

Improved Camera Distortion Correction and Depth Estimation for Lenslet Light Field Camera

Changkun Yang, Zhaoqin Liu, Kaichang Di, Yexin Wang, and Man Peng

Abstract

A light field camera can capture both radiance and angular information, providing a novel solution for depth estimation. The paper proposes two improved methods including distortion model optimization and depth estimation refinement for a lenslet light field camera. For distortion model optimization, a novel 14-parameter distortion model that involves sub-aperture images generation is applied to correct the light field camera images. For depth estimation refinement, an algorithm reducing the high influence of outliers on depth estimation in weak texture regions is proposed based on multi-view stereo matching using the cost volume. Experimental results show the projection error has decreased by about 30% and the depth root-mean-squared error on real world images has decreased by about 42% with our distortion correction method and depth estimation method compared with state of art algorithms. It verifies the correctness and effectiveness of our proposed methods and show significant improvement on accuracy of depth map estimation.

Introduction

Light field cameras have become popular in recent years in computational photography, computer vision, and the close range photogrammetry field because they can capture both the radiance and angular information in a single snapshot thanks to a micro-lens array placed between the main lens and sensor. Typical applications include industrial measurement (Heinze *et al.*, 2016), measurement of the growth of plants and animals (Apelt *et al.*, 2015), visual odometry (Dansereau *et al.*, 2011), simultaneous localization and mapping (SLAM) (Dong *et al.*, 2013). Light field cameras can be divided into two categories depending on the distance between the micro-lens array and sensor. In the first category called unfocused plenoptic cameras (Adelson and Wang, 1992; Ng *et al.* 2005), the distance is fixed to be equal to the micro-lens focal length, such as in the commercial products Lytro and Lytro Illum (Lytro, 2017). In the second one called focused plenoptic cameras (Lumsdaine and Georgiev, 2009; Perwaß and Wietzke, 2012), the distance can be changed, such as Raytrix cameras (Raytrix, 2017). In this study, we focus on unfocused plenoptic cameras and use a Lytro Illum camera.

Depth estimation is one of the most important research topics for light field camera image postprocessing. The light field images can be processed for multiple images from different views of the scene, namely sub-aperture images (which will be described in detail in another section). The depth

estimation is based on the disparities observed in the adjacent sub-aperture images, similar to stereo camera approaches. Camera calibration is a necessary prerequisite for accurate depth estimation. A number of methods have been proposed. For the unfocused plenoptic cameras, Dansereau *et al.* (2013) proposed a decoding, calibration, and rectification approach for lenslet light field cameras, in which a 15-parameter camera model was presented for calibration and distortion correction. Cho *et al.* (2013) calibrated a light field camera by searching for local maximization and estimating the rotation of the micro-lens array in the frequency domain based on Dansereau *et al.* (2013). Bok *et al.* (2014) proposed a more accurate calibration method for a micro-lens light field camera based on line features extracted from raw images directly. However, in these calibration algorithms, distortion corrections are all based on the radial distortion model, which does not fit well with the lenslet light field camera. The unfocused plenoptic camera calibration remains to be an important yet challenging task for precision improvement of the subsequent depth estimation. For focused plenoptic cameras, some methods on the metric calibration have been proposed (Heinze *et al.*, 2016; Zeller *et al.*, 2016; Strobl and Lingenauber, 2016), which are beyond the scope of this paper and will not be detailed in the following sections.

Recently a number of depth estimation algorithms for light field images have been proposed. Yu *et al.* (2013) explored the 3D geometry of line in a light field image and derived a disparity map using line matching between sub-aperture images. Wanner and Goldluecke (2013) proposed a local depth estimation algorithm using a structure tensor to compute local slopes in epipolar plane image (EPI). Tao *et al.* (2013) proposed a fusion approach that combined defocus and correspondence cues to estimate the scene depth using EPI, and the global smoothness of depth map was refined by Markov random fields. Tasic *et al.* (2014) proposed a depth estimation algorithm by defining a description of EPI texture and mapping this texture to scale-depth space. Sabater *et al.* (2014) proposed a depth estimation algorithm based on block-matching using the sub-aperture images without demosaicking. Compared with above algorithms, Jeon *et al.* (2015) and Zhang *et al.* (2015) achieved the sub-pixel shifts estimation of sub-aperture images using the phase shift theorem in the Fourier domain to obtain an accurate disparity map. In addition, Kim *et al.* (2013) estimated disparity maps using the 4D light field captured by a digital single lens reflex (DSLR) with movement. Chen *et al.* (2014) introduced a bilateral consistency metric on the surface camera to estimate stereo matches in the light image in the presence of occlusion. However, the baseline of the light field images used in Kim *et al.* (2013) and Chen *et al.* (2014) are much larger than the baseline of

Changkun Yang is with the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences; and also the University of Chinese Academy of Sciences, Beijing 100049, China.

Zhaoqin Liu, Kaichang Di, Yexin Wang, and Man Peng are with the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, No. 20A, Datun Road, Chaoyang District, Beijing 100101, China (liuzq@radi.ac.cn).

Photogrammetric Engineering & Remote Sensing
Vol. 85, No. 3, March 2018, pp. 197–208.
0099-1112/18/197–208

© 2018 American Society for Photogrammetry
and Remote Sensing
doi: 10.14358/PERS.85.3.197

the light field images captured by a lenslet light field camera; thus, their methods are not directly applicable for the lenslet light field camera.

In this paper, two improved methods including distortion model optimization and depth estimation refinement for a lenslet light field camera are proposed so as to achieve higher depth estimation accuracy. First, a novel 14-parameter optical distortion model is applied to correct the aberrations of the lenslet light field camera. Then, a depth estimation algorithm inspired by Jeon *et al.* (2015) is proposed to reduce the influence of outliers on the depth estimation for the light field images. The proposed depth estimation algorithm makes use of the phase shift theorem in the Fourier domain to estimate the sub-pixel shifts of sub-aperture images with an extremely narrow baseline as Jeon *et al.* (2015). Compared with Jeon *et al.* (2015), our depth estimation algorithm divides the center sub-aperture image into strong texture regions and weak texture regions by a strong texture confidence measure. Only costs in strong texture regions are calculated and refined by taking the center sub-aperture image as the guiding image to filter every cost slice to alleviate the coarsely scattered unreliable matches (He *et al.*, 2013). With the cost volume, the multi-label optimization propagates the depth estimation in strong texture regions to the weak texture regions. At last, the disparity refinement (Yang *et al.*, 2007) is used to recover a non-discrete depth map. To verify the effectiveness and accuracy of the proposed algorithm, qualitative and quantitative evaluations have been performed by comparing with Jeon's algorithm using complex real-world scenes. The experimental results show that the proposed algorithm is effective and of higher accuracy on depth estimation using a lenslet light field image.

Geometric Model of Light Field Camera and Improved Distortion Correction

Camera Model and Three-dimensional Measurement Principles

For the traditional camera, its geometric model is based on the collinearity equation. In contrast to a traditional camera, a light field camera has a micro-lens array placed between the main lens and the CCD array. The projection model of a lenslet light field camera is shown in Figure 1. All rays from an arbitrary point passing through the main lens and the micro-lens intersect the CCD array at multiple points. In Figure 1, the rays from the arbitrary point $P(X_c, Y_c, Z_c)$ in the camera coordinate system penetrate the main lens and intersect at (X, Y, Z) , which are the image coordinates in the camera coordinate system. The micro-lens image center (u_c, v_c) is defined as the intersection of the CCD, and a ray that passes through the image center and the micro-lens center. Moreover, the coordinates of a projected point (u, v) in a micro-lens image centered at (u_c, v_c) . The distance L_m is between the main lens and the micro-lens array, the distance L_c is between the main lens and the CCD array, (c_x, c_y) is the principal point.

The projection equation of the lenslet light field camera proposed by Bok *et al.* (2014) is as follows:

$$\begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = \frac{1}{K_1 Z_c + K_2} \begin{bmatrix} f_x X_c - Z_c \dot{u}_c \\ f_y Y_c - Z_c \dot{v}_c \end{bmatrix} \quad (1)$$

where f_x and f_y are the principal distances of the main lens. Moreover, $\Delta u = u - u_c$ and $\Delta v = v - v_c$ are the displacements from the micro-lens image center. Here, $\dot{u}_c = u_c - c_x$ and $\dot{v}_c = v_c - c_y$. Parameters K_1 and K_2 are additional intrinsic parameters calculated as follows (Bok *et al.*, 2014).

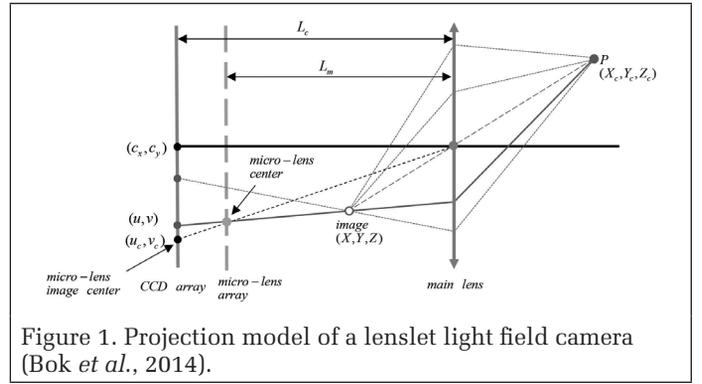


Figure 1. Projection model of a lenslet light field camera (Bok *et al.*, 2014).

$$K_1 \equiv -\frac{(L_m + F)L_c}{(L_m - L_c)F}, \quad K_2 \equiv \frac{L_m L_c}{L_m - L_c} \quad (2)$$

where F is the focal length of the main lens.

