A Generic Multi-Projection-Center Model and Calibration Method for Light Field Cameras

Qi Zhang [ID], Chunping Zhang, Jinbo Ling, Qing Wang [ID], *Member, IEEE*, and Jingyi Yu

Abstract—Light field cameras can capture both spatial and angular information of light rays, enabling 3D reconstruction by a single exposure. The geometry of 3D reconstruction is affected by intrinsic parameters of a light field camera significantly. In the paper, we propose a multi-projection-center (MPC) model with 6 intrinsic parameters to characterize light field cameras based on traditional two-parallel-plane (TPP) representation. The MPC model can generally parameterize light field in different imaging formations, including conventional and focused light field cameras. By the constraints of 4D ray and 3D geometry, a 3D projective transformation is deduced to describe the relationship between geometric structure and the MPC coordinates. Based on the MPC model and projective transformation, we propose a calibration algorithm to verify our light field camera model. Our calibration method includes a close-form solution and a non-linear optimization by minimizing re-projection errors. Experimental results on both simulated and real scene data have verified the performance of our algorithm.

Index Terms—Multi-projection-center (MPC) model, light field cameras, two-parallel-plane (TPP) representation, calibration

---

1 INTRODUCTION

THE micro-lens array (MLA) based light field cameras, including conventional light field camera [1] and focused light field camera [2], can capture radiance information of light rays in both spatial and angular dimensions, i.e., 4D light field [3], [4]. The data from light field camera is equivalent to narrow baseline images of traditional cameras with coplanar projection centers. The measurement of same point in multiple directions allows or strengthens the applications in computational photography and computer vision, such as digital refocusing [5], depth estimation [6], segmentation [7] and so on. Recent work also proposed the methods on light field registration [8] and stitching [9], [10] to expand the field of view (FOV). To support these applications, it is crucial to accurately calibrate light field cameras and establish exact relationship between the ray space and 3D scene.

It plays an important role to build a model for describing the ray sampling pattern of light field cameras. Previous approaches have dealt with imaging models on light field cameras in different optical designs [11], [12], [13], [14], [15]. The common points are based on the fact that the microlens is regarded as a pinhole model and the main-lens is described as a thin-lens model. However, some of open issues still remain in the models and methods. First, the

proposed models focus on angular and spatial information of rays, but the relationship between light field and 3D scene geometry is not explored. Second, very little work has considered a generic model before to describe light field cameras with different image formations [1], [2]. Third, existing intrinsic parameters of light field camera models are either redundant or incomplete such that corresponding solutions are neither effective nor efficient.

In the paper, we first propose a multi-projection-center (MPC) model based on two-parallel-plane (TPP) representation [3], [4]. Then we deduce the transformations between 3D scene geometry and 4D light rays. Based on geometry transformations in the MPC model, we characterize various light field cameras in a generic 6-intrinsic-parameter model and present an effective intrinsic parameter estimation algorithm. Experimental results on both virtual (simulated data) and physical (Lytro, Illum and a self-assembly focused) light field cameras have verified the effectiveness and efficiency of our model.

Our main contributions have three aspects, including

(1) We deduce the transformations to describe the relationship between light field and scene structure.
(2) We describe light field cameras with different image formations as a generic 6-parameter model without redundancy.
(3) We propose an effective intrinsic parameter estimation algorithm for light field cameras, including a closed-form linear solution and a nonlinear optimization.

The remainder of the paper is organized as follows. Section 2 summarizes related work on the models of light field cameras and calibration methods. Section 3 introduces our MPC model and the transformations between the 3D structure and 4D light field. Based on the theory of light field parameterization, a generic 6-intrinsic-parameter light

---

- *Q. Zhang, C. Zhang, J. Ling, and Q. Wang are with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China. E-mail: nwpuqzhang@gmail.com, {724495506, 312620756}@qq.com, qwang@nwpu.edu.cn.*
- *J. Yu is with ShanghaiTech University, Shanghai 200031, China. E-mail: jingyi.udel@gmail.com.*

field camera model is proposed. Section 4 provides the details of our calibration method and analyzes computational complexity of the closed-form solution. In Section 5, we present extensive results on the simulated and real scene light fields, demonstrating more accurate intrinsic parameter estimation than previous work [11], [13].

