{-1}\boldsymbol{h}_1)^{\mathsf{T}} \cdot (\boldsymbol{A}^{-1}\boldsymbol{h}_1) - (\boldsymbol{A}^{-1}\boldsymbol{h}_2)^{\mathsf{T}} \cdot (\boldsymbol{A}^{-1}\boldsymbol{h}_2) &= 0 \end{aligned}\right\} \Leftrightarrow \left.\begin{aligned} \boldsymbol{h}_1^{\mathsf{T}}\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_2 &= 0 \\ \boldsymbol{h}_1^{\mathsf{T}}\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_1 &= \boldsymbol{h}_2^{\mathsf{T}}\, \boldsymbol{\omega}_\infty\, \boldsymbol{h}_2 \end{aligned}\right\}, \tag{6}$$

with the so-called absolute conic

$$\boldsymbol{\omega}_\infty = \boldsymbol{A}^{-\mathsf{T}}\boldsymbol{A}^{-1} = \begin{bmatrix} p^2 & 0 & 0 \\ 0 & p^2 & 0 \\ 0 & 0 & f^2 \end{bmatrix}. \tag{7}$$

Eq. (6) enable the direct estimation ($\hat{}$) of the focal length f either using the orthogonality constraint ($\hat{f}_1$) or using the normalization constraint ($\hat{f}_2$):

$$\hat{f}_1 = \pm p \cdot \sqrt{\frac{-h_{11}h_{12} + h_{21}h_{22}}{h_{31}h_{32}}}, \quad \hat{f}_2 = \pm p \cdot \sqrt{\frac{h_{12}^2 + h_{22}^2 - h_{11}^2 - h_{21}^2}{h_{31}^2 - h_{32}^2}}, \tag{8}$$

for every single calibration image of the checkerboard plate. Note that $\boldsymbol{h}_1 = [\,h_{11}\ h_{21}\ h_{31}\,]^{\mathsf{T}}$ and $\boldsymbol{h}_2 = [\,h_{12}\ h_{22}\ h_{32}\,]^{\mathsf{T}}$. The only required data for the metric initialization of f are the side length p of the virtual sensor pixel featuring total focus and virtual depth images together with the geometry of the checkerboard calibration plate, *i.e.*, the N coordinates of its corners ${}_{\mathrm{O}}\boldsymbol{p}_i = [{}_{\mathrm{O}}x_i\ {}_{\mathrm{O}}y_i\ {}_{\mathrm{O}}z_i]^{\mathsf{T}},\ \forall i \in \{1, \ldots, N\}$. The requirement on the knowledge of the latter geometry could, however, be partially lifted [16]. Experiments show that the estimation $\hat{f}_2$ is slightly better conditioned with regard to the amount of perspective distortion included in the calibration images (orthogonal plate projections are widely discouraged for camera calibration [24]). We choose the median of all the $\hat{f}_2$ estimations for every C calibration images, arriving at a value $\hat{f} = \mathrm{median}(\hat{f}_{2_c}),\ \forall c \in \{1, \ldots, C\}$, that closely matches the nominal focal length of the lens unit.

The absolute extrinsic camera parameters $\{{}_{\mathrm{C}}\boldsymbol{R}^{O}, {}_{\mathrm{C}}\boldsymbol{t}^{O}\}$ can then be estimated for every calibration image c using both, the C homographies $\boldsymbol{H}_c$ and the approximated intrinsic matrix $\boldsymbol{A}$ including the virtual pixel size p and the newly estimated focal length $\hat{f}$ as follows: $\hat{\boldsymbol{r}}_1 = 1/s\boldsymbol{A}^{-1}\boldsymbol{h}_1,\ \hat{\boldsymbol{r}}_2 = 1/s\boldsymbol{A}^{-1}\boldsymbol{h}_2,\ \hat{\boldsymbol{r}}_3 = \hat{\boldsymbol{r}}_1 \times \hat{\boldsymbol{r}}_2,\ \hat{\boldsymbol{t}} = 1/s\boldsymbol{A}^{-1}\boldsymbol{h}_3$, and $s = ||\boldsymbol{A}^{-1}\boldsymbol{h}_1|| = ||\boldsymbol{A}^{-1}\boldsymbol{h}_2||$. These can also be used for a potential hand-eye calibration of the plenoptic camera [25].

At this point all parameters have been estimated with reasonable accuracy. It is known that the radial lens distortion parameters can be initialized at zero value for subsequent nonlinear optimization. If necessary, they can also be estimated in advance, perhaps stand-alone [26,27].