It should be noted that the displacements $\Delta u = 0$ and $\Delta v = 0$ for the rays through the center of lens, substituting $\Delta u = 0$ and $\Delta v = 0$ into Equation 1, it becomes the classical collinearity equation:

$$\begin{bmatrix} \dot{u}_c \\ \dot{v}_c \end{bmatrix} = \begin{bmatrix} f_x \frac{X_c}{Z_c} \\ f_y \frac{X_c}{Z_c} \end{bmatrix}. \quad (3)$$

One projection point in the raw image (sensor image) yields one equation in the form of Equation 1, so at least two corresponding points are needed to calculate their 3D coordinates. There are multiple corresponding points in the raw image that come from the same object point when it is out of focus. Therefore, the 3D coordinates of a target point can be computed by Equation 1.

Using the projection equation (1), the theoretical accuracy of three-dimensional measurement using light field images can be analyzed.

From Equation 1, taking two corresponding points in the u -direction as an example, their projection equations are as follows:

$$\begin{aligned} \Delta u_1 &= \frac{1}{K_1 Z_c + K_2} (f_x X_c - Z_c \dot{u}_{c1}) \\ \Delta u_2 &= \frac{1}{K_1 Z_c + K_2} (f_x X_c - Z_c \dot{u}_{c2}) \end{aligned} \quad (4)$$

From Equation 4, we obtain

$$\dot{u}_{c1} - \dot{u}_{c2} = \frac{(\Delta u_2 - \Delta u_1) K_2}{Z_c} + (\Delta u_2 - \Delta u_1) K_1. \quad (5)$$

We define $p = \dot{u}_{c1} - \dot{u}_{c2}$ and $\Delta U = \Delta u_2 - \Delta u_1$ to shorten Equation 5:

$$p = \frac{\Delta U K_2}{Z_c} + \Delta U K_1 \quad (6)$$

From Equation 6, Z_c can be solved by

$$Z_c = \frac{\Delta U K_2}{p - \Delta U K_1} \quad (7)$$

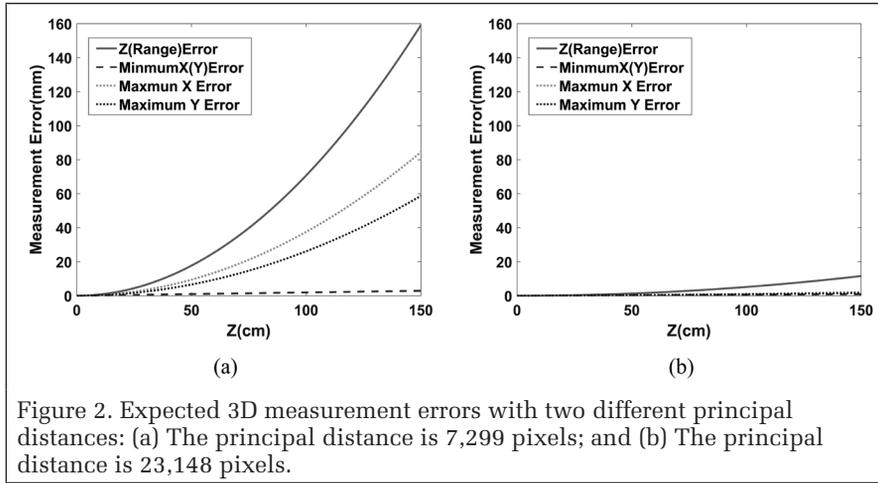


Figure 2. Expected 3D measurement errors with two different principal distances: (a) The principal distance is 7,299 pixels; and (b) The principal distance is 23,148 pixels.

Taking partial derivatives on both sides of Equation 7 and substituting Equation 6 into Equation 7, we get the measurement error in the Z direction as:

$$\sigma_{Z_c} = -\frac{\Delta UK_2}{(p - \Delta UK_1)^2} \sigma_p = -\frac{Z_c^2}{\Delta UK_2} \sigma_p \quad (8)$$

where σ_p is the actual matching error and ΔU is equivalent to the baseline of two sub-aperture images. Equation 8 states that σ_{Z_c} is proportional to Z_c^2 and σ_p but inversely proportional to ΔU . Hence, we can improve the measurement precision in the Z direction by reducing matching error σ_p or increasing ΔU .

From Equation 1, let $\Delta u = \Delta v = 0$. Then, we have:

$$\begin{aligned} X_c &= \frac{\dot{u}_c}{f_x} Z_c \\ Y_c &= \frac{\dot{v}_c}{f_y} Z_c \end{aligned} \quad (9)$$

From Equations 8 and 9, the accuracies (standard errors σ_{X_c} and σ_{Y_c}) of the coordinates in the X and Y directions can be calculated as follows:

$$\begin{aligned} \sigma_{X_c} &= \sqrt{\left(\frac{Z_c^2}{\Delta UK_2}\right)^2 \left(\frac{\dot{u}_c}{f_x}\right)^2 \sigma_p^2 + \left(\frac{Z_c}{f_x}\right)^2 \sigma_{\dot{u}_c}^2} \\ \sigma_{Y_c} &= \sqrt{\left(\frac{Z_c^2}{\Delta UK_2}\right)^2 \left(\frac{\dot{v}_c}{f_y}\right)^2 \sigma_p^2 + \left(\frac{Z_c}{f_y}\right)^2 \sigma_{\dot{v}_c}^2} \end{aligned} \quad (10)$$

where $\sigma_{\dot{u}_c}$ and $\sigma_{\dot{v}_c}$ are the measurement errors of the raw image coordinates in the horizontal and vertical directions, respectively. The measurement errors in X_c and Y_c rely on not only Z_c , but also the position of the object in the raw image. At a given Z_c , the error of X_c is minimized when $\dot{u}_c = 0$, while the largest errors of X_c lie on the left and right margins of the raw image. Similarly, at a given Z_c , the error of Y_c is minimized when $\dot{v}_c = 0$, while the largest errors of Y_c lie on the top and bottom margins of the raw image. Note that the formula for σ_{X_c} , σ_{Y_c} , and σ_{Z_c} is similar to that of the traditional stereo camera (Di and Li, 2007).

Using the standard errors σ_{X_c} , σ_{Y_c} , and σ_{Z_c} , the expected 3D measurement error can be calculated. In order to evaluate the accuracy of depth estimation, we calculated the expected 3D measurement errors for the Lytro Illum camera used in Hazirbas *et al.* (2017), whose principal distance is 7,299 pixels. In

this theoretical analysis, $\Delta U = 1$ pixel, $\sigma_p = 0.14$ pixel, and $\sigma_{\dot{u}_c} = \sigma_{\dot{v}_c} = 14$ pixels, because the size of the raw image is 14 times the size of the sub-aperture image for the Lytro Illum camera. In addition, in order to study the measurement performance in different principal distance, a Lytro Illum camera with the principal distance of 23148 pixels is also analyzed. Figure 2 shows the expected 3D measurement errors for a Lytro Illum camera with the principal distance of 7,299 pixels and the principal distance of 23,148 pixels.

From Equation 8, the expected measurement error σ_{Z_c} is 70.7 mm at 100 cm distance, and σ_{Z_c} is 159.1 mm at 150 cm distance for the Lytro Illum camera with principal distance of 7,299 pixels, as shown in Figure 2a. The expected measurement error σ_{Z_c} is 5.24 mm at 100 cm distance, and σ_{Z_c} is 11.79 mm at 150 cm distance for the Lytro Illum camera with the principal distance of 23,148 pixels, as shown in Figure 2b. Therefore, the expected working range for the light-field camera for measurement purpose in metric range is within 1.5 meters.

Optical Distortion Correction

For the traditional camera, the collinearity equations with additional parameters as radial and decentering distortion is used to calibrate cameras (Brown, 1966; Fryer, 1996). The existing calibration methods (Dansereau *et al.*, 2013; Bok *et al.*, 2014) for the lenslet field camera are also based on the traditional radial distortion model:

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \left(1 + k_1 r^2 + k_2 r^4\right) \begin{bmatrix} x \\ y \end{bmatrix} \quad (11)$$

where (\tilde{x}, \tilde{y}) are the undistorted coordinates, (x, y) are the distorted coordinates, and k_1, k_2 are two radial distortion parameters. Because the radial distortion model is based on a pinhole camera model, in which the camera aperture is described as a point and the rays always pass through the point. For a light field camera, the rays that do not pass through the center of the main lens cannot fit well to the model. Therefore, it is not completely suitable to correct the light field images using the radial distortion model. In this paper, a 14-parameter distortion model that considers aberrations is applied to correct distortions for the lenslet light field camera. It needs to be noted that the decentering distortion parameters p_1 and p_2 are not included in our distortion model. There are two reasons: first, the effect of these parameters is very small to the center sub-aperture image of the light field cameras (Lytro Illum) according to our experiments, only about 0.001 pixels or less in the root-mean-squared error (RMSE) of projection errors. Second, the decentering distortion model does not fit well to the rays and do not pass through the center of the main lens.

The flowchart of the distortion correction strategy is shown in Figure 3. It consists of the following steps: (i) solving interior parameters and exterior orientation parameters by light field camera calibration using Bok's calibration method (Bok *et al.*, 2014), as to be described; (ii) computing undistorted checkerboard corners based on the interior parameters and distorted micro-lens image centers; (iii) getting distorted checkerboard corners by matching between the raw image and the center sub-aperture image; (iv) determining 14 initial distortion coefficients using the undistorted and distorted checkerboard corners based on 14-parameter distortion

model; and (v) solving 14 optimized distortion coefficients by recalibration with distorted checkerboard corners, 14 initial distortion coefficients, interior parameters, and exterior orientation parameters.

Light Field Camera Calibration

The lenslet light field camera is firstly calibrated by Bok's method (Bok et al., 2014) to get the interior parameters ($K_1, K_2, f_x, f_y, c_x, c_y, k_1, k_2$) and exterior orientation parameters R, t . The Camera Calibration Toolbox for MATLAB (https://sites.google.com/site/yun-subok/lf_geo_calib) is used to complete the calibration with the raw images of checkerboard pattern captured from the different distances and orientations. The Lytro Illum camera and an example of a checkerboard pattern raw image from Bok's calibration dataset which are available online (Bok, 2017) are shown in Figure 4. The size of the raw image is 7,7285,368 pixels.