2 RELATED WORK

To acquire 4D light field, there are various imaging systems developed from traditional camera. Wilburn et al. [16] present a camera array to obtain light field with high spatial and angular resolutions. Classic calibration approach is employed for the camera array [17]. More general, in traditional multi-view geometry framework, multiple cameras in different poses are defined as a set of unconstrained rays, which is known as Generalized Camera Model (GCM) [18]. The ambiguity of the reconstructed scene is discussed in traditional topics [19]. However, such applications on the camera array are limited by its high cost and complex control. In contrast, the MLA enables a single camera to record 4D light field more conveniently and efficiently, though the baseline and spatial resolution are relatively smaller than camera array. Compared to the camera array, multiple projection centers of MLA-based light field camera are aligned on a plane strictly due to physical design. Recent work devotes to intrinsic parameter calibration of light field cameras in two designs [1], [2], which are quite different according to the image pattern of micro-lenses.

The main difference of light field cameras is the relative position of main lens's imaging plane and the MLA plane [20]. It determines rays' distribution from the same point, which affects the way to extract sub-apertures from raw image, i.e., the micro-lens images [21], [22]. However, the measurements of the same point in multiple directions are obtained in different types of light field cameras, equivalent to the data of GCM. Therefore, the light field camera model can use classic multi-view geometry theory for reference.

Recently, some state-of-the-art methods have proposed models on conventional light field camera, where multiple viewpoints or sub-apertures are convenient to be synthesized. Dansereau et al. [11] present a model to decode pixels into rays for a Lytro camera, where a 12-free-parameter transformation matrix is related to reference plane outside the camera (in nonlinear optimization, 10 intrinsic parameters and 5 distortion coefficients are finally estimated). However, the calibration method using traditional camera calibration algorithm is not effective, also there are redundant parameters in the decoding matrix. Bok et al. [13] formulate a geometric projection model consisting of a main lens and a MLA (their extended work has been published in IEEE TPAMI [23]). Intrinsic parameters are estimated by conducting raw images directly and an analytical solution is deduced. Moreover, Thomason et al. [15] try to deal with the misalignment of the MLA and estimated its position and orientation.

Apart from this, other researchers have explored models on the focused light field camera, where multiple projections of the same point are convenient to be recognized. Johannsen et al. [12] propose to calibrate intrinsic parameters of the focused light field camera. By reconstructing 3D

TABLE 1
Notation of Symbols in the Paper

Term	Definition
$L(i, j, u, v)$	Indexed pixel of raw image inside the camera
$L(I, J, U, V)$	Virtual (conjugate) light field outside the camera
$L(s, t, x, y)$	Decoded physical light field
$(k_i, k_j, k_u, k_v, u_0, v_0)$	Intrinsic parameters
$\mathbf{X}_w$	3D point in the world coordinates
$\mathbf{X}_d$	3D point reconstructed by $L(i, j, u, v)$
$\mathbf{X}_c$	3D point reconstructed by $L(s, t, x, y)$
$\mathbf{R}_{3\times3} = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]$	Rotation matrix of extrinsic parameter
$\mathbf{t}_{3\times1} = (t_x, t_y, t_z)^\top$	Translation vector of extrinsic parameter
$\mathbf{M}_{2n\times4}$	Measurement matrix of n rays
$\mathbf{P}_{4\times4}$	Homogenous projection matrix
$\mathbf{A}_{3\times3}$	Non-homogenous projection matrix partitioned from $\mathbf{P}$
$\mathbf{H}_{4\times3}$	Homography matrix decided by intrinsic and extrinsic parameters only
$\mathbf{d} = (k_1, k_2, k_3, k_4)^\top$	Distortion vector

points from the parallax in adjacent micro-lens images, the parameters (including depth distortion) are estimated. However, the geometry center of micro image is on its micro-lens's optical axis in the camera model. This assumption causes inaccuracy on reconstructed points and estimated results are finally compensated by the coefficients of depth distortion. Hahne et al. [24] further discuss the influence of above-mentioned assumption, i.e., the deviation of micro-lens and its image. Heinze et al. [25] apply a similar model with Johannsen et al. [12] and deduce a linear initialization for intrinsic parameters.

