On the Calibration of Focused Plenoptic Cameras

Ole Johannsen[1], Christian Heinze[2], Bastian Goldluecke[1], and Christian Perwaß[2]

[1] Heidelberg Collaboratory for Image Processing, University Heidelberg
[2] Raytrix GmbH, Schauenburgerstr. 116,
24116 Kiel, Germany

Abstract. Plenoptic cameras provide a robust way to capture 3D information with a single shot. This is accomplished by encoding the direction of the incoming rays with a microlens array (MLA) in front of the camera sensor. In the *focused plenoptic camera*, a MLA acts like multiple small cameras that capture the *virtual scene* on the focus plane of a main lens from slightly different angles, which enables algorithmic depth reconstruction. This *virtual depth* is measured on the camera side, and independent of the main lens used. The connection between actual lateral distances and virtual depth, however, does depend on the main lens parameters, and needs to be carefully calibrated. In this paper, we propose an approach to calibrate focused plenoptic cameras, which allows a metric analysis of a given scene. To achieve this, we minimize an energy model based upon the thin lens equation. The model allows to estimate intrinsic and extrinsic parameters and corrects for radial lateral as well as radial depth distortion.

Keywords: focused plenoptic camera, plenoptic 2.0, metric calibration, calibration, depth distortion, Raytrix

1 Introduction

While normal 2D cameras only record the intensity of light at a certain position on the image sensor, plenoptic cameras capture the complete 4D lightfield on position on the image sensor, plenoptic cameras capture the complete 4D lightfield on the sensor plane. The 4D lightfield is an intensity function that not only depends on the position on the imaging plane, but also the incident direction. This additional information allows an algorithmic 3D reconstruction of the captured scene [19, 4, 10, 11].

The idea of plenoptic cameras originates in the early 20th century. First described using a grid of pinholes inside a camera by Ives in 1903 [6], Lippmann in 1908 proposed the use of microlenses in front of the image plane [8]. Emerging from these concepts, a lot of research on plenoptic cameras has been done, varying from setups with multiple independent cameras over such ones with microlenses in front of an image sensor. Several improvements to the design have
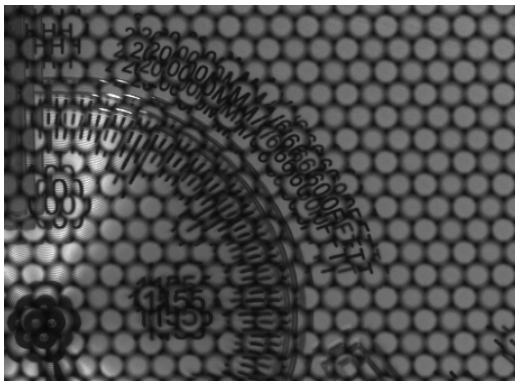
Fig. 1. Detail of a raw image captured by a plenoptic 2.0 camera by Raytrix. Objects closer to the camera are visible in more microlens images. The different types of microlenses can be distinguished in this image by comparing the sharpness of the projections.

been proposed, for example, cameras manufactured by Raytrix[3] employ multiple types of microlenses to accomplish a larger depth of field. Based on the different types of input data, a multitude of different concepts for estimating depth, like EPI stacks [19, 18, 4, 2], focus stacks [10, 11], or multiview stereo approaches [7, 1] has been introduced.

For any kind of 3D camera, it is of interest to be able to metrically measure the depth and determine the extent of a captured object. Hence, a lot of work has been put into calibrating plenoptic cameras. Vaish et al. [16] and Svoboda et al. [15] work on calibrating plenoptic multi-camera arrays while Dansereau et al. [5] deal with the calibration of unfocused lenslet-based plenoptic cameras like the ones commercially available from Lytro[4].

This paper concentrates on the metric depth reconstruction with *focused* lenslet-based plenoptic cameras or *plenoptic 2.0* cameras [9] as commercially available from Raytrix [12] (see figure 3). The idea of depth reconstruction with a focused plenoptic camera is to find the virtual point an object point is focused on by the main lens. Since points at different distances to the main lens are focused upon different distances behind the main lens, a reconstruction of the original 3D scene is possible.

The location a scene point is projected to on the sensor side of the camera does not only depend on the point's distance to the camera, but also on the focal length and focus distance of the main lens. A metric surveying of objects is impossible without knowing the exact parameters of the camera. Our contribution is to introduce a model of how to estimate those parameters to calibrate a plenoptic 2.0 camera.

-----

[3] see `www.raytrix.de`
[4] see `www.lytro.com`

The paper is organized as follows. First, a brief introduction to camera models is given and afterwards, the depth estimation algorithm implemented in a Raytrix camera is described in terms of these models. In section 2.3, distortion models are introduced, which allow for the correction of lens errors. Here we propose a novel type of distortion model, which corrects for distortion in depth dimension. The effect of this type of distortion is that objects located on a plane perpendicular to the optical axis are projected to different virtual depths depending on their position in the frame. Section 3 describes the final calibration model, after which we show results on a number of different camera configurations. Finally, we give an outlook on possible future work.