The distortion model used in Bok's calibration method (Bok et al., 2014) is the traditional radial distortion model with distortion coefficients k_1 and k_2 . Using the distortion coefficients, the center sub-aperture image to be described can be corrected, which is used to get distorted checkerboard corners by matching with the raw image to be described.

The interior parameters ($K_1, K_2, f_x, f_y, c_x, c_y, k_1, k_2$) and exterior orientation parameters obtained in this subsection are utilized as fixed values to recalibrate camera to get 14 optimized distortion coefficients.

Sub-Aperture Image Generation

A sub-aperture image is a collection of camera rays that pass through the common point on the main lens. Taking the same pixel underneath each micro-lens, the sub-aperture images from different angles can be obtained. As shown in Figure 5a, the green lines, blue lines and yellow lines come from different angles respectively, each of them can make up a sub-aperture image. Figure 5b is an overhead view of Figure 5a.

In this paper, a new method for the sub-aperture images generation is proposed. The micro-lenses are hexagons and arranged like a honeycomb, but the sub-aperture image is an orthogonal grid. Therefore, the center

(red dots in Figure 5b) of the virtual micro-lenses (red dashed box in Figure 5a) are evenly spaced and aligned horizontally and vertically in CCD array. The first center of the virtual micro-lens is at the coordinate (7.5, 7.5) in the CCD array. The space between these centers is the same size as those of the true micro-lenses (14 pixels). Every virtual micro-lens

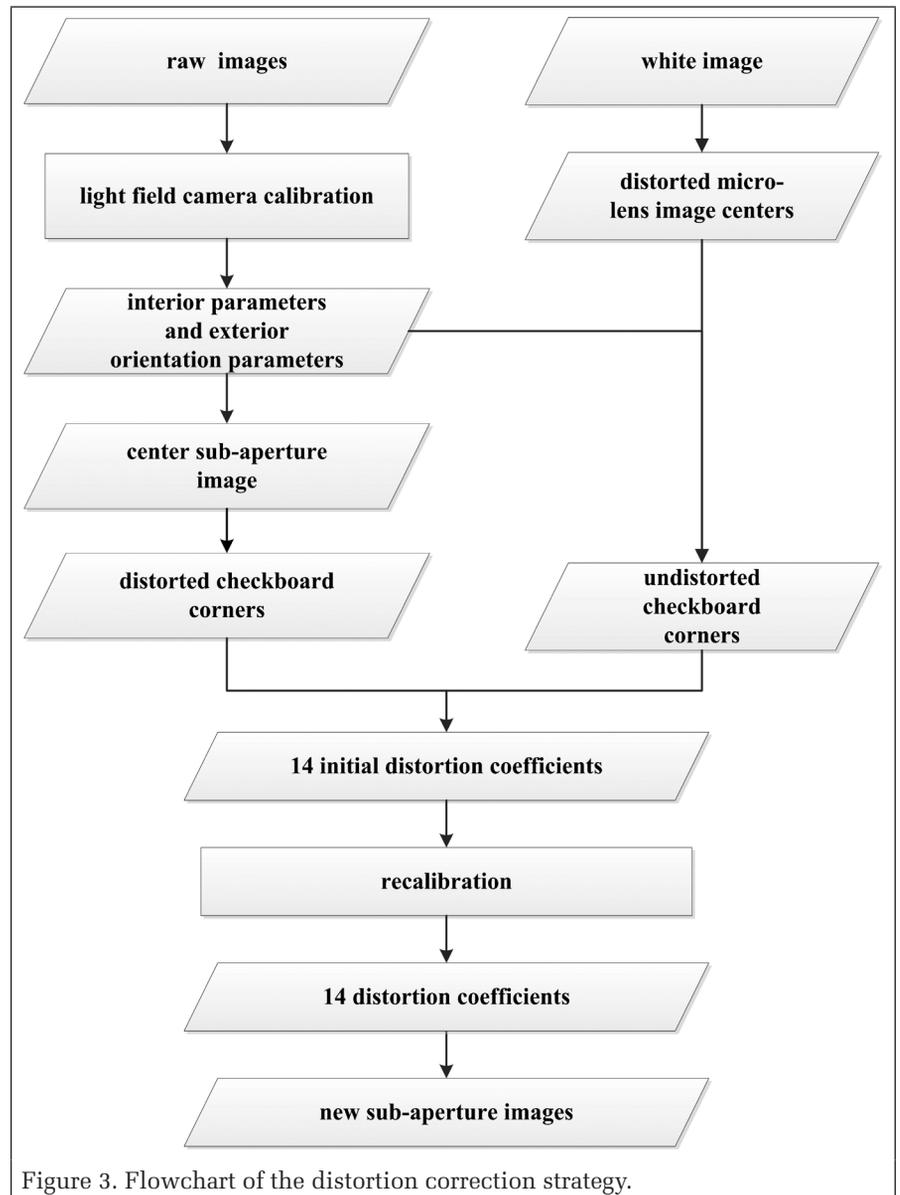
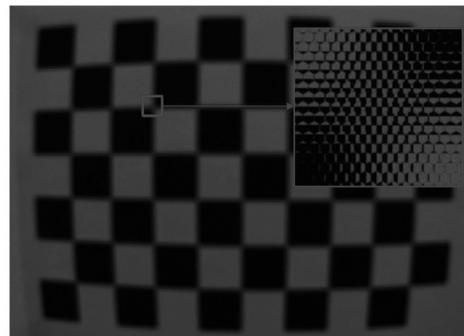


Figure 3. Flowchart of the distortion correction strategy.



(a)



(b)

Figure 4. (a) The Lytro Illum camera, and (b) An example of a raw image from Bok's dataset for calibration

corresponds to a pixel (the red grid in Figure 5b) in a sub-aperture image. Taking the same pixel underneath each virtual micro-lens, sub-aperture images from different angles can be obtained. The corresponding relationship between the coordinate (u_s, v_s) in the sub-aperture image and the virtual micro-lens image center (u_c, v_c) in the raw image is as follows:

$$\begin{aligned} u_c &= (u_s - 1) * s + u_0 \\ v_c &= (v_s - 1) * s + v_0 \end{aligned} \quad (12)$$

where s is a scaling factor whose value equals the space between the centers of two virtual micro-lenses (here, $s = 14$). Moreover, (u_0, v_0) is the coordinate of the first center of the virtual micro-lenses in the CCD array (here, $u_0 = 7.5 \text{ pixel}$, $v_0 = 7.5 \text{ pixel}$).

The value of a pixel underneath virtual micro-lens can be calculated by barycentric interpolation among the pixels with same displacements from adjacent micro-lens image centers underneath three nearest micro-lenses to the virtual micro-lens. For example, as shown in Figure 5a, the value of the green grid underneath the virtual micro-lens (red dashed box) can be calculated by barycentric interpolation among the green pixels underneath three nearest micro-lenses to the virtual micro-lens. In practice, the size of the sub-aperture images, which corresponds to the number of the virtual micro-lenses, can be arbitrary. However, we set the size of the sub-aperture images to be the same as the size of the raw image scaled by the diameter of a real micro-lens (14 pixels).

Acquisition of Undistorted and Distorted Checkboard Corners
The white image shown in Figure 6 is an image taken through a white diffuser, or of a white scene by the light field camera

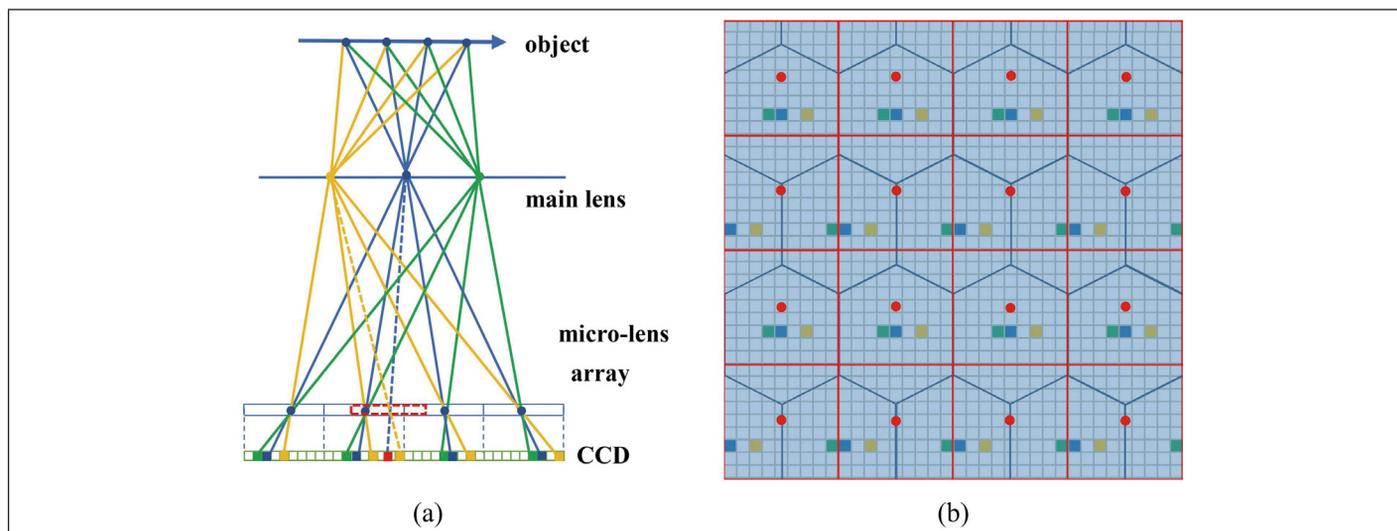


Figure 5. (a) Imaging model of light field camera (side view). The blue rectangles represent micro-lenses. The red dashed box represents a virtual micro lens; and (b) Overhead view of (a). The aqua grid represents the CCD pixels, and the overlaid blue hexagons represent the micro-lenses. The red dots are our virtual micro-lens image centers. The green, blue, and yellow squares represent the pixels with same displacements from adjacent micro-lens image centers, respectively. Each red grid cell represents one pixel in the sub-aperture image.