In a word, previous light field camera models are either redundant or complex, which leads to a non-unique solution of intrinsic parameter estimation or inaccuracy of decoding light field. An unreliable camera model is also a bottleneck that might impede light field applications for computer vision and computational photography, especially on light field registration, stitching and enhancement. To support further applications, a general light field camera model capable of representing rays and scene geometry more concisely is in urgent need.

3 MULTI-PROJECTION-CENTER MODEL

In this section, we first propose the MPC model based on the TPP representation of light field. Then we deduce the transformation matrix to relate 3D scene geometry and 4D rays. Finally, we utilize the MPC model to describe the image formation of light field cameras and define generic intrinsic parameters, including conventional and focused light field cameras. Table 1 gives the notation of symbols used in the following sections.

3.1 The Coordinates of MPC Model

As shown in Fig. 1, there are three coordinates in the MPC model, i.e., 3D world coordinates $O_w X_w Y_w Z_w$, 3D camera coordinates $OXYZ$, 4D TPP coordinates $o_{st}st - o_{xy}xy$ ($o_{st}st$ for the view plane and $o_{xy}xy$ for the image plane). In general, the transformation between world and camera coordinates is related by extrinsic parameters $[\mathbf{R}|\mathbf{t}]$. The spacing between two parallel planes of traditional TPP representation is normalized as 1 to describe a set of rays [3], [4].
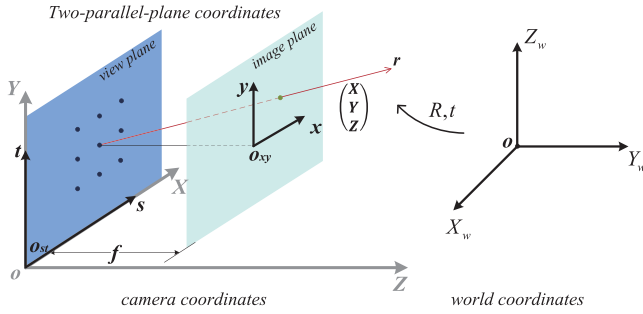
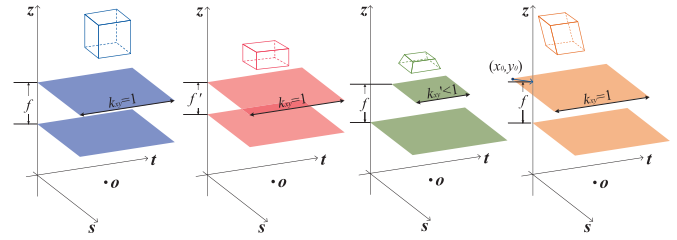Fig. 1. An illustration of three coordinates in the MPC model.

Fig. 2. Examples of transformations between 3D structure and 4D light field of a Lambertian cube. The leftmost is an original cube and others are the projected ones with the changing of parameter f, scaling in the image plane k_{xy} ($=k_x=k_y$) and translation $(0,0,x_0,y_0)^\top$ respectively.

Although it is complete and concise, to derive the transformation between 3D structure and 4D rays in light field cameras, we prefer a model consisting of two parallel planes with the spacing f.

Let $L_f(s,t,x,y)$ denote light field in the MPC model with the spacing f. Then the ray is parameterized by two planes, i.e., $Z=0$ and $Z=f$. Let $Z=0$ denote the view plane $o_{st}st$ and $Z=f$ denote the image plane $o_{xy}xy$. In the MPC model, $\mathbf{r}=(s,t,x,y)^\top$ defines a ray passing $(s,t,0)^\top$ and $(x,y,f)^\top$, where $(s,t,0)^\top$ is the projection center and $(x,y,f)^\top$ is the corresponding projection.

Given a projection center $(s,t,0)^\top$ (i.e., the (s,t)th view or sub-aperture) and the 3D point $\mathbf{X}=(X,Y,Z,1)^\top$, we can get the image projection $\mathbf{x}=(x,y,1)^\top$ in the local coordinate of the (s,t)th view,

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & -fs \\ 0 & f & 0 & -ft \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (1)$$

Since there are multiple projection centers $(s,t,0)^\top$, $s,t=1,2,\ldots,N$, the 3D point $\mathbf{X}$ can be observed for $N\times N$ times. Obviously, when the spacing f changes to 1 and there is only one projection center $(0,0,0)^\top$ on the view plane, the image formation degenerates into traditional central-projective camera model [19].