Last, the optimal ($\star$) parameters $\widehat{\boldsymbol{\Omega}}_\star$ including $\hat{f}_\star$, ${}_{\mathrm{C}}\hat{x}_{\mathrm{r}\star}$, ${}_{\mathrm{C}}\hat{y}_{\mathrm{r}\star}$, $\hat{k}_{1\star}$, $\hat{k}_{2\star}$ as well as the C extrinsic transformations $\{{}_{\mathrm{C}}\widehat{\boldsymbol{R}}^{O}_{c\star}, {}_{\mathrm{C}}\hat{\boldsymbol{t}}^{O}_{c\star}\}$ can be estimated on the basis of the maximum likelihood criterion, *i.e.*, by sensibly minimizing the discrepancies between the erroneous measurements $\tilde{\boldsymbol{p}}$ and the expected, distorted projections ${}_{\mathrm{S}}\hat{\boldsymbol{p}}_{\mathrm{d}}$ of the actual corners of the calibration plate ${}_{\mathrm{O}}\boldsymbol{p}$ as follows:

$$\widehat{\boldsymbol{\Omega}}_\star = \arg\min_{\widehat{\boldsymbol{\Omega}}} \sum_{c=1}^{C} \sum_{i} \left\| {}_{\mathrm{S}}\tilde{\boldsymbol{p}}^{\{c,i\}} - {}_{\mathrm{S}}\hat{\boldsymbol{p}}_{\mathrm{d}}^{\{c,i\}}\left(\widehat{\boldsymbol{\Omega}},\ p,\ {}_{\mathrm{O}}\boldsymbol{p}_i\right) \right\|^2. \tag{9}$$

The expected projections ${}_{\mathrm{S}}\hat{\boldsymbol{p}}_{\mathrm{d}}$ depend on the calibration parameters $\boldsymbol{\Omega}$ to be optimized, on the side length p, and on the known geometry of the corners of the calibration plate ${}_{\mathrm{O}}\boldsymbol{p}$. The calibration parameters $\boldsymbol{\Omega}$ are initialized as explained above.

A word of caution regarding the formulation of the lens distortion model as in Ref. [28]: First, the only formulation that is correct on a physical ground, viz. based on Snell's refraction law, applies

to virtual, undistorted projections derived from the actual scene (u→d formulation). Many authors apply the formulation the other way around (d→u formulation), which is wrong on a strict, physical ground and can be potentially misinterpreted when it comes to using the parameterized model. Experiments show, however, that the d→u formulation does come very close by the physically-conform u→d formulation [23]. Second, it is necessary to state the dimensions of the lens distortion model along with delivering the calibration results, *i.e.*, whether it has been estimated on normalized directional coordinates or on projected pixels or millimeters. Third, it is a good idea to use ${}_{\mathrm{C}}x_{\mathrm{r}}$ and ${}_{\mathrm{C}}y_{\mathrm{r}}$ to detach the camera's principal point from the origin of lens distortion, which is equivalent to releasing the first degree of freedom of lens decentering distortion [18].

2.4. Stage #2: Depth calibration

Metric calculation of depths in front of a plenoptic camera demands metric knowledge about its inner lengths b and h. These parameters connect virtual depths with actual ranges in S_{C}, cf. Eq. (3). It is by the relationship between virtual and actual depths that we will be able to estimate these parameters in a novel way.

In Section 2.3 we removed the third row of the system of equations in Eq. (2)

$$ {}_{\mathrm{C}}z_{\mathrm{f}} = \frac{{}_{\mathrm{C}}z}{-{}_{\mathrm{C}}z/f+1} \Big/ {}_{\mathrm{C}}z = r_{31}\,{}_{\mathrm{O}}x + r_{32}\,{}_{\mathrm{O}}y + r_{33}\,{}_{\mathrm{O}}z + {}_{\mathrm{C}}z^{O}, \tag{10} $$

obtaining Eq. (4). Eq. (10), however, can readily be executed *after* optimal estimation of the other intrinsic parameters in Eq. (9), obtaining optimal depth values ${}_{\mathrm{C}}\hat{z}_{\mathrm{f}\star}^{\{c,\,i\}}$ for all features and C calibration images. These values, together with the virtual depths $\tilde{v}^{\{c,\,i\}}$ acquired during calibration, can be used to estimate the inner lengths h and b by using Eq. (3). First, we initialize these values by ordinary least squares on all available data:

$$ \underbrace{\begin{bmatrix} \tilde{v}^{\{1,1\}} & 1 \\ \tilde{v}^{\{1,2\}} & 1 \\ \vdots & 1 \end{bmatrix}}_{\boldsymbol{C}} \begin{bmatrix} b \\ h \end{bmatrix} = \underbrace{\begin{bmatrix} {}_{\mathrm{C}}\hat{z}_{\mathrm{f}\star}^{\{1,1\}} \\ {}_{\mathrm{C}}\hat{z}_{\mathrm{f}\star}^{\{1,2\}} \\ \vdots \end{bmatrix}}_{\boldsymbol{d}} \Rightarrow \begin{bmatrix} \hat{b} \\ \hat{h} \end{bmatrix} = \mathrm{inv}\big(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\big)\boldsymbol{C}^{\mathrm{T}}\boldsymbol{d}. \tag{11} $$

Note that these parameters can be estimated using a sole calibration image but using more data is beneficial. Earlier approaches initialized these parameters using privileged information from the manufacturer, together with the focus distance that can be gauged from the lens unit, which is both inconvenient and error-prone.

After that, the whole thin lens camera model can be used to optimize only these two parameters $\hat{\boldsymbol{\Phi}} = \{\hat{b}, \hat{h}\}$ as follows:

$$ \hat{\boldsymbol{\Phi}}_{\star} = \arg\min_{\hat{\boldsymbol{\Phi}}} \sum_{c=1}^{C} \sum_{i} \left\| {}_{\mathrm{C}}\tilde{\boldsymbol{p}}_{\mathrm{f}}^{\{c,\,i\}}\big(\hat{\boldsymbol{\Phi}}\big) - {}_{\mathrm{C}}\hat{\boldsymbol{p}}_{\mathrm{f,d}}^{\{c,\,i\}}\big(p,\ {}_{\mathrm{O}}\boldsymbol{p}_i,\ \hat{\boldsymbol{\Omega}}_{\star}\big) \right\|^2, \tag{12} $$

making use of the optimal parameters $\hat{\boldsymbol{\Omega}}_{\star}$ obtained in Eq. (9). This optimization is well conditioned and converges in a few steps. Note that we opt for minimizing 3-D Euclidean distances in S_{C} between focused depths *within* the camera because these distances feature Gaussian noise, refer to Fig. 5. Gaussian noise is certainly not expected from reconstructed actual depths in front of the camera as the relationship in Eq. (10) is highly nonlinear. It is conceivable, however, that a minimization of reprojection errors on raw plenoptic images delivers even more accurate results.

Depth images from plenoptic cameras are also affected by systematic depth errors. The authors in Ref. [9] suggest that these are in part consequence of the Petzval field curvature aberration, which describes a slight change of focal distance for oblique projections. The authors model this distortion in a similar way to the radial lens distortion, affecting projection depth ${}_{\mathrm{C}}z_{\mathrm{f}}$ instead of its lateral position $\{{}_{\mathrm{C}}x,\ {}_{\mathrm{C}}y\}$. Further, the authors introduce a linear dependency of this model w.r.t. the magnitude of the virtual depth. Since this type of distortion is in accordance with the nature of the lens used, we propose a more general depth distortion model:

$$ \begin{aligned} {}_{\mathrm{C}}\tilde{z}_{\mathrm{f,d}} &= {}_{\mathrm{C}}\tilde{z}_{\mathrm{f,u}} + \alpha \cdot (\hat{x}_{\mathrm{C}}/\hat{z}_{\mathrm{C}}) + \beta \cdot (\hat{y}_{\mathrm{C}}/\hat{z}_{\mathrm{C}}) \\ &\quad + \sum_{i=1}^{\infty} (\gamma_i + \delta_i\, {}_{\mathrm{C}}\tilde{z}_{\mathrm{f,u}}) \cdot \left(\sqrt{(\hat{x}_{\mathrm{C}}/\hat{z}_{\mathrm{C}})^2 + (\hat{y}_{\mathrm{C}}/\hat{z}_{\mathrm{C}})^2} \right)^{i}. \end{aligned} \tag{13} $$

Eq. (13) models a skewed paraboloid with factors α and β that parameterize a linear bias dependent on the lateral projection position (*i.e.*, a planar slope, which could be consequence of inner skewness of the camera components), and γ_i and δ_i parameterize the linear dependency of the depth distortion w.r.t. the undistorted, focused projection depth ${}_{\mathrm{C}}\tilde{z}_{\mathrm{f,u}} = (\tilde{v} \cdot \hat{b} + \hat{h})$ and the absolute lateral distance. These parameters can be added to the set of unknown parameters $\hat{\boldsymbol{\Phi}}$. As explained in the next Section 3.1, our experiments using a short focal length lens of 12.5 mm feature a stronger planar distortion w.r.t. the lateral projection position, an underlying parabolic depth distortion component, and a steep higher-degree paraboloid in 7th degree to cope with the strong peripheral distortion. It is known that multi-focus plenoptic cameras work best using long focal lengths and short focus distances [19]. Further research is required in this concern.