2   Theoretical Background

The concept behind calibrating a camera lens combination is that multiple images of a known object (e.g. a checkerboard or a dot grid) are taken. The known dimensions of the model and the extracted projections on the sensor are then leveraged to estimate the intrinsic and extrinsic camera parameters. The extrinsic camera parameters describe the position and rotation of the model points in relation to the camera, while the intrinsic camera parameters parametrize the projection through the lens. Additionally, correction parameters like lens distortion must be calculated.

In this chapter, we introduce the theoretical background for these topics. First, a short introduction to camera models is given, which is an important prerequisite to understand depth estimation with a plenoptic camera. Second, we shed some light on lens distortion models and introduce the proposed depth distortion model.

2.1   Camera Models

In order to describe the perspective projections and optical properties of cameras, multiple camera models have been proposed over time. One of the most simple models is the pinhole camera model, which is often used in computer graphics. In this model, points in the object space are projected through the optical center (the pinhole) onto an image plane resulting in a 2D representation of the 3D object space.

While this model is useful for generating 2D images in computer graphics or even for calibrating 2D cameras [20], any depth information is lost by the projection. Hence, a more physically motivated model has to be chosen to describe a plenoptic camera. The thin lens model describes the distance between an object in front of the camera and the virtual point where the light emitted from that object is focused on behind the lens.

Looking at figure 2, one can see rays emerging from the two object points O_1 and O_2. Each ray is bent according to the focal point of the lens and focused upon I_1 and I_2, respectively. The intercept theorem leads to the thin lens equation, which describes the connection between an object at a distance b along the
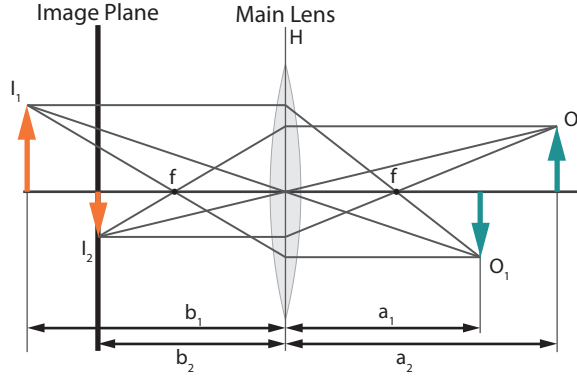
Fig. 2. Thin Lens Camera Model. Object points O_i are projected through the lens and focused to virtual image points I_i behind the lens. If a virtual image point lies upon the image plane, the point is *in focus* and therefore captured sharply, while points away from the focus plane are projected in front or behind the image plane and hence appear blurred. This effect is called depth of field (DOF).

optical axis from the camera, the distance of its focused representation a and the focal length f of the lens as

$$\frac{1}{f} = \frac{1}{a} + \frac{1}{b}. \tag{1}$$

As one can see in figure 2, this model also describes the effect of depth of field (DOF): an object that is at the focus distance of the camera (in this case a_2) is focused precisely onto the image plane resulting in a sharp representation of that point in the image, while an object in front (or behind) that focus plane is focused behind (or in front) of the image plane (O_1 and I_1). Hence, the light of an object point that is out of focus influences a larger area, resulting in a blurred representation of that point.

The thin lens model only holds for lenses whose thickness is negligible in comparison to its focal length. As this must not always be the case, the thin lens model can be extended to the thick lens model. This model includes an offset (the lens thickness) to the model, hence, the thin lens model is a special case of the thick lens model. It can be shown that any combination of lenses – in particular any single lens – can be approximated by the thick lens model, although multiple effects like aberrations or distortion are not included. While with normal camera lenses, the difference between both models is negligible, especially in the field of microscopy the thin lens model can not be used to describe the optics of the lens.

Consequently, the thick lens model would be the perfect model to build the basis for the desired calibration algorithm. Unfortunately, the given information is generally not sufficient to distinguish between the thin lens and the thick lens model, as this requires knowledge of the absolute distance between a point

in front and one behind the camera. Hence, the lens thickness is impossible to estimate by only using calibration targets. For this reason, we will use the thin lens model for the remainder of the paper in order to model the main lens.

If the coordinate system is placed with its origin at the intersection of the main lens axis with the lens plane and the z-axis running along the main lens axis, the thin lens projection can be expressed as matrix multiplication in homogeneous coordinates

$$\tilde{i} = Ao, i = \frac{\tilde{i}}{\tilde{i}_4},$$

with $i = \begin{pmatrix} i_1 \\ i_2 \\ i_3 \\ 1 \end{pmatrix}, \tilde{i} = \begin{pmatrix} \tilde{i}_1 \\ \tilde{i}_2 \\ \tilde{i}_3 \\ \tilde{i}_4 \end{pmatrix}, o = \begin{pmatrix} o_1 \\ o_2 \\ o_3 \\ 1 \end{pmatrix}$ and $A := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{pmatrix}.$ (2)

Above, the point o is the object point in front of the camera, i the virtual point onto which o is projected by the main lens, and A the projection matrix. The multiplication Ao leads to the representation $\tilde{i}$ of i in homogenous coordinates, from which i can easily be recovered.