3.2 Transformation between Geometry and Rays

It is known that different directional rays from one point enable 3D reconstruction. Let the ray $\mathbf{r}$ intersect at the point $\mathbf{X}$ in the 3D space, we can get the relationship between the ray and 3D point by the triangulation,

$$\underbrace{\begin{bmatrix} f & 0 & -x & -fs \\ 0 & f & -y & -ft \end{bmatrix}}_{\mathbf{M}} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{0}, \quad (2)$$

where $\mathbf{M}$ is a $2n\times 4$ matrix consisting of n rays and the MPC parameter f.

If two rays $\mathbf{r}_i=(s_i,t_i,x_i,y_i)^\top$ and $\mathbf{r}_j=(s_j,t_j,x_j,y_j)^\top$ are from one 3D point $\mathbf{X}$, they can be represented by the following two equivalent forms,

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{x_i-x_j}\begin{bmatrix} s_jx_i-s_ix_j \\ t_i(x_i-x_j)-y_i(s_i-s_j) \\ f(s_j-s_i) \end{bmatrix} \quad (3)$$

and

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{y_i-y_j}\begin{bmatrix} s_i(y_i-y_j)-x_i(t_i-t_j) \\ t_jy_i-t_iy_j \\ f(t_j-t_i) \end{bmatrix}. \quad (4)$$

3.3 3D Projective Transformation

In fact, a linear transformation on the coordinates of $\mathbf{r}$ causes 3D projective distortion on the reconstructed point $\mathbf{X}$ [19], deduced from Eqs. (3) and (4). As shown in Fig. 2, we show three examples of linear transformations, including the changing of f, scaling in the image plane k_{xy} ($k_{xy}=k_x=k_y$) (in general there are 4 scaling factors k_s,k_t,k_x,k_y, two in the view plane and two in the image plane respectively), and translation in the image plane of specific view $(0,0,x_0,y_0)^\top$ (generally $(s_0,t_0,x_0,y_0)^\top$ in both planes). The details are derived as follows.

(1) If we change f into f', the imaging point $(x,y,f)^\top$ passed by $\mathbf{r}$ becomes $(x,y,f')^\top$ and the intersection of rays becomes $\mathbf{X}'$. Substituting it into Eqs. (3) and (4), we have

$$\mathbf{X}'=\mathbf{P}_1(f')\mathbf{X}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & f'/f & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (5)$$

where $\mathbf{X}$ and $\mathbf{X}'$ are in the homogeneous coordinates.

(2) Let $\mathbf{r}$ become $\mathbf{r}'=(s+s_0,t+t_0,x+x_0,y+y_0)^\top$, thus there is a transformation on the rays caused by the offset $\mathbf{m}=(s_0,t_0,x_0,y_0)^\top$. Substituting it into Eqs. (3) and (4), we can get the transformation matrix between $\mathbf{X}$ and $\mathbf{X}'$,

$$\mathbf{X}'=\mathbf{P}_2(\mathbf{m})\mathbf{X}=\begin{bmatrix} 1 & 0 & x_0/f & s_0 \\ 0 & 1 & y_0/f & t_0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (6)$$

(3) Let $\mathbf{r}$ become $\mathbf{r}'=(k_ss,k_tt,k_xx,k_yy)^\top$, thus there is a transformation caused by the scaling vector $\mathbf{k}=(k_s,k_t,k_x,k_y)^\top$. Then the transformation matrix between $\mathbf{X}$ and $\mathbf{X}'$ is,