3. Results

Next, the implementation of the calibration procedure is described. The parameters are then validated using a commercial range measuring table.

3.1. Calibration

We use a Raytrix R5-C-K color camera featuring 4.2 Megarays and a 12.5 mm Canon lens that yields total focus and depth images with 1 MP each. We choose a low post-processing level for depth images in order not to hallucinate information by using regularization terms. Note that $p = 2 \cdot 5.5\,\mu\mathrm{m}$ is twice the side length of the actual Baumer camera pixels because we are using 1 MP virtual images instead of the raw 4 MP images.

First, eight tilted calibration images including feature ranges between 11 and 55 cm have been taken, see Fig. 3. The total focus images in the top row are processed with either DLR CalDe or DLR CalDe++ [22], which flawlessly detect all 2023 visible corners and assigns them to their known coordinates in S_{O}. *All* of these corners will be used for calibration, *i.e.*, data filtering is not applied. After that, the feature projections are represented around the central point of the image instead of the upper-left corner.

The initialization of the parameters requires the computation of homographies, which is performed on normalized pixel and Euclidean coordinates. In detail, C linear equation systems (one for every image c, $\forall c \in \{1, \ldots, C\}$) are solved in a least-square sense using the single value decomposition. The homographies are then transformed back to their original dimensions by matrix multiplication [23]. The 2-D reprojection root mean square (RMS) error using homographies amounts to 3.3 pixels. Next, we estimate the focal length of the lens using Eq. (8) with the result of $\hat{f} = 12.01$ mm. It is clear that, in the absence of radial distortion correction, the estimated focal length diverges from reality as the apparent scaling factor is compromised. This initial value is, however, valid for subsequent nonlinear optimization of the whole model. Next, the C absolute extrinsics of the camera w.r.t. the calibration plate $\left\{{}_{\mathrm{C}}\boldsymbol{R}_c^{O},\ {}_{\mathrm{C}}\boldsymbol{t}_c^{O}\right\}$ are estimated as explained in Section 2.3. A pinhole

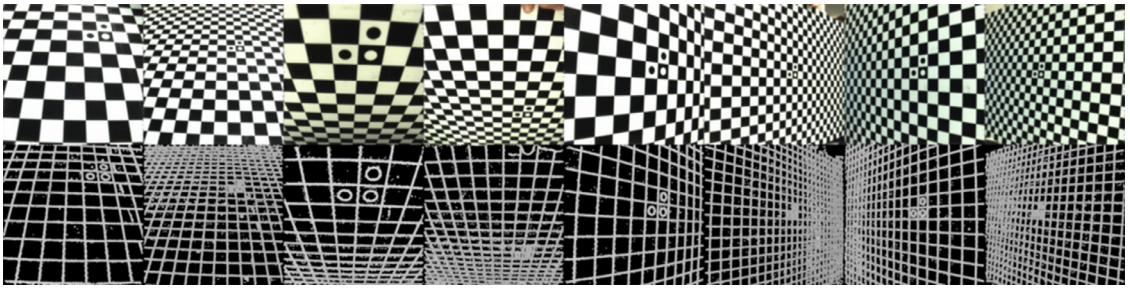
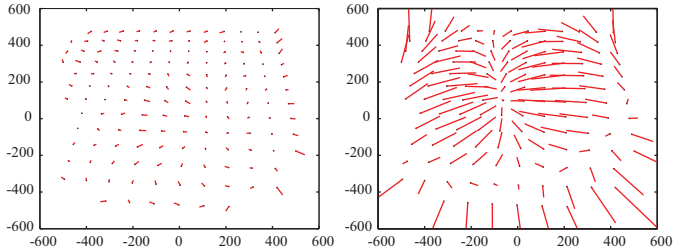
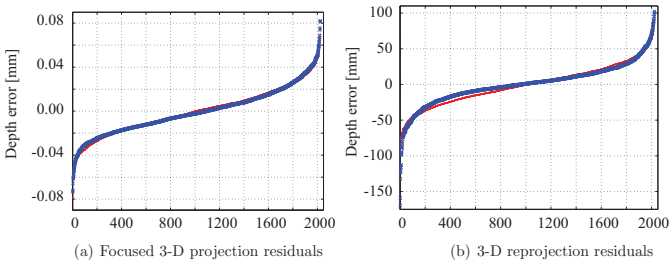


Fig. 3. Total focus brightness calibration images (top) and depth images (bottom).