2.2 Depth Estimation
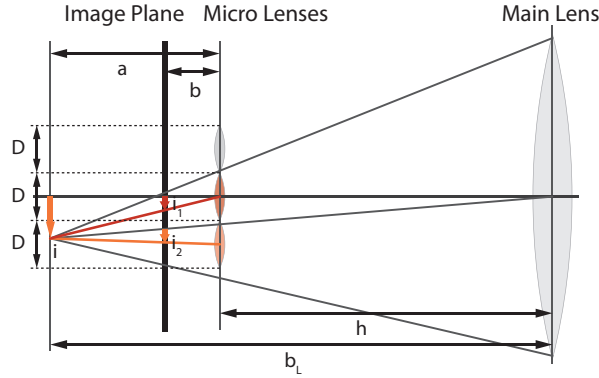
Fig. 3. Schematics of a plenoptic camera. An object point (not shown) is projected through the main lens (right side) onto the virtual image point i. The resulting light cone is split up by the microlenses and focused onto the image plane at i_1 and i_2. Thus, depth estimation can be performed by finding corresponding projections and calculating the parallax.

The image in a plenoptic 2.0 camera is captured through a microlens array, see figure 1. This allows algorithms to estimate the virtual depth of a given scene.
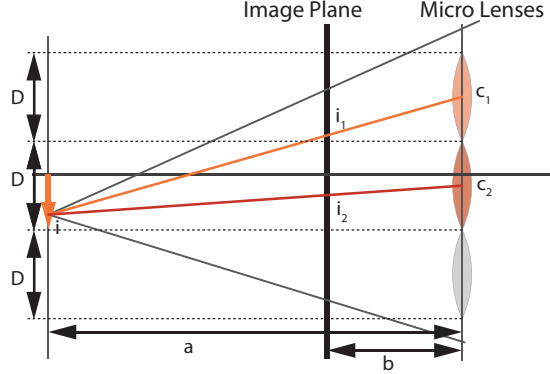
Fig. 4. The Raytrix plenoptic camera estimates depth with correspondence search in the microlens domains. The result of the internal algorithm is a virtual depth $v := a/b$ for each pixel given in multiples of the distance b between microlens array and image plane.

As described by the thin lens model, an object point is focused behind the main lens onto a virtual image point. The cone of rays between this virtual point and the lens is split up by the microlenses and focused at different points onto the image plane, see figure 3. As only the position of these projections is of interest, the pinhole model is sufficient to describe the effect of the microlenses.

Looking at just the microlens array, in figure 4 one can see how the projection i_1 and i_2 arise as the projections of the main lens' virtual image point i onto the image plane. The depth reconstruction works similar to that in stereo cameras: a pair of microlenses is chosen, corresponding points are identified and triangulated. To identify matching projections in multiple lenses, pixel patches are compared. The usually chosen photoconsistency measure is the sum of absolute differences over small pixel patches along the epipolar lines. For this to work, sufficient object structure and image contrast are needed.

We define the *virtual depth* v of the image point i as the distance between i and the MLA, given in multiples of the distance b between the sensor and the MLA, as returned by the depth estimation algorithm of the Raytrix camera. Thus, $v = \frac{a}{b}$, see figure 3. As one can see in figure 4, the virtual depth can be computed from the detected correspondences using the intercept theorem. Assuming the distance $D = \|c_1 - c_2\|$ between the microlens centers c_1 and c_2 to be known, the intercept theorem leads to

$$v := \frac{a}{b} = \frac{\|i - c_1\|}{\|i_1 - c_1\|} = \frac{D}{\|i_1 - i_2\|}. \tag{3}$$

To calibrate this model, we need to estimate this distance b as well as the distance h between the main lens and the MLA. Knowledge of both allows virtual depths to be transformed to metric distances: $b_L = h + v \cdot b$, which can be projected in front of the camera by the thin lens equation (see eq. 1).

The virtual depth can also be used to reconstruct the refocused image [13]. This refocused image is equivalent to a picture a common 2D camera would have taken. Due to the increasing redundancy for higher virtual depths, the relative resolution decreases. As each point has to be projected at least 4 times – twice for every dimension of the image – for depth estimation to work the resolution of the refocused image is one quarter of the raw image.

2.3 Distortion models

Lateral distortion. Distortion describes errors in the geometric projection through a lens. It represents a deviation from the ideal rectilinear projection which maps straight lines to straight lines. Distortion can follow many patterns, but in general it is primarily radially symmetric due to the symmetric design of a lens.