$$\mathbf{X}'=\mathbf{P}_3(\mathbf{k})\mathbf{X}=\begin{bmatrix} k_s & 0 & 0 & 0 \\ 0 & k_t & 0 & 0 \\ 0 & 0 & k_s/k_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (7)$$
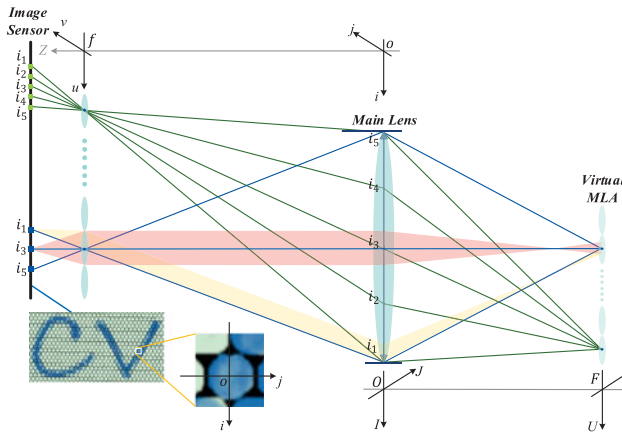
Fig. 3. Optical path of a conventional light field camera [1]. There are two MPC coordinates inside the camera and in the outer world with linear transformation, i.e., $L_f(i,j,u,v)$ and $L_F(I,J,U,V)$ respectively.

and

$$\mathbf{X}' = \mathbf{P}_4(\mathbf{k})\mathbf{X} = \begin{bmatrix} k_s & 0 & 0 & 0 \\ 0 & k_t & 0 & 0 \\ 0 & 0 & k_t/k_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (8)$$

In particular, Eqs. (7) and (8) hold when $k_s/k_t = k_x/k_y$.

As shown in the left-most of Fig. 2, there is a scene with a Lambertian cube recorded by a MPC model. The observation of the cube in multiple directions is 4D light field. If the coordinates are linearly transformed and the light intensity keeps constant, the intersections of rays will be transformed by a 3D projection matrix. Therefore, the cube will be projected by transformation parameters (the right three of Fig. 2).

3.4 The MPC Model in Light Field Cameras

Light field cameras are improved from traditional cameras. They record real world scene in different but similar ways. In traditional cameras, the central projection process of a 2D image is a dimension reduction of 3D space [19]. In light field camera, 3D structure projected by the main lens is arranged by the design of light path on the image sensor. The processes of multiple center projections are analyzed as follows.

On the one hand, as for a conventional light field camera, the sampling pattern of light field is shown in Fig. 3. The pixel $(u,v)^\top$ of $N \times N$ sub-aperture images is extracted from the micro-lens of $(u,v)^\top$. The sub-aperture image of the (i,j) view is extracted from the pixels $(i,j)^\top$ in the local micro-lens image coordinates, as shown in Fig. 3. Obviously, there are two light fields, i.e., $L_f(i,j,u,v)$ inside the camera and $L_F(I,J,U,V)$ in the outer world. Considering the projection of main lens, there is a 3D projective distortion between the 3D points reconstructed from L_f and L_F.

On the other hand, as for the focused light field cameras, two sampling patterns of light field in two different optical paths are shown in Fig. 4. The micro-lenses project the distorted 3D scene inside the camera on the image sensor, where the image range is controlled by the aperture of main lens and the distance of components. The light field inside the camera can be decoded by the pixels of image sensor
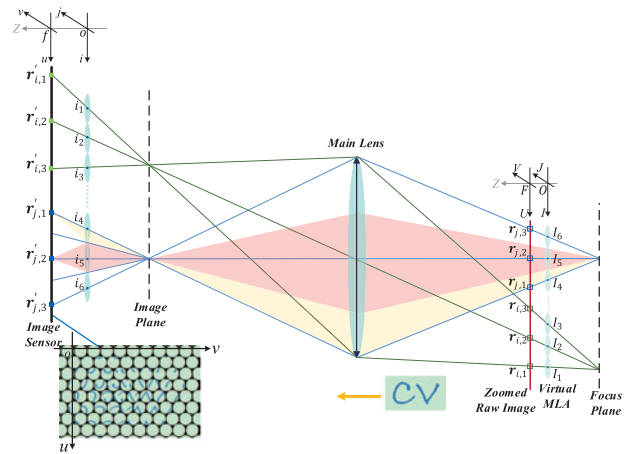
(a) Telescopic case
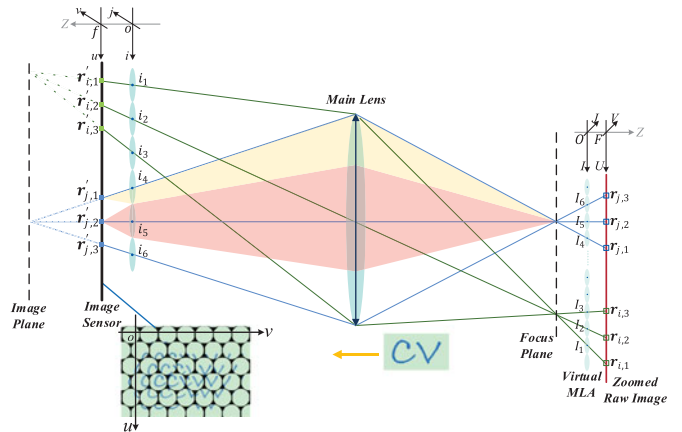
(b) Binocular case