(a) Stepwise calibration, Stage #1 (b) Joint optimization

Fig. 4. Reprojection residual errors in S_S using the radial lens distortion parameters obtained either following our approach in Section 2.3 (a) or from joint optimization (b). Red arrows are 50×. Our approach yields $\hat{k}_{1\star} = -0.1893$ and $\hat{k}_{2\star} = 0.2020$ with distortion origin at ${}_C\hat{x}_{r\star} = -0.023$ and ${}_C\hat{y}_{r\star} = 0.006$ in z-normalized camera coordinates. The resulting RMS error is 0.39 pixels. The joint optimization yields $\hat{k}_{1\star} = -0.118$ and $\hat{k}_{2\star} = 0.158$ with origin at ${}_C\hat{x}_{r\star} = 0.0258$ and ${}_C\hat{y}_{r\star} = -0.0008$. The resulting RMS error is 3.0 pixels. The inaccuracy after joint optimization renders total focus brightness images virtually useless.



(a) Focused 3-D projection residuals (b) 3-D reprojection residuals

Fig. 5. Ordered set of 2023 residual distance errors (blue crosses) against a Gaussian distribution with the same mean and variance (red dots). Subfigure (a) shows the distribution of distances between focused projections within the camera $\left({}_C\tilde{\boldsymbol{p}}_{f,d} - {}_C\hat{\boldsymbol{p}}_{f,d}\right)$, whereas subfigure (b) shows distances between reprojected (from virtual depths) and estimated (using the parameterized camera and scene models) actual 3-D points in front of the camera $\left({}_C\tilde{\boldsymbol{p}} - {}_C\hat{\boldsymbol{p}}\right)$. The data in (a) follow a Gaussian distribution, thus optimal estimation by leveraging the maximum likelihood criterion is warranted. By contrast, note the long tail at the left-hand side of (b), which means that some depth measurements using noisy virtual depths are markedly beyond ground truth—a known circumstance in stereo vision.

camera model featuring $\hat{f}$ and $\left\{{}_C\hat{\boldsymbol{R}}_c^O,\ {}_C\hat{\boldsymbol{t}}_c^O\right\}$ shows a reprojection RMS error of 3.6 pixels.

At this point the thin lens camera model optimization described in Eq. (9) takes place. The optimization is performed using the `lsqnonlin` method in MATLAB®, a Levenberg–Marquardt implementation. After 14 iterations and 8 s it successfully optimizes the model's parameters yielding a reprojection RMS error of 0.39 pixels. Such a low reprojection residual indicates a highly accurate parametrization and pose estimation—in consideration of the megapixel size of the images and the fact that images are filled with features to the brim. The optimal focal length $\hat{f}_\star$ amounts to 12.76 mm and the lens distortion parameters are detailed in Fig. 4 (a).

By way of contrast, we also implemented a plain (non-iterative) joint optimization of *all* intrinsic parameters (including b and h) using all virtual images. The optimization delivers erroneous radial distortion parameters because they compensate for noisy virtual depth values and a potential depth distortion model featuring radial components, see Fig. 4 (b). This effect can be mathematically interpreted as bad conditioning and/or local minima because a constrained optimization featuring fixed optimal intrinsic values as obtained by our method does yield a lower residual cost. In fact, the authors in Ref. [9] encountered the same difficulties. Therefore, they proposed an empirical optimization approach using iterative, constrained optimization steps within the framework of sequential quadratic programming.

Second, the calibration of the inner lengths of the camera between the MLA and the image plane (b) and between the main lens and the MLA (h) is conducted as in Section 2.4. The initialization of these parameters by solving the system of equations in Eq. (11) uses the optimal depth values ${}_C\hat{z}_{f\star}^{\{c,\,i\}}$ for all features and C calibration images together with the virtual depths $\tilde{v}^{\{c,\,i\}}$ in the bottom row of Fig. 3. It is convenient to filter the depth values in case of missing pixels or noise artifacts. The author in Ref. [12] averages the virtual depth values of a whole image blob. Instead, we opt for using the median value within a radius of 5 pixels around the measured projections ${}_S\tilde{\boldsymbol{p}}^{\{c,\,i\}}$. The values estimated by the linear least-squares method for the distances $\hat{b}$ and $\hat{h}$ are 0.397 and 11.969 mm, respectively.