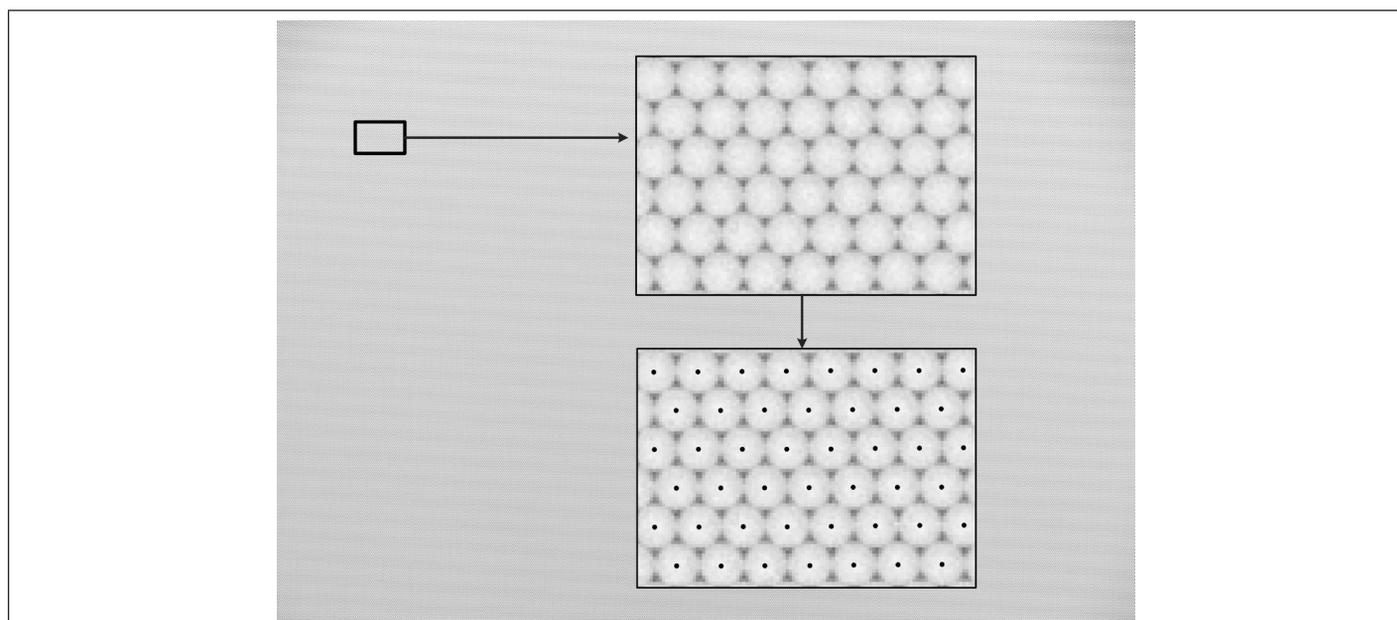


Figure 6. A white image showing the effect of vignetting. The brightest spots in the micro-lens image are extracted as the micro-lens image centers.

(Dansereau *et al.*, 2013). Because of vignetting, the brightest spot in each micro-lens image approximates its center.

The distorted micro-lens images centers can be extracted using the white image (Bok *et al.*, 2014). The undistorted micro-lens images centers are calculated by removing distortion from the distorted checkerboard corners with distortion coefficients k_1 and k_2 . The distorted micro-lens images centers and checkerboard corners on the raw image are shown in Figure 7

The undistorted checkerboard corners (\hat{u}_i, \hat{v}_i) in the raw image can be calculated using the equation:

$$\begin{bmatrix} \hat{u}_i \\ \hat{v}_i \end{bmatrix} = \begin{bmatrix} \hat{u}_{cl} + \Delta u_i \\ \hat{v}_{cl} + \Delta v_i \end{bmatrix} \quad (13)$$

where $(\hat{u}_{cl}, \hat{v}_{cl})$ are the corresponding undistorted micro-lens images centers, and $(\Delta u_i, \Delta v_i)$ are the corresponding displacements of the checkerboard corners from the undistorted micro-lens centers. In order to get $(\hat{u}_{cl}, \hat{v}_{cl})$, the displacements Δu and Δv varied from -4 pixels to 4 pixels in intervals of 0.5 pixels. Then, the corresponding undistorted micro-lens image centers (\hat{u}_c, \hat{v}_c) are computed according to the following equation, which can be got from Equation 1:

$$\begin{bmatrix} \hat{u}_c \\ \hat{v}_c \end{bmatrix} = \begin{bmatrix} \frac{f_x X_c - \Delta u (K_1 Z_c + K_2)}{Z_c} \\ \frac{f_y Y_c - \Delta v (K_1 Z_c + K_2)}{Z_c} \end{bmatrix} \quad (14)$$

where (X_c, Y_c, Z_c) are the 3D coordinates of checkerboard corners in the camera coordinate system, which can be calculated by transforming the world coordinate (X_w, Y_w, Z_w) of the checkerboard corners with the rotation matrix \mathbf{R} and the translation vector \mathbf{t} from the above calibration:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \mathbf{t} \quad (15)$$

The calculated (\hat{u}_c, \hat{v}_c) may not be included in the undistorted micro-lens image centers. Therefore, the undistorted micro-lens image centers nearest to (\hat{u}_c, \hat{v}_c) are chosen as $(\hat{u}_{cl}, \hat{v}_{cl})$. The corresponding $(\Delta u_i, \Delta v_i)$ are calculated by Equation 1. Finally, (\hat{u}_i, \hat{v}_i) are obtained by Equation 13).

The distorted checkerboard corners on the raw image (\hat{u}_a, \hat{v}_a) are extracted by matching the center sub-aperture image with a raw image based on least-squares matching. The checkerboard corners to be matched in the center sub-aperture image are extracted through corner detection. The initial checkerboard corners to be matched in the raw image are derived by adding distortion on the undistorted checkerboard corners (\hat{u}_i, \hat{v}_i) with distortion parameters k_1 and k_2 .

All extracted distorted checkerboard corners (\hat{u}_a, \hat{v}_a) are shown in Figure 8.

Recalibration with the 14-parameter Distortion Model

The light field camera records the information of the rays from multiple directions, and therefore various types of aberrations need to be considered to correct distortions. A geometric model that explicitly describes how an aberrated

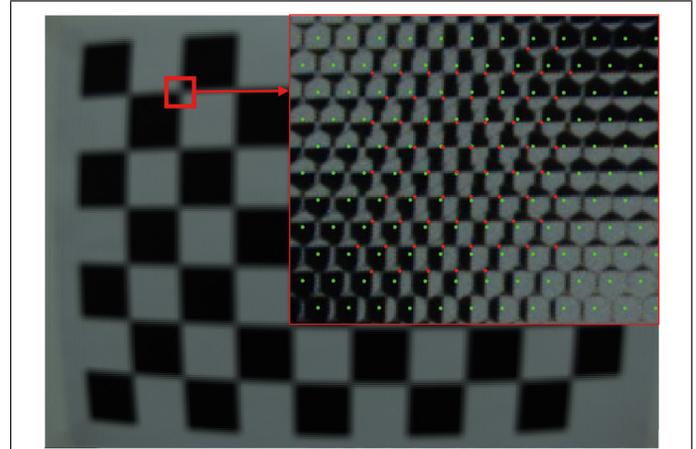


Figure 7. A raw image of a checkerboard pattern. The green dots represent the distorted micro-lens image centers. The red dots represent the distorted checkerboard corners.

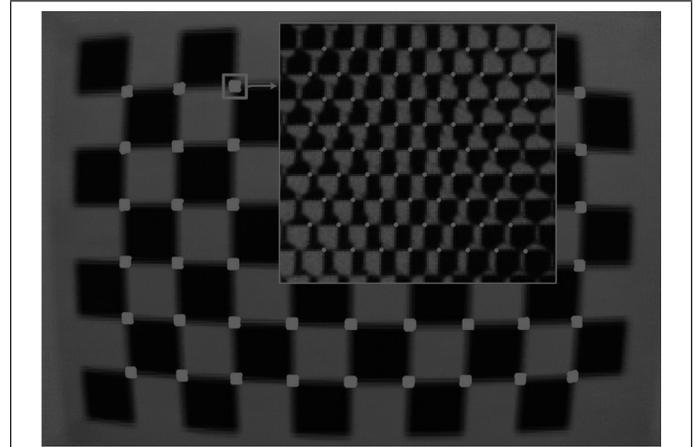


Figure 8. Distorted checkerboard corners in the raw image.

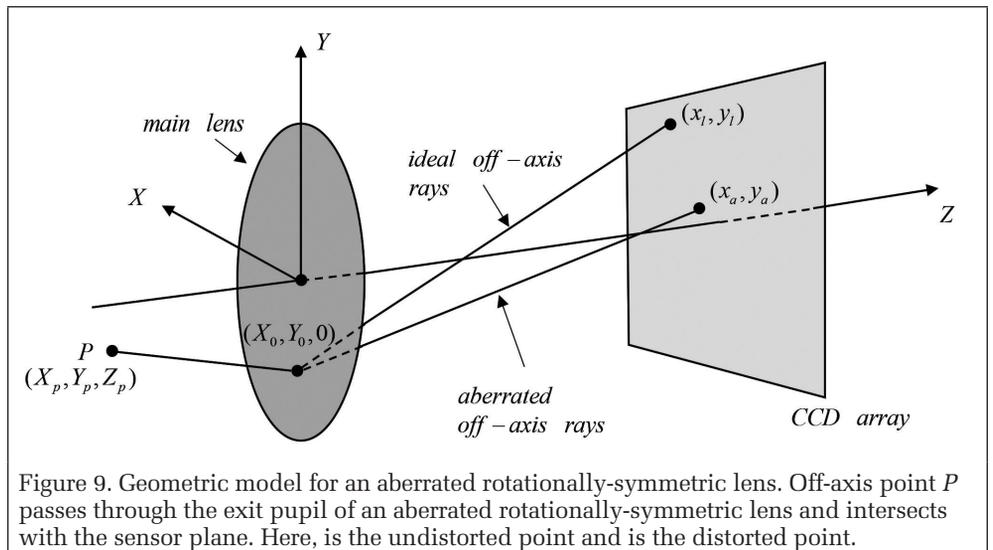


Figure 9. Geometric model for an aberrated rotationally-symmetric lens. Off-axis point P passes through the exit pupil of an aberrated rotationally-symmetric lens and intersects with the sensor plane. Here, (x_i, y_i) is the undistorted point and (x_a, y_a) is the distorted point.