However, in applications the distortion is not always perfectly radially symmetric. Therefore, multiple extensions of the pure radial distortion model have been introduced. One of the most prominent and widely used is Brown's distortion correction model from 1966 [3],

$$
\begin{aligned}
x_{\mathrm{u}} =&(x_{\mathrm{d}} - x_{\mathrm{c}})(1 + k_1 r^2 + k_2 r^4 + \cdots) \\
&+ (p_1(r^2 + 2(x_{\mathrm{d}} - x_{\mathrm{c}})^2) + 2p_2(x_{\mathrm{d}} - x_{\mathrm{c}})(y_{\mathrm{d}} - y_{\mathrm{c}}))(1 + p_3 r^2 + \cdots), \\
y_{\mathrm{u}} =&(y_{\mathrm{d}} - y_{\mathrm{c}})(1 + k_1 r^2 + k_2 r^4 + \cdots) \\
&+ (p_2(r^2 + 2(y_{\mathrm{d}} - y_{\mathrm{c}})^2) + 2p_1(x_{\mathrm{d}} - x_{\mathrm{c}})(y_{\mathrm{d}} - y_{\mathrm{c}}))(1 + p_3 r^2 + \cdots).
\end{aligned}
\tag{4}
$$

Here the point (x_u, y_u) is the undistorted and (x_d, y_d) the distorted image point. The parameters k_i describe the radial distortion and the parameters p_i the tangential distortion. $(x_{\mathrm{c}},\ y_{\mathrm{c}})$ is an offset as the origin of the distortion is not necessarily the center of the image. The radius r is defined by the Euclidean distance to the origin of distortion $\sqrt{(x_{\mathrm{d}} - x_{\mathrm{c}})^2 + (y_{\mathrm{d}} - y_{\mathrm{c}})^2}$.

Brown deducts this formula from the thin prism model, which states that any skew of lenses inside a lens can be obtained by combining a perfect lens which is also perfectly aligned with a thin prism. The distortion model was designed to be an approximation to ray-tracing. In general, the even degrees of the polynomials are predominant. Hence, in most calibration algorithms, only these degrees are taken into account. The distortion model we use includes the coefficients k_1, k_2, p_1, p_2 and p_3.

Distortion in direction of the optical axis. A second kind of distortion is that of distortion in direction of the optical axis which in this paper we term the *depth distortion*. It originates from the Petzval field curvature which describes a slight change of focal length for points at greater distance from the optical axis [14]. This kind of distortion does not effect the position of a point in lateral direction, but the distance it is projected to in depth dimension. Hence, in 2D imaging it would only result in a slight blur at the image corners. But as one can observe in figure 5, the effect leads to a depth distortion for the plenoptic
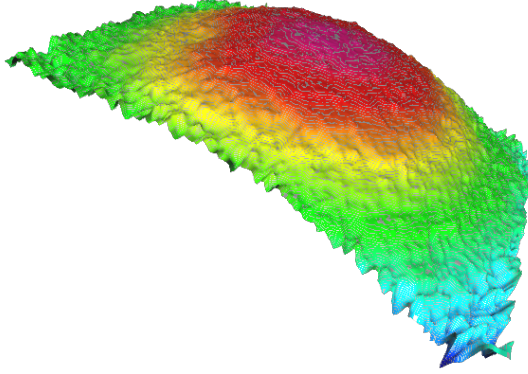
Fig. 5. Color-coded 3D representation of a plane captured by a Raytrix camera with a standard Nikon AF Nikkor 50mm 1:1.8D lens. The depth distortion bends the plane, with the effect becoming stronger with increasing radius from the center. The difference between the estimated depth at the center and the border is about 0.5 virtual depth units, corresponding to approximately 3mm if reprojected in front of the camera.

camera. As this kind of distortion behaves similarly to the lateral distortion, we model it with a structurally similar formula depending on the off-center radius r.

Although in this way, this kind of distortion can be easily compensated for, another aspect has to be taken into account as well. The depth distortion changes over the depth range of the camera. Therefore, the distortion does not only depend on the radius but also on the virtual depth. We think the reason for the change in the depth distortion over the depth range is that the main lens is optimized in a way that the effect of Petzval field curvature is minimal at the image plane, while the focused plenoptic camera calculates depth *behind* this plane. Hence, the main lens correction is not sufficient in this scenario.

Further experimental analysis of the depth distortion suggests that the distortion changes linearly with the virtual depth. Figure 6 visualises this behavior. Thus, we suggest calculating the depth distortion depending on the radius which is linearly adapted by the virtual depth,

$$r' = r \cdot (s_1 + v_d \cdot s_2) \tag{5}$$

$$v_u = v_d + t_1 r' + t_2 r'^2 +_3 r'^4, \tag{6}$$

where v_u is the undistorted and v_d the distorted virtual depth, respectively, while s_i are the radial and t_i the depth distortion coefficients to be determined during calibration.