Next comes the second nonlinear optimization in Eq. (12). This optimization is well conditioned, taking 2 iterations and 1 s. The final values for the inner lengths are $\hat{b}_\star = 0.432$ mm and $\hat{h}_\star = 11.850$ mm (both are actually negative, owing to the formulation choice). The nonzero parameters of the depth distortion model are also estimated: $\alpha = -0.080$, $\beta = -0.044$, $\gamma_2 = -0.127$, $\gamma_7 = -190.03$, and $\delta_7 = 14.82$. Note that we minimize 3-D Euclidean distances between focused projections in S_C within the camera, refer to Fig. 5. The optimization engine used is, again, MATLAB®'s `lsqnonlin`.

3.2. Validation

We collect independent range data in order to validate the last section's results. To this end we use a commercial range measuring table and a known planar calibration pattern, see Fig. 6. The camera is mounted on the vertical axis and is shifted in range from 10 to 90 cm in 80 steps of 1 cm.

Note that it is also possible to estimate range accuracy without an external measuring system. On the one hand, 3-D reprojections are computed from both, virtual depth images (e.g. of the calibration pattern) and the optimal intrinsic parameters, using Eqs. (2) and (3). On the other hand, 3-D estimations of the same features' structure can be computed using total focus brightness images, the local model of the calibration pattern, and the

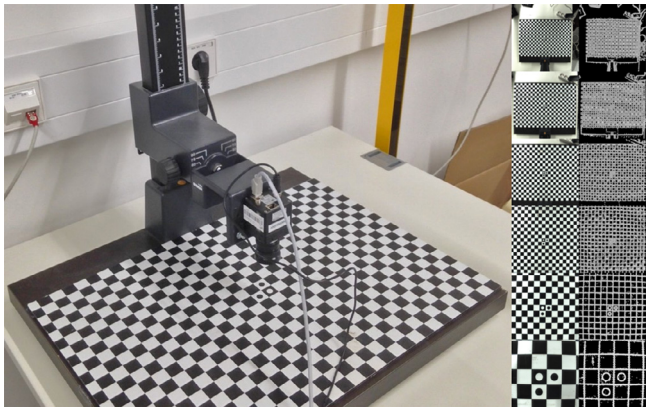
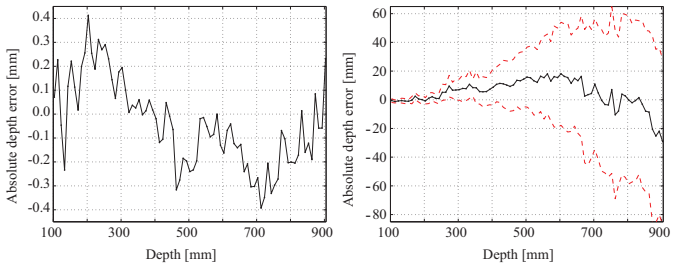


Fig. 6. Left: Camera mounted on the vertical axis of a range measuring table. Right: Some total focus brightness and virtual depth images of the validation set consisting of 81 image pairs and range measurements.



(a) Model-based structure estimation (b) Virtual depth measurements

Fig. 7. Absolute range accuracy (solid) and standard deviation (dashed) w.r.t. registered ground truth for all features and ranges. Since the range measurements are in a different reference frame, these have been registered to S_C using an aleatory, camera-based range datum. We preferably take a datum from model-based structure estimation, as they are sub-millimetrically accurate in the whole validation range (a). The same transformation has been used to register both measurement sets (a) and (b) because they share the same thin lens camera model. Using that transformation and the measurements, the camera origin at S_C (*i.e.*, the optical center of the lens) can be pinpointed within the camera lens at approx. 1 cm distance in front of the sensor chip, exactly as expected. It is worth noting, however, that the thin lens model is just an approximation, whose origin is not a physically distinctive point.

intrinsic parameters estimated in Section 2.3. The latter method is preferably in the form of a nonlinear optimization similar to Eq. (9), now with fixed, optimal intrinsic parameters. The method is known as model-based structure estimation and is highly accurate when using regular pinhole cameras with an adequate angular field of view [24]. Unfortunately, high accuracy is yet to be verified in the context of plenoptic cameras. In this section we also consider the accuracy of model-based structure estimation using total focus brightness images.