Table 1. Primary aberrations (Mahajan, 1984).

Wave aberration terms	Aberration name
$W_1(e, l) = e l^4$	Spherical
$W_2(e, l) = e e l^2 (e \cdot l)$	Coma
$W_3(e, l) = e e \cdot l l^2$	Astigmatism
$W_4(e, l) = e e l^2 l ^2$	Field curvature
$W_5(e, l) = l ^2 (e \cdot l)$	Distortion

Table 2. Secondary aberrations (Mahajan, 1984).

Wave aberration terms	Aberration name
$W_6(e, l) = e e l^6$	Spherical*
$W_7(e, l) = e e l^4 (e \cdot l)$	Coma*
$W_8(e, l) = e e l^2 (e \cdot l)^2$	Astigmatism*
$W_9(e, l) = (e \cdot l)^3$	Arrows
$W_{10}(e, l) = e e l^4 l ^2$	Lateral spherical
$W_{11}(e, l) = e e l^2 l ^2 (e \cdot l)$	Lateral coma
$W_{11}(e, l) = e e l^2 l ^2 (e \cdot l)$	Lateral astigmatism
$W_{13}(e, l) = e e l^2 l ^4$	Lateral field curvature*
$W_{14}(e, l) = l ^4 (e \cdot l)$	Lateral distortion*

*The word "secondary" is associated with these aberrations, e.g., secondary spherical.

Table 3 Correlation coefficients for the aberration parameters.

Parameter 1	Parameter 2	Correlation coefficient (%)
α_1	α_4	62.93
α_1	α_6	89.19
α_2	α_5	66.16
α_2	α_7	73.70
α_2	α_{14}	57.10
α_3	α_4	51.47
α_3	α_8	45.27
α_3	α_{12}	83.80
α_4	α_6	50.35
α_4	α_{12}	42.57
α_4	α_{13}	81.73
α_5	α_{14}	94.24
α_6	α_{10}	46.23
α_8	α_{10}	55.10
α_9	α_{11}	56.52
α_{11}	α_{14}	42.76
α_{12}	α_{13}	57.45

rotationally-symmetric lens bends a ray is introduced (Figure 9). There are two basic rays: an ideal off-axis ray that describes how a thin lens bends a ray at a general point $(X_0, Y_0, 0)$ on the exit pupil, and the aberrated off-axis ray that passes through the same point on the pupil but deviates from the ideal one.

For a rotationally-symmetric lens, its aberrations include primary, secondary, and tertiary aberrations (Mahajan, 1984). It was found experimentally that the first two aberrations are sufficient to correct the light field image accurately. Aberration $V(e, l)$ can be modeled as a function of the pupil point $e = (X_0, Y_0)$ and the undistorted point $l = (x_l, y_l)$ (Mahajan, 1984). The function can be written as a power series (Tang and Kutulakos, 2013), our 14-parameter distortion model is represented as:

$$V(e, l) = \sum_k a_k \frac{\partial W_k(e, l)}{\partial e} \quad (16)$$

where $a_k (k = 1, 2, \dots, 14)$ are the distortion coefficients, $W_k(e, l)$ are the wave aberration terms.

For primary (or Seidel) aberration, the aberration terms are listed in Table 1.

For secondary (or Schwarzschild) aberrations, the aberration terms are listed in Table 2.

The aberration items $\frac{\partial W_k(e, l)}{\partial e}$ are obtained by taking

the partial derivative of the wave aberration items $W_k(e, l)$ with respect to e . For instance, the term $\frac{\partial W_5(e, l)}{\partial e} = \left(\frac{\partial W_5(e, l)}{\partial X_0}, \frac{\partial W_5(e, l)}{\partial Y_0} \right)$,

where $\frac{\partial W_5(e, l)}{\partial X_0} = x_l(x_l^2 + y_l^2)$, and $\frac{\partial W_5(e, l)}{\partial Y_0} = y_l(x_l^2 + y_l^2)$. The

term $\frac{\partial W_{14}(e, l)}{\partial e} = \left(\frac{\partial W_{14}(e, l)}{\partial X_0}, \frac{\partial W_{14}(e, l)}{\partial Y_0} \right)$, where

$\frac{\partial W_{14}(e, l)}{\partial X_0} = x_l(x_l^2 + y_l^2)^2$, and $\frac{\partial W_{14}(e, l)}{\partial Y_0} = y_l(x_l^2 + y_l^2)^2$.

The coefficients a_5 and a_{14} corresponds to the traditional radial distortion coefficients k_1 and k_2 .

The high correlations between parameters may affect the numerical solution and the estimated parameters (Clarke and Fryer, 1998). We analyze the correlation between the relevant parameters. Correlation coefficients are obtained directly from the covariance matrix using a given set of observation equations and the functional model (Equation 16). The corners used to compute the distortion coefficients using Equation 16 are 37,233 in total from 20 images captured from different distances and orientations. The parameters which showed a high level of correlation are shown in Table 3.

From Table 3, the radial distortion parameters a_5 and a_{14} have the highest correlation coefficients 94.24%. There are different extents of correlation between other parameters. But the high correlation coefficients don't imply that one parameter can dropped without consequence (Clarke *et al.*, 1998). It is found experimentally that RMSE values of projection errors can achieve minimization using all of 14 parameters to correct distortion. Therefore, the 14-parameter distortion model is applied in the paper.

With the 14-parameter distortion model, the field light camera is recalibrated using distorted checkerboard corners, 14 initial distortion coefficients, interior parameters $(K_1, K_2, f_x, f_y, c_x, c_y)$ and exterior orientation parameters.

The initial values of 14 distortion parameters are obtained through minimizing the following objective function:

$$G = \sum_{\min} (x_a - x_l - V_x(e, l)^2 + y_a - y_l - V_y(e, l)^2) \quad (17)$$

where $V(e, l)$ is calculated according to Equation 16. For a lenslet light field camera, the point $e = (X_0, Y_0)$ on the exit pupil can be calculated from the following equation (Bok *et al.*, 2014):

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} \dot{u}_a / f_x \\ \dot{v}_a / f_y \end{bmatrix}, \quad (18)$$

where $(\Delta u, \Delta v)$ are the displacements from the micro-lens image centers. The distorted checkerboard corners $l_a = (x_a, y_a)$ and undistorted checkerboard corners $l = (x_l, y_l)$ are computed using:

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} \dot{u}_a / f_x \\ \dot{v}_a / f_y \end{bmatrix}, \begin{bmatrix} x_l \\ y_l \end{bmatrix} = \begin{bmatrix} \dot{u}_l / f_x \\ \dot{v}_l / f_y \end{bmatrix}, \quad (19)$$

The extracted distorted checkerboard corners in raw image are denoted by $m = [\dot{u}_a, \dot{v}_a]^T$. The coordinates of the checkerboard corners in the world coordinate system is denoted by $m = [X_w, Y_w, Z_w]^T$. If the images used to calibrate are taken from N poses and there are C checkerboard corners in each image used to calibrate, the distortion coefficients $(a_1, a_2, \dots, a_{14})$ can be obtained through an iterative optimization by minimizing the following objective function:

$$\sum_{i=1}^N \sum_{j=1}^C m_{ij} - \check{m}(K_1, K_2, f_x, f_y, c_x, c_y, a_1, a_2, \dots, a_{14}, R_i, t_i, M_j)^2 \quad (20)$$

where \check{m} is the projection of checkerboard corner M_j in image i calculated by adding distortion $V(e, l)$ on the undistorted checkerboard corners (\dot{u}_j, \dot{v}_j) , and (\dot{u}_j, \dot{v}_j) are calculated by Equation 13, $(\dot{u}_{cl}, \dot{v}_{cl})$ in Equation 13 is calculated by removing distortion from the corresponding distorted micro-lens image centers with radial distortion parameters a_5 and a_{14} . The corresponding $(\Delta u_j, \Delta v_j)$ in Equation 13 are calculated by Equation 1. For Equation 20, it is a nonlinear minimization problem, which is solved with the Levenberg-Marquardt algorithm as implemented in More (1978).

Improved Depth Estimation

It is found experimentally that there are a number of outliers on depth estimation in the weak texture region due to the influence of narrow baseline, noise and aliasing for a lenslet light field camera. The proposed depth estimation method is developed based on Jeon's method (Jeon *et al.*, 2015) with a

simple improvement strategy to reduce the influence of outliers on depth estimation. The cost volume in Jeon's algorithm is refined. In our proposed algorithm, the center sub-aperture image is divided into strong and weak texture regions by a strong texture confidence measure. Costs are only calculated in strong texture regions. With the cost volume, the multi-label optimization propagates the correct depth in the strong texture regions to the weak texture regions. At last, the disparity refinement (Yang *et al.*, 2007) is used to recover a non-discrete depth map.

Strong Texture Confidence

The strong texture confidence measure C_e is used to test which regions of the center sub-aperture images the depth estimation seems promising. We define it as:

$$C_e(u, v) = \sum_{u' \in \mathcal{N}(u, v)} |I_c(u, v) - I_c(u', v)| + \sum_{v' \in \mathcal{N}(u, v)} |I_c(u, v) - I_c(u, v')| \quad (21)$$

where $\mathcal{N}(u, v)$ is a 1D window in the center sub-aperture image I_c . The size of the neighborhood can be small (seven pixels in our experiment) as it is supposed to measure only the local color variation.