3   The Calibration Model

The main idea behind the calibration is to capture a target with a known pattern (in our case a dot pattern with a known grid size – see figure 7(a)). These dots
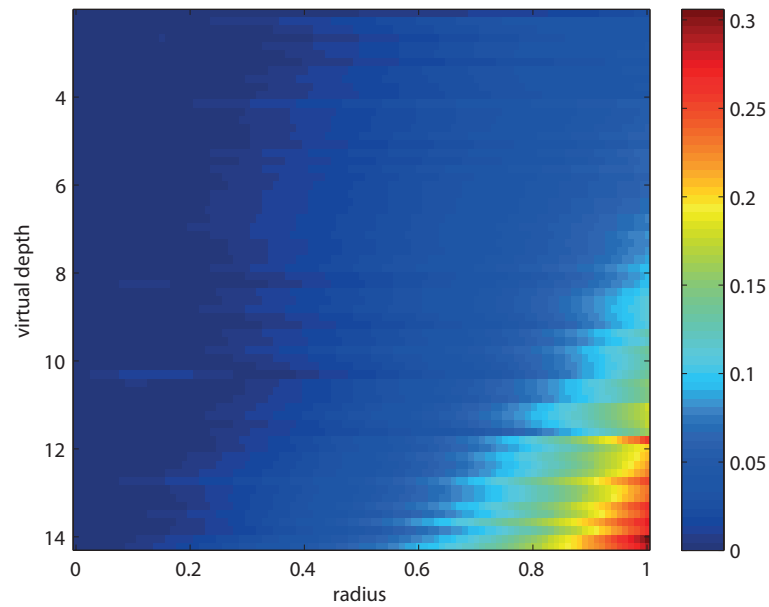
Fig. 6. Depth radial distortion for different virtual depths. The depth distortion increases in magnitude with increasing radius from the image center, but also with increasing virtual depth. The dependence on virtual depth is approximately linear.
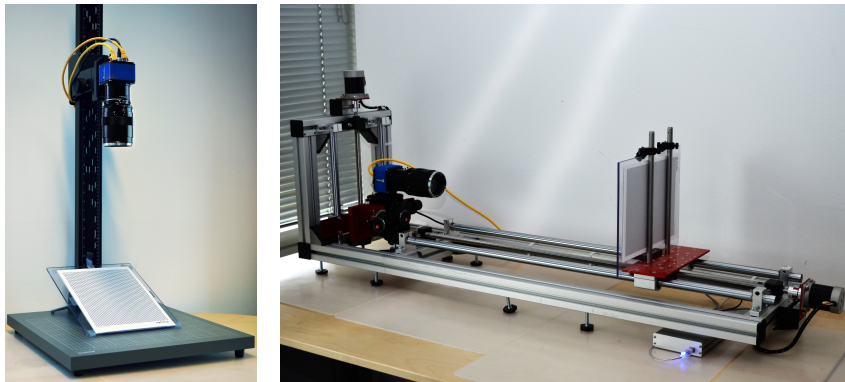
(a) Target capture setup   (b) Linear axis used to measure the ground truth

Fig. 7. Experimental setup to (a) capture individual targets and to (b) measure the ground truth depth of a target plane.

are detected in the image and their virtual depth is calculated. We now need to establish the model parameters such that the detected virtual points are equal to the projection of a model grid with the correct size.

To detect these virtual points first the virtual depth of the given scene and the refocused image have to be calculated as described by Perwaß[13]. The dot pattern has to be detected and positions have to be converted to metric distances. The pixel size is known but due to the different resolution of the depth map and the refocused image it has to be adapted.

As in 2D calibration approaches, the unknown variables can be divided into two groups: the extrinsic and the intrinsic camera parameters, see figure 8. The former describe the position and rotation of the calibration target (or the camera, depending on the point of view), the latter the characteristics of the modeled plenoptic camera like focal length and the distance between the image plane and the modeled lens. A parameter specific to plenoptic cameras is the distance between the microlens array and the image plane (corresponding to the factor between virtual and metric depth). Due to imperfect lenses, we also have to compensate for distortion. For that, the distortion models previously introduced will be used.

The parameters are estimated by minimizing the residual between the rotated, translated and projected model points (which is equivalent to a rotated and translated camera) and the measured image points. The distortion model is applied to the measured points as they are supposed to rectify images. The residual, which depends on the unknowns summarized in figure 8, is given by

$$R = \sum_{j=1}^{C} \sum_{i=1}^{N_j} (dist(p_{ji}) - \pi_A(T_j R_j m_{ji}))^2, \tag{7}$$

where C is the number of targets used for calibration, N_j is the number of points found on target j. For each i, j, the vector p_{ji} represents the measured point, m_{ji} the corresponding model point, R_j the rotation matrix of target j and T_j the translation matrix of that target. The projection π_A is computed according to the thin lens camera model and depends on the projection matrix A, i.e. ultimately only on the focal length f, see equation 2. $dist(\cdot)$ represents both the lateral (see section 2.3) as well as the depth distortion (see equation 6).

As can be seen in figure 8, we need 15 parameters to describe the intrinsic characteristics of the camera and 6 for the position of each target. Hence, for C targets, a total of $6C + 15$ parameters are used.

As the overall energy is non-convex and has local minima, for more robust results, the optimization proceeds in several steps. First, only the initial pose is estimated. For this, the parameters for focal length f, focus distance h and multilens array distance b are initialized with their theoretical quantities from the technical specifications of lenses and cameras. As these values are usually close to the final results, this is a good choice to start the optimization with. However, due to small deviations in manufacturing, they might not be exact, and we find that optimization is still required.

	parameter	model variable	number of unknowns
intrinsic	focal length	f	1
	focus distance	h	1
	distance between MLA and sensor	b	1
	distortion offset	x_c, y_c	2
	lateral distortion	k_i, p_i	5
	depth distortion	s_i, t_i	5
ext.	rotation	R_j	3
	translation	T_j	3
	total		21

Fig. 8. Parameters to be optimized for during calibration.