Fig. 7 (a) shows the absolute range accuracy w.r.t. registered ground truth of model-based structure estimation using the total focus brightness images of the Raytrix camera, the known model of the calibration plate ${}_O\boldsymbol{p}_i$, and the thin lens camera model parameterized using Section 3.1 including $\hat{f}_\star$, $\hat{k}_{1\star}$, $\hat{k}_{2\star}$, ${}_C\hat{x}_{\mathrm{r}\star}$, and ${}_C\hat{y}_{\mathrm{r}\star}$. Range data shows highest accuracy with a standard deviation of 0.18 mm in the range between 10 and 90 cm—irrespective of the actual depth, which matches the precision of the measuring system. Note that this validation range is larger than the distances used for camera calibration. Therefore, model-based structure estimation using the thin lens camera model that has been parameterized using the method presented in Section 2.3 does indeed yield highly accurate range estimations.

Fig. 7 (b) in turn shows the same absolute range accuracy plot, now using the light-field range measurements reprojected from virtual depths. All corner measurements within an angular field of view of 30° have been used (between 9 and 450 depending on the camera height, cf. Fig. 6). In this way we include the problematic depth distortion model in the validation results. We obtain highly accurate measurements in short range between 10 and 25 cm with an accuracy of 1 mm. From that point through to 90 cm an unsteady bias appears within the range of ± 2 cm. Note that due to the overall unbiased results, these range measurements do not require any bias correction factor, cf. Ref. [9] with a correction factor of 25 cm. These are good accuracy values considering the extended measurement range beyond calibration data and the highly noisy nature of virtual depths.

4. Conclusion

In this work we introduce a novel method for the calibration of focused plenoptic monocular cameras that leverages long-established practice for the calibration of standard monocular cameras [13,14]. We decouple the calibration of the traditional capabilities of plenoptic cameras from the calibration of their novel features related with depth estimation. In this way, the higher noise levels of the latter novel features will not affect the estimation of traditional parameters like the focal length and the radial lens distortion. Further advantages are: First, different robustification methods can be applied to the input data (either total focus or depth images) in accordance to their specific propensity toward outliers. Second, both subtasks are simpler, enabling novel, rapid initialization schemes for all parameters where the only required physical data are the metric size of the sensor elements (pixels) and the local geometry of the calibration pattern. Third, neither the correlated lens and depth distortion models nor the inner lengths f, b and h will get entangled during optimization. In addition, we address particular details on the modeling of this sort of cameras and suggest modifications in the choice of the minimization space of the depth distortion model.

Experiments show the rapid convergence of the approach along with its accuracy on independent ground-truth validation data.

Future work comprises the study of the depth distortion model and the skewness of single camera components with regard to the lens unit used. The inclusion of the geometry of the MLA in the calibration algorithm should be addressed as well as the consideration of the three different types of microlenses used in the MLAs of Raytrix cameras [12]. In addition, a calibration approach based on raw plenoptic images would generalize this formulation to jointly calibrate unfocused and focused plenoptic cameras [29].

References

[1] E. Hecht, Optics, 3rd edition, Addison Wesley, 1998.
[2] E.H. Adelson, J.R. Bergen, The Plenoptic function and the elements of early vision, Comput. Models Vis. Process. 91 (1) (1991) 3–20.
[3] Lytro, Inc.Lytro™ Camera. URL https://www.lytro.com/.
[4] Raytrix GmbH. Raytrix™ R5 Camera. URL http://www.raytrix.de/.
[5] Pelican Imaging Co. Pi Cam™. URL http://www.pelicanimaging.com/.
[6] A. Lumsdaine, T.G. Georgiev, The focused plenoptic camera, in: Proceedings of International Conference on Computational Photography (ICCP), 2009.
[7] C.M. Christensen, The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail, Harvard Business School Press, Boston, Massachusetts, 1997.
[8] D.G. Dansereau, O. Pizarro, S.B. Williams, Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 2013, pp. 1027–1034.
[9] O. Johannsen, C. Heinze, B. Goldluecke, C. Perwass, On the calibration of focused plenoptic cameras, in: Proceedings of GCPR Workshop on Imaging New Modalities, 2013.
[10] T. Luhmann, C. Jepping, B. Herd, Untersuchung zum messtechnischen Genauigkeitspotenzial einer Lichtfeldkamera, in: Publikationen der DGPF, Band 23, Hamburg, Germany, 2014.