Fig. 4. Optical paths of focused light field cameras with different designs [2]. There is a conjugate MPC coordinate $L_F(I,J,U,V)$ in the outer world with the inner one $L_f(i,j,u,v)$.

$(u,v,f)^\top$ and their corresponding optical centers of microlens $(i,j,0)^\top$, i.e., $L_f(i,j,u,v)$. In addition, $(i,j)^\top$ is determined by the layout of MLA, as shown in Fig. 5b. By the transformation on the coordinate of $L_f(i,j,u,v)$ we have discussed in Section 3.3, the outside light field $L_F(I,J,U,V)$ is obtained, which is the conjugate MPC coordinate outside the camera. The real world scene can be reconstructed by the light field $L_F(I,J,U,V)$ without projective distortion.

Let $L(i,j,u,v)$ denote indexed pixels of light field cameras with $f = 1$. Moreover, $L(i,j,u,v)$ is a set of indexed pixels and not a physical light field. In conventional light field camera, $L(i,j,u,v)$ are the sub-aperture images indexed by the (i,j) view. In the focused light field cameras, $L(i,j,u,v)$ are micro-lens images indexed by their relative positions on the raw image. Obviously, by a linear transformation on the $L(i,j,u,v)$, we can conduct $L_F(I,J,U,V)$ and eliminate 3D projective distortion caused by the main lens. However, to parameterize 4D light field without redundancy, the spacing of two parallel planes should be 1. Let $L(s,t,x,y)$ denote the normalized light field. According to Eqs. (5) to (8), the normalization is a linear operation on the coordinates, and transformation matrices $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ are all identity matrices. It means that indexed pixels $L(i,j,u,v)$ can be transformed to physical rays $L(s,t,x,y)$ in real world scene by linear transformations as we discussed before. The indexed

(a) Conventional light field camera
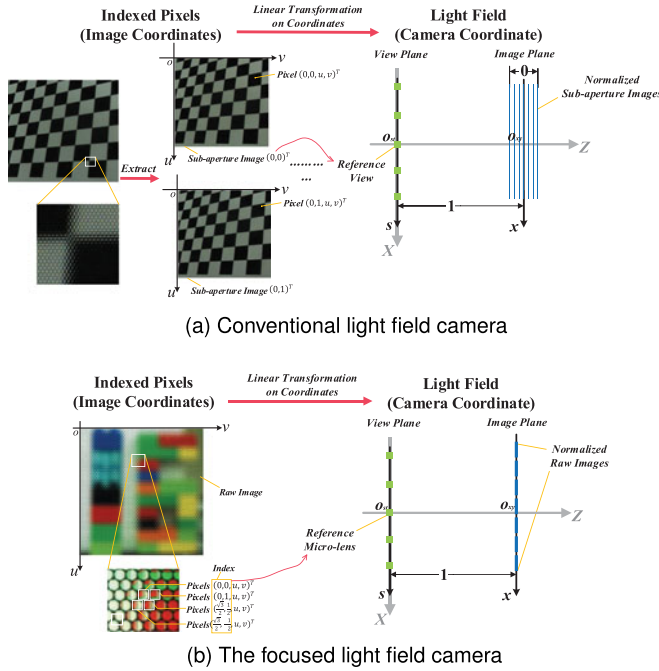

(b) The focused light field camera

Fig. 5. The indexed pixels $L(i, j, u, v)$ and decoded physical light field $L(s, t, x, y)$ of light field cameras in two designs.

pixels $L(i, j, u, v)$ and decoded physical light field $L(s, t, x, y)$ of light field cameras in two different designs are shown in Fig. 5, where pixels and physical rays are related by intrinsic parameters.