Afterwards, all model parameters except for the distortion are estimated. If a good initialization for the distance b between the MLA and the sensor is unknown and therefore might be far off, we suggest an iterative approach to determine this parameter: should the change in b be above a certain threshold after optimizing the whole model for the first time, all parameters except b are reset to the initial values. This is suggested especially for higher focal length as the algorithm might otherwise converge to local optima which are far off.

Finally, the distortion parameters are computed. As the undistorted model already tried to (erroneously) compensate for the distortion, we suggest iterating this part with a linear scaling factor to the lateral distortion, which is gradually phased out during optimization iterations.

For all optimization steps above, the Matlab implementation of the sequential quadratic programming (SQP) algorithm was used, which proved to be sufficiently accurate and efficient. Depending on the number of targets and the number of points on each of these the approximate computation time is between one to fifteen minutes.

4   Results

In order to evaluate the calibration, we use a linear axis as shown in figure 7(b) to generate ground truth data. A target showing a random noise pattern is placed perpendicular to the optical axis of the camera and moved along this axis, and we estimate the virtual depth of the captured scene. Due to possible slight skew of the target and depth distortion, a skewed paraboloid was fitted to the data and only the depth measured at the extrema of the paraboloid – i.e. the center of the distortion – was used for the virtual depth. This paraboloid was also used to estimate the depth distortion for different virtual depths. In addition, the actual real-world distance between the image sensor and the target was measured. This way, it is possible to compare the real distance between camera and object to the one obtained from calibrating the camera system.

To compare this ground truth data with the calibration parameters from the optimization, the estimated distances for various virtual depths were reprojected
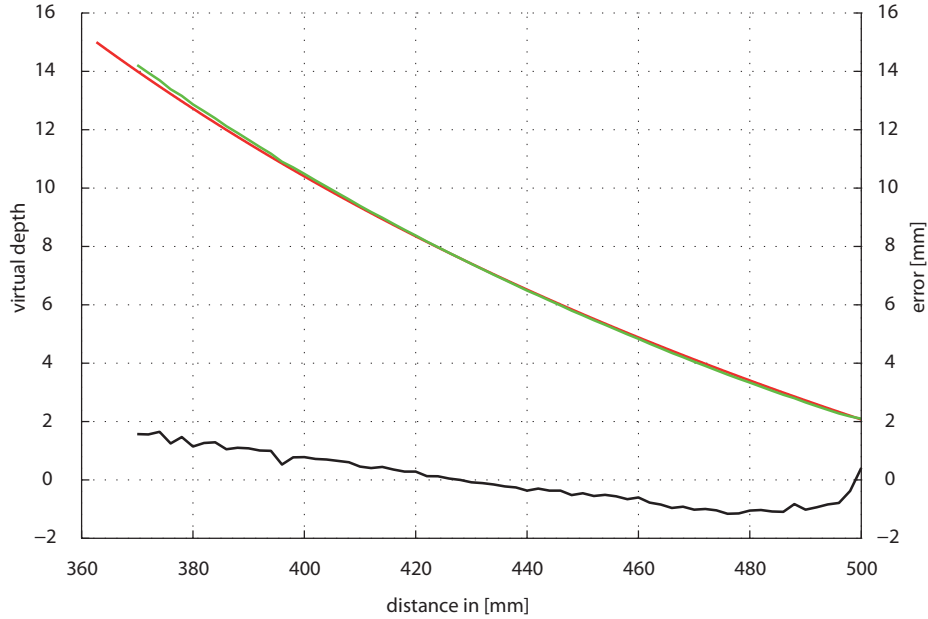
Fig. 9. A comparison shows that the measured ground truth data (green) and the data calculated from the calibration (red) correspond quite well (left axis). The difference between both graphs is plotted in black (right axis). The average relative error is only $0.36mm$, while the absolute error (see text) amounts to $245mm$. Three targets with a total of 2055 points were used for this calibration.

in front of the camera according to the thin lens model and compared to the measured data. We observe that absolute distances can generally not be estimated, which is the expected behaviour. On the one hand, the relation between the virtual depth and the relative change in metric distances is very similar if focal length, focus distance and distance to the object are increased simultaneously. On the other hand, as the lens thickness is neglected, it is not assured that the model can fit reality. Hence, there is a fixed offset of as much as $20cm$ between the calculated and the measured sensor-target distance. This error does not improve by using more points or targets.