[11] N. Zeller, F. Quint, U. Stilla, Calibration and accuracy analysis of a focused plenoptic camera, in: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3, 2014, pp. 205–212.
[12] C. Heinze, Design and Test of a Calibration Method for the Calculation of Metrical Range Values for 3D Light Field Cameras, Fachgebiet Mikroprozessortechnik und Elektronik, Faculty of Engineering, Fachhochschule Westküste and Faculty of Engineering and Computer Science, Hamburg University of Applied Sciences, 2014 Master's thesis.
[13] Z. Zhang, A Flexible new Technique for Camera Calibration, IEEE Trans. Pattern Anal. Mach. Intell. 22 (11) (2000) 1330–1334.
[14] P.F. Sturm, S.J. Maybank, On plane-based camera calibration: a general algorithm, singularities, applications, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Fort Collins, CO, USA, 1999, pp. 432–437.
[15] K.H. Strobl, G. Hirzinger, More accurate camera and hand-eye calibrations with unknown grid pattern dimensions, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Pasadena, CA, USA, 2008, pp. 1398–1405.
[16] K.H. Strobl, G. Hirzinger, More accurate pinhole camera calibration with imperfect planar target, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1st IEEE Workshop on Challenges and Opportunities in Robot Perception, Barcelona, Spain, 2011, pp. 1068–1075.
[17] R.Y. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, IEEE J. Robotics Automation 3 (4) (1987) 323–344.
[18] G.P. Stein, Internal Camera Calibration Using Rotation and Geometric Shapes, Massachussets Institute of Technology, Artificial Intelligence Laboratory, 1993 Master's thesis. AITR-1426.
[19] C. Perwass, L. Wietzke, Single lens 3D-Camera with extended depth-of-field, in: Proceedings of SPIE, Digital Photography VIII, Vol. 8291, 2012, p. 829108.
[20] J. Mallon, P.F. Whelan, Which pattern? Biasing aspects of planar calibration patterns and detection methods, Pattern Recognition Letters 28 (8) (2007) 921–930.
[21] J. Heikkilä, O. Silvén, A Four-step Camera Calibration Procedure with Implicit Image Correction, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Juan, Puerto Rico, 1997, pp. 1106–1112.
[22] K.H. Strobl, W. Sepp, S. Fuchs, C. Paredes, M. Smíšek, K. Arbter, DLR CalDe and DLR CalLab (2005). URL http://www.robotic.dlr.de/callab/
[23] K.H. Strobl, A flexible approach to close-range 3-D modeling, Dissertation, in: Institute for Data Processing, Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, Munich, Germany, 2014.
[24] K.H. Strobl, W. Sepp, G. Hirzinger, On the issue of camera calibration with narrow angular field of view, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), St. Louis, MO, USA, 2009, pp. 309–315.
[25] K.H. Strobl, G. Hirzinger, Optimal hand-eye calibration, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Beijing, China, 2006, pp. 4647–4653.
[26] F. Devernay, O.D. Faugeras, Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured enviroments, Machine Vision and Applications 13 (1) (2001) 14–24.
[27] J.A. Sánchez, E.A. Destefanis, L.R. Canali, Plane-based camera calibration without direct optimization algorithms, in: IV Jornadas Argentinas de Robótica, Córdoba, Argentina, 2006.
[28] J. Weng, P. Cohen, M. Herniou, Camera calibration with distortion models and accuracy evaluation, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (10) (1992) 965–980.
[29] A. Lumsdaine, T.G. Georgiev, G. Chunev, Spatial analysis of discrete plenoptic sampling, in: Proceedings of SPIE, Digital Photography VIII, Vol. 8299, 2012, p. 829909.

Klaus H. Strobl is a research scientist at the Institute of Robotics and Mechatronics of the German Aerospace Center (DLR) in Oberpfaffenhofen, Germany. His research interests focus on computer vision, 3-D graphics, camera calibration, and mobile robotics. Klaus studied electrical engineering (automatic control) at the Universidad de Navarra (Spain), the Vienna University of Technology (Austria), the Technische Universität München (Germany), and the Norwegian University of Science and Technology (Norway). He held a visiting researcher position at the Department of Computing, Imperial College London in 2009 and earned his Ph.D. summa cum laude in electrical engineering in 2014 at Technische Universität München.

Martin Lingenauber is a researcher at the Department of Perception and Cognition of the Robotics and Mechatronics Center, German Aerospace Center (DLR) in Oberpfaffenhofen, Germany. He received his Dipl.-Ing. degree in aerospace engineering in 2010 from the Technical University of Berlin. From 2010 until 2011 he was with the control systems department of the European Space Agency (ESA) in Noordwijk where he worked on testing components for attitude control systems and on new vision sensor concepts. Since 2011 he is with the Institute of Robotics and Mechatronics and works on robot vision for space applications and plenoptic cameras for robotics.