In order to get more reliable strong texture region, the guided filter (He *et al.*, 2013) is applied to smooth C_e with the center sub-aperture image. Then, a threshold is set to C_e , resulting in a binary confidence mask M_e , which marks the strong texture regions as shown in Figure 10.

Cost Volume Computation

Our algorithm only computes the matching costs for the pixels in the center sub-aperture images with $M_e(u, v) = 1$. The matching costs for the pixels with $M_e(u, v) = 0$ is set to a constant value. The cost volume $C(u, v, l)$ computation is defined as:

$$C(u, v, l) = \alpha C_A(u, v, l) + (1 - \alpha) C_G \quad (22)$$

where $\alpha \in (0, 1)$ is the balancing parameter. The sum of absolute differences $C_A(u, v, l)$ is defined as:

$$C_A(u, v, l) = \begin{cases} \sum_{(u, v) \in \Omega} \min(|I_c(u, v) - I_o(u + d_s, v + d_t)|, \tau_1) & \text{if } M_e(u, v) = 1 \\ \tau_1 & \text{if } M_e(u, v) = 0 \end{cases} \quad (23)$$

where I_c is the center sub-aperture image, I_o are the other sub-aperture images, and τ_1 in the first equation in Equation 23 represents a truncation used for removing the bad matches, $C_A(u, v, l)$ in the second one in Equation 23 one can be arbitrary constant value. Here, it is set to τ_1 . Moreover, Ω is a small square region centered at (u, v) . The sub-pixel shift vectors d_s, d_t of the multi-view sub-aperture images are defined as:

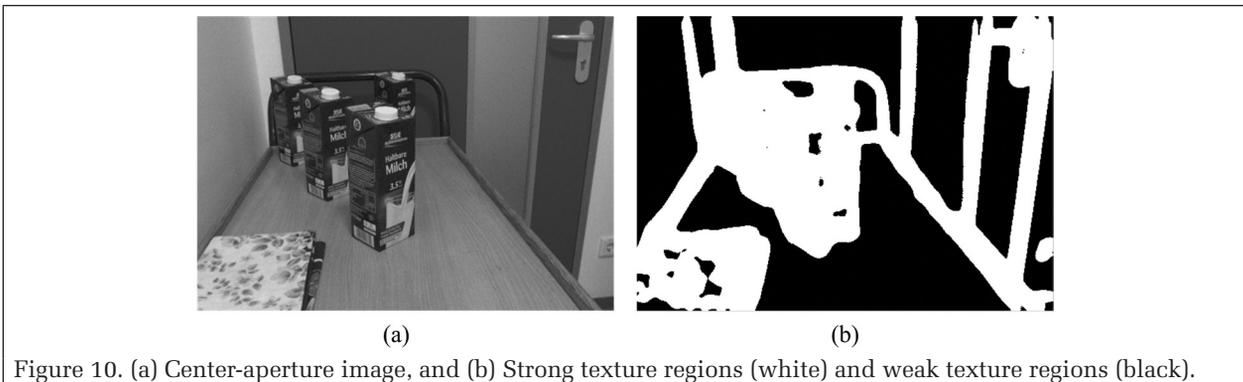


Figure 10. (a) Center-aperture image, and (b) Strong texture regions (white) and weak texture regions (black).

$$\begin{aligned} d_s &= kl(s - s_c) \\ d_t &= kl(t - t_c) \end{aligned} \quad (24)$$

where k is the unit of label in pixels, (s_c, t_c) is the center view, and (s, t) are the other views. The sum of gradient differences C_G is defined as

$$C_G(u, v, l) = \begin{cases} \sum_{(u, v) \in \Omega} \beta \min(|\nabla_u I_c(u, v) - \nabla_u I_o(u + d_s, v + d_t)|, \tau_2) \\ + (1 - \beta) \min(|\nabla_v I_c(u, v) - \nabla_v I_o(u + d_s, v + d_t)|, \tau_2) & \text{if } M_e(u, v) = 1 \\ \tau_2 & \text{if } M_e(u, v) = 0 \end{cases} \quad (25)$$

where $\beta \in (0, 1)$ control the relative importance of two direction gradient differences, and τ_2 in the first equation in Equation 25 represents a truncation used for removing the bad matches, $C_G(u, v, l)$ in the second one in Equation 25 can be arbitrary constant value. Here, it is set to τ_2 . In order to alleviate the coarsely scattered unreliable matches, the guided filter (He *et al.*, 2013) is used in every cost slice to get the refined cost volume $C'(u, v, l)$.

Multi-label Optimization

Multi-label optimization using graph cuts (Kolmogorov *et al.*, 2002) is performed to propagate and correct the disparities. The optimal disparity map is obtained through minimizing

$$l_r = \operatorname{argmin} \sum_{(u, v)} C'(u, v, l(u, v)) + \lambda_1 \sum_{(u, v) \in I} \|l(u, v) - l_a(u, v)\| + \lambda_2 \sum_{(u', v') \in N_c} \|l(u, v) - l(u', v')\| \quad (26)$$

where l is the center sub-aperture image, $C'(u, v, l(u, v))$ is the refined matching cost using guided filter, $\|l(u, v) - l_a(u, v)\|$ is the data fidelity, and $\|l(u, v) - l(u', v')\|$ is the local smoothness, which is used to propagate the correct disparities to the weak texture regions.

Disparity Refinement

After the multi-label optimization, the discrete disparity map is obtained. In order to get a continuous disparity map, the

Table 4. Interior parameters with Bok's method for Bok's dataset.

Parameters	K_1	$K_2(mm)$	$f_x(pixel)$	$f_y(pixel)$	$c_x(pixel)$	$c_y(pixel)$	k_1	k_2
Value	3.373	2428.955	7134.867	7128.613	3842.742	2719.563	0.334	0.105

Table 5. Distortion coefficients with our method for Bok's dataset.

Coefficient	Value	Coefficient	Value
α_1	7.4569e-6	α_8	-5.6999e-6
α_2	2.9391e-5	α_9	8.2009e-6
α_3	-3.2600e-5	α_{10}	-2.1190e-6
α_4	5.3997e-5	α_{11}	-7.5619e-6
α_5	-0.3228	α_{12}	1.0090e-4
α_6	-1.3753e-6	α_{13}	3.9430e-5
α_7	-5.6920e-6	α_{14}	0.1730

Table 6. Comparisons of reprojection errors based on Bok's (left) and our (right) models using Bok's dataset (unit: pixel).

Δv	Δu				
	-3	-2	0	2	3
-3	0.255 0.151	0.207 0.154	0.177 0.164	0.195 0.158	0.236 0.142
-2	0.218 0.154	0.185 0.157	0.167 0.154	0.174 0.168	0.196 0.159
0	0.199 0.161	0.185 0.156	0.168 0.137	0.175 0.166	0.182 0.172
2	0.230 0.159	0.203 0.164	0.184 0.161	0.187 0.163	0.216 0.155
3	0.278 0.163	0.236 0.163	0.203 0.163	0.224 0.158	0.270 0.154

method proposed by Yang *et al.* (2007) is applied. The continuous disparity l^m is calculated as:

$$l^m = l - \frac{C(l_+) - C(l_-)}{2(C(l_+) + C(l_-) - 2C(l))} \quad (27)$$

where $l_+ = l + 1$ and $l_- = l - 1$ are the adjacent cost slices of l , l is the discrete disparity with the minimal cost, and l^m is the disparity with the minimum cost, which is derived from the least square quadratic fitting over three costs: $C(l_-)$, $C(l)$ and $C(l_+)$.

Experimental Results

Distortion Correction Result

After calibrating the camera using raw images, the sub-aperture images can be generated based on the interior parameters and the distortion parameters. The exterior orientation parameters from the calibration result transform the checkerboard coordinate to the camera coordinate. The transformation from the camera coordinate to micro-lens image centers coordinate is computed using Equation 14. The transformation from micro-lens image centers coordinate to sub-aperture coordinate is computed using the inverse transformation of Equation 12). Corner features are extracted from sub-aperture images independently, and the root-mean-squared error (RMSE) values of projection errors can be computed. To verify the effectiveness of our distortion model, we applied our correction method and Bok's method (Bok *et al.*, 2014) on the Lytro Illum calibration dataset provided by Bok (2017). The calibration result obtained by Bok's method and our method for Bok's datasets are shown in Tables 4 and 5. Table 6 shows the comparisons of reprojection errors based on Bok *et al.* (2014) and our models using Bok's dataset.

In addition, we apply our correction method and Bok's method (Bok *et al.*, 2014) on our dataset to further verify the effectiveness of our distortion model. Our dataset is captured using a Lytro Illum camera at distances from 210 mm to 250 mm, which includes 12 images of a checkerboard pattern whose grid size is 7.22 mm. The camera is calibrated using Bok's geometric calibration toolbox (Bok, 2017), and the ray re-projection error in 3D is 0.088 mm. The calibration result is shown in Table 7. The distortion coefficients obtained by our method for our datasets are shown in Tables 8 and 9 and show the comparisons of reprojection errors based on Bok *et al.* (2014) and our models using our dataset.

From Table 6 with Table 9, the projection errors of the sub-aperture images corrected by our method are clearly smaller. Especially, the reprojection errors of the sub-aperture images with larger displacement are significantly improved using our correction method. The RMSE in Table 6 for Bok's and ours is 0.208 and 0.155 pixel, respectively. The RMSE in Table 9 for Bok's and ours is 0.635 and 0.424 pixel, respectively. The average projection error for the two datasets has decreased by about 30% with our correction method.