However, if we correct for this fixed offset and only consider the relative differences between the various virtual depths, the results fit reality very well. Figure 9 shows both the measured depth curve (green) and the shifted estimated curve (red). Two kinds of errors can be seen. On the one hand there is some noise present in the ground truth. This is due to measurement errors and the standard deviation of the estimation. On the other hand, systematic errors can be seen: the error is not completely random. This might be due to inaccuracies while identifying the marks on the targets, or because of improperly printed targets – we have experimented with this, and an error of 1-2% in the distance between the spots would explain the difference between the two curves quite closely.
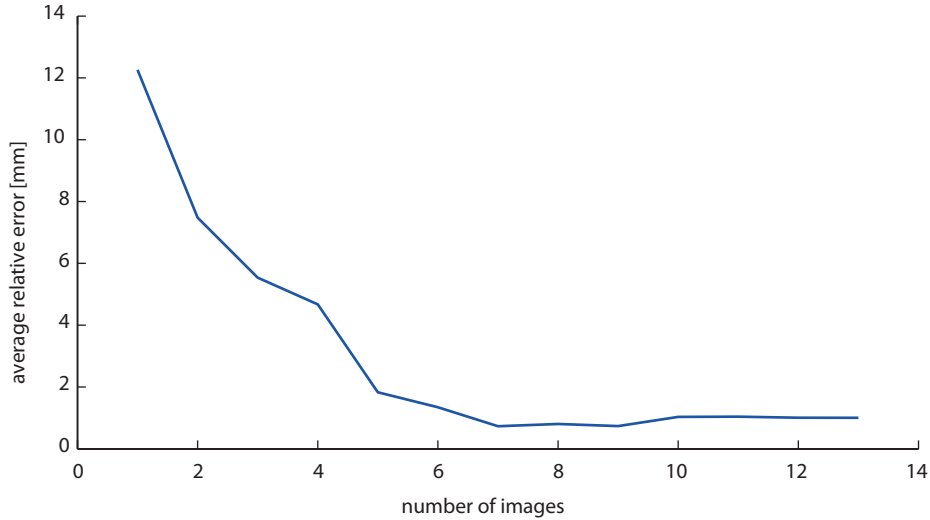
Fig. 10. The number of images of targets used for calibration plotted against the average relative error. The data has been generated using multiple camera lens combinations. Due to only few samples with more than 9 images used the curve is not as monotonous as expected.

We now discuss to what extend other parameters influence the quality of the calibration. First of all, the number of targets used for the calibration plays an important role. The more targets used, the better and more robust the results, see figure 10. This is not surprising, as more data generally works towards reducing the influence of deviations in measurements on the error. More input data may also be generated by using a finer dot grid on the targets. Experiments show that although the quality of the estimate improves with the number of points per target, the number of targets is generally more important. In general, we suggest using at least 3 targets with a few hundred points each.

We have compared multiple lenses on the same camera, and the results for the distance between the sensor and the MLA are robust, that is this distance is approximately the same for all lenses. Another point that is important for calibrating plenoptic cameras is that the actual focal length and focus distance change for different focal settings of the lens. As one can see in figure 11, the effective focal length of a Zeiss Makro Planar T* 100mm f/2 ZE is between $110mm$ and $130mm$ depending on the focus distance. This corresponds to the observed behaviour when mounted on a DSLR.

Furthermore, one can see that the results for the R29 are better than that of the R5. This is due to several reasons. First, because the R29 creates larger and sharper images as well as has a higher depth resolution. Second, the R5's ground truth data is noisier than that of the R29. Third, the wide angle lenses used with the R5 are more difficult to calibrate.

camera	dot spacing	declared on lens		computed from calibration		difference
		focal length	focus distance	focal length	focus distance	
R5M		25	300	22.60	300.63	4.22
		25	500	24.59	552.55	7.21
		50	500	55.93	607.89	1.25
		50	700	51.29	741.58	7.33
		75	900	79.12	1036.24	3.09
R29M	4	50	440	55.39	444.82	1.81
	16	50	440	56.07	450.45	1.27
		50	500	64.11	491.10	1.60
		100	440	107.35	332.85	2.00
		100	500	131.37	508.46	1.13
	4	100	700	119.50	734.66	1.86
	16	100	700	120.15	741.30	1.90

Fig. 11. Declared versus calculated values depending on focal length and focus distance for different cameras and lenses and spacing of the target dots. If no dot spacing distance is given above, the values were averaged over targets with spacing between $2mm$ and $8mm$. All distances are given in millimeters. In addition, the average relative difference between the ground truth and the projected depth curve is given.

Figure 12 shows how depth values on a plane are corrected by the depth distortion estimation. While figure 12(a) shows initial erroneous depth estimates in virtual depth for different virtual depths and radii, correcting this data by the introduced distortion model as shown in 12(b) leads to considerably improved results. There is close to no systematic error present. This is particularly remarkable as the correction works on the whole depth range, while the targets used for the calibration only cover part of it.

5   Conclusion

We have presented a way to compute a metric calibration for plenoptic 2.0 cameras. Based upon the thin lens equation, a quadratic residual was determined, which can be minimized by standard optimization techniques. In addition, distortion is taken into account. Here, a newly developed model was presented, which corrects for depth distortion over varying virtual depths and radii.

We have tested our method on a number of different camera models available from Raytrix. The results show that although the absolute distance between an object and the camera can not be precisely estimated, the relative distances between the virtual depth levels are calculated correctly. The depth distortion model leads to an accurate correction of the data over the complete depth range, even if calibration targets only cover part of it.