Depth Estimation Result

The effectiveness of our distortion correction method and depth estimation

method is qualitatively and quantitatively evaluated on the Hazirbas' dataset (Hazirbas et al., 2017), which is a real-scene indoor benchmark composed of 4D light-field images captured by Lytro Illum camera and with ground truth depth obtained from a registered RGB-D sensor. In order to verify the effectiveness of our distortion model, the sub-aperture images corrected by Bok's (Bok et al. 2014) and our methods are used to estimate depth with the state-of-the-art algorithm developed by Jeon et al. (2015). Jeon's source code is released on the website (https://sites.google.com/site/hgjeoncv/home/depthfromlf_cvpr15). Moreover, in order to verify the effectiveness of our depth estimation method, we have implemented it to compare with the method developed by Jeon et al. (2015) by using the same sub-aperture images corrected by our correction method. We fix the user-defined parameters for a fair comparison for the different correction method and dedicate to obtain the best results by optimizing parameters for different depth estimation method. The code is running in MATLAB on an Intel i7 3.4 GHZ and 24 GB memory PC. The parameters experimentally perform well with $\alpha = 0.5$, $\tau_1 = 0.5$, $\tau_2 = 0.5$, label_number=40. The parameters λ_1 and λ_2 are empirically selected as $\lambda_1 = 2$, $\lambda_2 = 0.009$ for Bok's and our correction method, $\lambda_1 = 4$, $\lambda_2 = 0.01$ for our depth estimation method. The absolute error map $|d_{estimate} - d_{groundtrue}|$ and the corresponding error distribution histogram (the error of greater than 500 mm is set to 500 mm) are shown in Figure 11. Bok's plus Jeon's represents Bok's correction method and Jeon's depth estimation method. Ours plus Jeon's represents our correction method and Jeon's depth estimation method. Ours plus depth represents our correction method and our depth estimation method.

In order to compare the results in detail, parts of the images in the red and green rectangles shown in the error maps are magnified in Figure 11. Comparing the red squares in Figure 11b and 11c in the three scenes, it can be seen that the depth errors in the strong texture areas decrease using our correction methods, which reflects that our correction method is better than Bok's in correcting the distortions. Comparing to the green squares in Figure 11b and 11c using depth estimation method of Jeon *et al.*, we can find that using our method, the depth errors shown in Figure 11d decrease also in the weak texture regions. This is because our depth estimation method reduces the effects of outliers in the weak texture regions. Accordingly, the reduction in depth error due to our distortion correction method and the error reduction caused by our depth estimation method are both reflected in the error distribution histogram. As can be seen from the histograms shown in Figure 11, the concentrations trend of the bins toward smaller error in the error distribution histograms is obvious by using our methods. From histograms in Figure 11b

Table 7. Interior parameters derived by the calibration method (Bok et al., 2014).

Parameters	K_1	$K_2(mm)$	$f_x(pixel)$	$f_y(pixel)$	$c_x(pixel)$	$c_y(pixel)$	k_1	k_2
Value	-147.167	26708.272	23148.233	23116.376	3879.220	2710.382	0.686	-0.717

Table 8. Interior parameters with our method for our dataset.

Coefficient	Value	Coefficient	Value
a_1	-1.5343e-5	a_8	1.6164e-6
a_2	1.3736e-5	a_9	8.7593e-6
a_3	2.1034e-4	a_{10}	-1.9066e-5
a_4	7.3973e-4	a_{11}	3.1570e-5
a_5	-0.3484	a_{12}	0.0030
a_6	-2.7023e-6	a_{13}	-0.0013
a_7	-1.2063e-6	a_{14}	1.7863

Table 9. Comparisons of reprojection errors based on Bok's (left) and our (right) models using our dataset (unit: pixel).

Δv	Δu				
	-3	-2	0	2	3
-3	0.686 0.443	0.585 0.484	0.595 0.396	0.663 0.493	0.734 0.521
-2	0.691 0.460	0.607 0.389	0.511 0.337	0.593 0.379	0.717 0.522
0	0.686 0.370	0.605 0.422	0.277 0.274	0.519 0.408	0.649 0.432
2	0.662 0.364	0.618 0.348	0.572 0.401	0.567 0.343	0.708 0.547
3	0.791 0.477	0.685 0.383	0.551 0.397	0.632 0.521	0.781 0.373

Table 10. RMSE of the absolute error (unit: mm).

	Bok's+Jeon's	Ours+Jeon's	Ours+depth
Scene 1	232.38	200.27	82.89
Scene 2	220.97	181.24	128.40
Scene 3	122.91	118.04	95.72

Table 11. Median of the absolute error (unit: mm).

	Bok's+Jeon's	Ours+Jeon's	Ours+depth
Scene 1	203.72	147.50	62.66
Scene 2	171.50	88.30	81.06
Scene 3	101.11	84.46	73.12

to 11c, c, and from 11c to 11d, the changes of the bins concentrations verify the effectiveness of error reduction of both our distortion correction and depth estimation methods.

The quantitative evaluations of our distortion correction and depth estimation methods are performed by computing RMSE and median of the absolute error. The RMSE for the above three scenes are shown in Table 10. The median of the absolute errors for the above three scenes are shown in Table 11.

From the middle two columns of Table 10, we can find that the magnitude of RMSE reduction is not prominent, the average RMSE reduction of three scenes is about 12%; this is because the distortion correction only reduces the depth error in the strong texture region. The last two columns of Table 10 show that with the same distortion correction method, our depth estimation method improves the accuracy of depth estimation in weak texture regions, resulting in the obvious reduction of depth errors (about 42% for the average RMSE reduction of three scenes). The images we used in Hazirbas' datasets are captured at the range of 80~150 cm. The RMSE for our depth estimation method act in accordance with the theoretical analysis in Figure 2a, which verifies the correctness of our method. From Table 11, we can find the depth errors move toward smaller values using our correction and depth estimation methods, which is consistent with movement trend in the errors distribution histograms shown in Figure 11. All the experiment results demonstrate the effectiveness of our

proposed distortion correction and depth estimation methods.

Though our methods produce obvious improvements on depth estimation. There are still some small regions with large depth errors in our depth estimation results. For example, from Figure 11d in scene 1, we can find that the depth errors within the region shown in the blue square are large. This is because that the region is occluded all around and the correct depth cannot be propagated into it. From Figure 11d in scene 2, the large error within the region shown in the blue square are mainly resulted from the influence of the shadow. The effects caused by occlusion and shadow are our future research work.

Conclusions

In this paper, we proposed a distortion correction and depth estimation method for lenslet light field cameras. The improvements include two aspects: (i) a novel 14-parameter distortion model was used to correct the light field camera images and was demonstrated to be superior to the traditional two parameter radial distortion model; and (ii) a depth estimation method is proposed to reduce the impact of outliers in weak texture regions. In addition, a new algorithm was proposed for sub-aperture image generation and the theoretical measurement error equation for the lenslet light field camera is derived. Distortion correction and depth estimation experiments were carried out to validate the feasibility and effectiveness of our proposed method. The experimental results show that our method outperforms the state-of-art depth estimation method with respect to accuracy.

According to the theoretical measurement analysis and experimental results, the unfocused plenoptic camera is suitable for measurement within a couple of meters at a millimeter level precision. In the future, with the foreseeable development of the light field camera, e.g., by increasing the number of micro-lenses and/or increasing the size of CCDs, lens improvement, we believe that it will be widely used in more applications.

Acknowledgments

This research is funded by National Natural Science Foundation of China (Project no. 41471388), National Basic Research Program of China (Project no. 2013CB733202), and National Key Research and Development Program of China (Project no.2016YFB0502102). The authors thank Yunsu Bok and

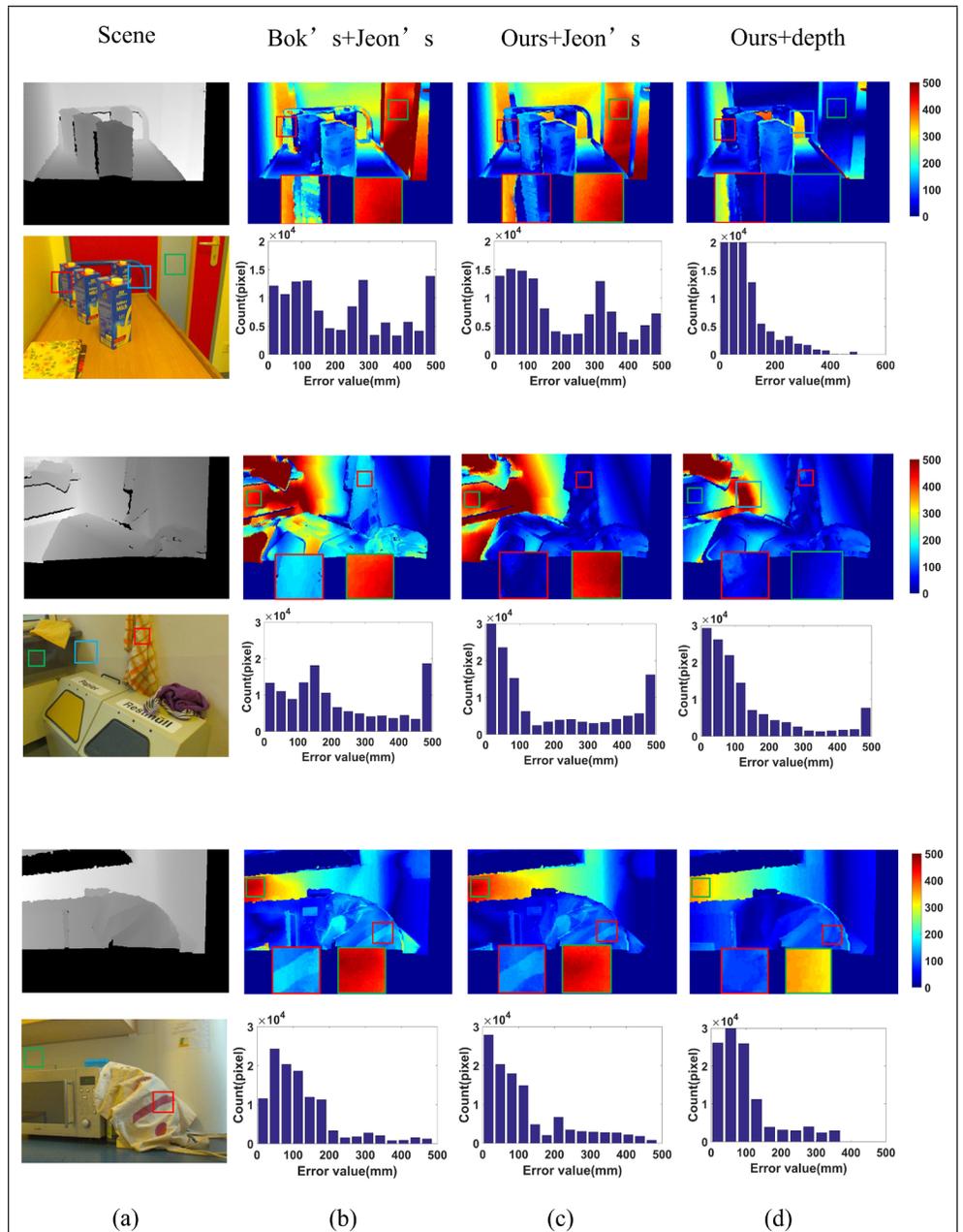


Figure 11. Absolute error map (unit:mm) and error distribution histograms. Every two rows represent a separate scene. For each two rows: (a) center sub-aperture image (bottom) and ground truth depth map (top); (b-c) Absolute error maps (top) and error distribution histograms (bottom) from Bok's and our correction methods both using Jeon's depth estimation method; and (d) Absolute error map (top) and error distribution histogram (bottom) using our correction method and our depth estimation method.

Hae-Gon Jeon of the Korea Advanced Institute of Science and Technology (KAIST) for help with their projection model and for making the calibration toolbox available on the Web.

References

- Ackermann, F., 1984. Digital image correlation: Performance and potential application in photogrammetry, *The Photogrammetric Record*, 11(64):429-439.
- Adelson, E.H, and J.Y.A. Wang, 1992. Single lens stereo with a plenoptic camera, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99-106.

- Apelt, F., D. Breuer, Z. Nikoloski, M. Stitt, and F. Kragler, 2015. Phytotyping4D: a light-field imaging system for non-invasive and accurate monitoring of spatio-temporal plant growth, *The Plant Journal*, 82(4): 693-706.
- Bok, Y., Yunsu Bok's Homepage, URL: https://sites.google.com/site/yunsubok/lf_geo_calib (last data accessed: 03 October 2018).
- Bok, Y., H.G. Jeon, and I.S. Kweon, 2014. Geometric calibration of micro-lens-based light-field cameras using line features, *Proceedings of the European Conference on Computer Vision*, 05-12 September, Zurich, Switzerland, pp. 47-61.
- Brown, D.C., 1966. Decentering Distortion of Lenses, *Photogrammetric Engineering*, V. XXXII, Number 3, pp. 444-462.
- Chen, C., H. Lin, Z. Yu, S.B. Kang, and J. Yu, 2014. Light field stereo matching using bilateral statistics of surface cameras, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 23-28 June, Columbus, Ohio, pp. 1518-1525.
- Cho, D., M. Lee, S. Kim, and Y.W. Tai, 2013. Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 01-08 December, Sydney, NSW, Australia, pp. 3280-3287.
- Clarke, T.A., and J.G. Fryer, 1998. The development of camera calibration methods and models, *The Photogrammetric Record*, 16 (91):51-66.
- Clarke, T.A., X. Wang, and J.G. Fryer, 1998. The principal point and CCD cameras, *The Photogrammetric Record*, 16(92):293-312.
- Dansereau, D.G., I. Mahon, O. Pizarro, S.B. Williams, 2011. Plenoptic flow: Closed-form visual odometry for light field cameras, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 25-30 September, San Francisco, California, pp.4455-4462.
- Dansereau, D.G., O. Pizarro, and S.B. Williams, 2013. Decoding, calibration and rectification for lenselet-based plenoptic cameras, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 23-28 June, Portland, Oregon, pp. 1027-1034.
- Di, K., and R. Li, 2007. Topographic mapping capability analysis of Mars exploration rover 2003 mission imagery, *Proceedings of the 5th International Symposium on Mobile Mapping Technology (MMT 2007)*, Padua, Italy, May 28-31. (CD-ROM).
- Dong, F., S.H. Ieng, X. Savatier, R. Etienne-Cummings, and R. Benosman, 2013. Plenoptic cameras in real-time robotics, *The International Journal of Robotics Research*, 32(2):206-217.
- Fryer, J.G., 1996. Camera Calibration (K.B. Atkinson, editor), *Close Range Photogrammetry and Machine Vision*, Bristol, Whittle Publishing, pp. 156-179.
- Hazirbas, C., L. Leal-Taixé, and D. Cremers, 2017. Deep depth from focus, *arxiv preprint arXiv: 1704. 01085*.
- He, K., J. Sun, and X. Tang, 2013. Guided image filtering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397-1409.
- Heinze, C., S. Spyropoulos, S. Hussmann, and C. Perwass, 2016. Automated robust metric calibration algorithm for multifocus plenoptic cameras, *IEEE Transactions on Instrumentation and Measurement*, 65(5):1197-1205.
- Honauer, K., O. Johannsen, D. Kondermann, and B. Goldluecke, 2015. A dataset and evaluation methodology for depth estimation on 4D light fields, *Proceedings of the Asian Conference on Computer Vision*, 21-23, Nov 2015, Taipei, Taiwan, pp. 19-34.
- Jeon, H.G., J. Park, G. Choe, J. Park, Y. Bok, Y.W. Tai, and I.S. Kweon, 2015. Accurate depth map estimation from a lenslet light field camera, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 07-12 June 2015, Boston, Massachusetts, pp. 1547-1555.
- Kim, C., H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, 2013. Scene reconstruction from high spatio-angular resolution light fields, *Proceedings of the ACM Transactions on Graphics (ACM SIGGRAPH)*, 32(4),73-1.
- Kolmogorov, V., and H. Zabih, 2002. Multi-camera scene reconstruction via graph cuts, *Proceedings of the European Conference on Computer Vision (ECCV)*, 28-31 May, Copenhagen, Denmark, pp. 8-40.
- Liang, C., and K., Ramamoorthi, 2015. A light transport framework for lenslet light field cameras, *ACM Transactions on Graphics (TOG)*, 34(2):16.
- Lin, H., C. Chen, S.B. Kang, and J. Yu, 2015. Depth recovery from light field using focal stack symmetry, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 07-13 December 2015, Santiago, Chile, pp. 3451-3459.
- Lumsdaine, A., and T. Georgiev, 2009. The focused plenoptic camera, *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, 16-17 April, San Francisco, California, pp. 1-8.
- Lytro camera, URL: <http://www.lytro.com> (last date accessed: 03 October 2018).
- Mahajan, V.N. 1984. *Optical Imaging and Aberrations: Ray Geometrical Optics*, SPIE press, Bellingham, Washington, 484 p.
- More, J., 1978. The Levenberg-Marquardt algorithm: Implementation and theory, *Numerical Analysis*, Springer, Berlin, Heidelberg, pp.105-116.
- Ng, R., M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, 2005. Light field photography with a hand-held plenoptic camera, *Computer Science Technical Report (CSTR)*, 2(11):1-11.
- Perwaß, C., and L. Wietzke, 2012. Single lens 3d-camera with extended depth-of-field, *Proceedings of Human Vision and Electronic Imaging XVII*, Burlingame, California.
- Raytrix 3d light field camera technology, URL: <http://www.raytrix.de/>. (last date accessed: 03 October 2018).
- Sabater, N., M. Seifi, V. Drazic, et al., 2014. Accurate disparity estimation for plenoptic images. *Proceedings of the European Conference on Computer Vision*, Springer International Publishing, 05-12 September, Zurich, Switzerland, pp. 548-560.
- Strobl, K., M. Lingenauber, 2016. Stepwise calibration of focused plenoptic cameras, *Computer Vision and Image Understanding*, 145:140-147.
- Tao, M.W., S. Hadap, J. Malik, and R. Ramamoorthi, 2013. Depth from combining defocus and correspondence using light field cameras, *Proceedings of International Conference on Computer Vision (ICCV)*, 01-08 December 2013, Sydney, NSW, Australia, pp. 673-680.
- Tosic, I., and K. Berkner, 2014. Light field scale-depth space transform for dense depth estimation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 23-28 June, Columbus, Ohio, pp. 435-442.
- Tang, H., and K.N. Kutulakos, 2013. What does an aberrated photo tell us about the lens and the scene?, *Proceedings of International Conference on Computational Photography (ICCP)*, 19-21 April. Cambridge, Massachusetts, pp. 1-10.
- Wanner, S., and B. Goldluecke, 2013. Variational light field analysis for disparity estimation and super-resolution, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):606-619.
- Yang, Q., R. Yang, J. Davis, and D. Nistér, 2007. Spatial-depth super resolution for range images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 17-22 June, Minneapolis Minnesota, pp.1-8.
- Yu, Z., X. Guo, H. Ling, A. Lumsdaine, and J. Yu, 2013. Line assisted light field triangulation and stereo matching, *Proceedings of International Conference on Computer Vision (ICCV)*, 1-8 December, Sydney, NSW, Australia, pp. 2792-2799.
- Zeller, N., C.A. Noury, F. Quint, C. Teulière, U. Stilla, and M. Dhome, 2016. Metric calibration of a focused plenoptic camera based on a 3D calibration target, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science*, Volume III-3, XXIII ISPRS Congress, Prague, Czech Republic, pp. 449-456.
- Zhang, Z., Y. Liu, and Q. Dai, 2015. Light field from micro-baseline image pair, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 07-12 June, Boston, Massachusetts, pp. 3800-3809.