In the future, we plan to extend the presented model. Investigating another lateral distortion model might be interesting, which describes the tangential distortion as a tilt of the main lens. Wang et al. [17] show that in terms of quality,

(a) Uncorrected depth distortion.   (b) Corrected depth distortion.
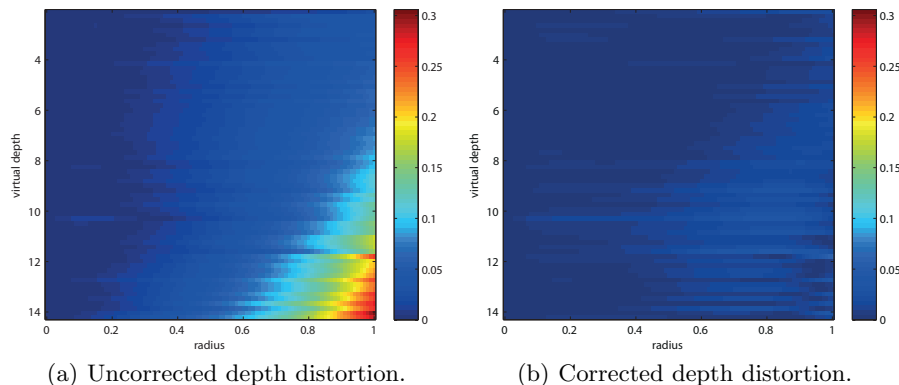
Fig. 12. Depth distortion over different radii and virtual depths. The depth distortion increases with higher virtual depths and radii. The error is given in virtual depth units.

this is equivalent to the standard model. However, due to the information on the tilt of the main lens, the correction of the depth distortion as well as the general depth calibration might be improved.

Another factor which has been neglected are the microlenses itself. First, calibration might improve if done for each type of microlens separately. Second, experiments suggest that distortion within the microlenses influences the depth estimation. Hence, we think that an extension of the model which can correct for the lateral distortion within the microlenses can further improve results.

References

1. Bishop, T., Favaro, P.: The light field camera: Extended depth of field, aliasing, and superresolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(5), 972–986 (2012)
2. Bolles, R., Baker, H., Marimont, D.: Epipolar-plane image analysis: An approach to determining structure from motion. International Journal of Computer Vision 1(1), 7–55 (1987)
3. Brown, D.C.: Decentering distortion of lenses. Photometric Engineering 32(3), 444–462 (1966)
4. Criminisi, A., Kang, S., Swaminathan, R., Szeliski, R., Anandan, P.: Extracting layers and analyzing their specular properties using epipolar-plane-image analysis. Computer vision and image understanding 97(1), 51–85 (2005)
5. Dansereau, D., Pizarro, O., Williams, S.: Calibration and rectification for lenselet-based plenoptic cameras. In: Proc. International Conference on Computer Vision and Pattern Recognition (2013)
6. Ives, F.: Parallax stereogram and process of making same (04 1903)
7. Kolmogorov, V., Zabih, R.: Multi-camera scene reconstruction via graph cuts. In: Proc. European Conference on Computer Vision. pp. 82–96 (2002)
8. Lippmann, G.: Épreuves réversibles donnant la sensation du relief. J. Phys. Theor. Appl. 7(1), 821–825 (1908)

9. Lumsdaine, A., Georgiev, T.: The focused plenoptic camera. In: In Proc. IEEE International Conference on Computational Photography. pp. 1–8 (2009)
10. Nayar, S., Nakagawa, Y.: Shape from focus. IEEE Transactions on Pattern Analysis and Machine Intelligence 16(8), 824–831 (1994)
11. Perez, N., Luke, J.: Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera. In: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video. pp. 1–4. IEEE (2009)
12. Perwass, C., Wietzke, L.: The next generation of photography (2010), `www.raytrix.de`
13. Perwass, C., Wietzke, L.: Single lens 3D camera with extended depth-of-field. In: Proc. SPIE 8291, Human Vision and Electronic Imaging XVII (2012)
14. Riedl, M.: Optical design fundamentals for infrared systems. SPIE Optical engineering Press (2001)
15. Svoboda, T., Martinec, D., Pajdla, T.: A convenient multicamera self-calibration for virtual environments. Presence: Teleoper. Virtual Environ. 14(4), 407–422 (2005)
16. Vaish, V., Wilburn, B., Joshi, N., Levoy, M.: Using plane + parallax for calibrating dense camera arrays. In: Proc. International Conference on Computer Vision and Pattern Recognition (2004)
17. Wang, J., Shi, F., Zhang, J., Liu, Y.: A new calibration model of camera lens distortion. Pattern Recognition 41(2), 607–615 (2008)
18. Wanner, S., Fehr, J., Jähne, B.: Generating EPI representations of 4D light fields with a single lens focused plenoptic camera. Advances in Visual Computing pp. 90–101 (2011)
19. Wanner, S., Goldluecke, B.: Globally consistent depth labeling of 4D light fields. In: Proc. International Conference on Computer Vision and Pattern Recognition. pp. 41–48 (2012)
20. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence (2000)