In summary, we can transform an indexed pixel of raw image $L(i, j, u, v)$ into a normalized physical light field $L(s, t, x, y)$ by a decoding matrix $\mathbf{D}$ that is consisting of intrinsic parameters $(k_i, k_j, k_u, k_v, u_0, v_0)$.

$$
\begin{bmatrix} s \\ t \\ x \\ y \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} k_i & 0 & 0 & 0 & 0 \\ 0 & k_j & 0 & 0 & 0 \\ 0 & 0 & k_u & 0 & u_0 \\ 0 & 0 & 0 & k_v & v_0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{=: \mathbf{D}} \begin{bmatrix} i \\ j \\ u \\ v \\ 1 \end{bmatrix}. \tag{9}
$$

Let $\mathbf{X}_d = [X_d, Y_d, Z_d, 1]^\top$ and $\mathbf{X}_c = [X_c, Y_c, Z_c, 1]^\top$ denote two 3D points reconstructed by $L(i, j, u, v)$ and $L(s, t, x, y)$ respectively. According to Eq. (9), the relationship between $\mathbf{X}_d$ and $\mathbf{X}_c$ is

$$
\underbrace{\begin{bmatrix} 1/k_i & 0 & -u_0/k_i & 0 \\ 0 & 1/k_j & -v_0/k_j & 0 \\ 0 & 0 & k_u/k_i & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{=: \mathbf{P}} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} X_d \\ Y_d \\ Z_d \\ 1 \end{bmatrix}, \tag{10}
$$

where $\mathbf{P} = \mathbf{P}_1(1)\mathbf{P}_3(\mathbf{k})\mathbf{P}_2(\mathbf{m})$ is determined by intrinsic parameters in the decoding matrix $\mathbf{D}$. Here, $\mathbf{m} = (0, 0, -u_0, -v_0)^\top$ and $\mathbf{k} = (1/k_i, 1/k_j, 1/k_u, 1/k_v)^\top$, which are totally decided by the mapping from indexed pixels to real world light rays.

In addition, the light field inside a conventional light field camera (in Fig. 3) can also be parameterized by the MPC model that is consisting of image sensor and the MLA. However, considering the convenience of extracting sub-

aperture images and the difficulty on detecting points on raw image in a conventional light field camera, we prefer to discuss the data as a set of sub-aperture images. Conversely, for the focused one, we model the parameterization plane by the raw image plane and discuss the raw image directly.

4 LIGHT FIELD CAMERA CALIBRATION

We verify our light field camera model by intrinsic parameter calibration. We will provide the details of how to solve intrinsic parameters, including a linear closed-form solution and a nonlinear optimization to minimize the re-projection error. In our method, the prior scene points are supported by a planar calibration board in different poses.

4.1 Linear Initialization

After necessary preprocessing, the micro-lens images are recognized [11], [21], [26], i.e., $L(i, j, u, v)$. We assume that the prior 3D point $\mathbf{X}_w$ in the world coordinates is related to the 3D point $\mathbf{X}_c$ in the MPC coordinates by a rigid motion, $\mathbf{X}_c = \mathbf{R}\mathbf{X}_w + \mathbf{t}$, with the rotation $\mathbf{R} \in SO(3)$ and translation $\mathbf{t} = (t_x, t_y, t_z)^\top \in \mathbb{R}^3$. Let $\mathbf{r}_i$ denote ith column vector of $\mathbf{R}$. The relationship among $\mathbf{R}$, $\mathbf{t}$, $\mathbf{X}_w$ and intrinsic parameters $\mathcal{P} = (k_i, k_j, k_u, k_v, u_0, v_0)$ is obtained by Eqs. (2) and (10).

$$
\mathbf{MP} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{X}_w = \mathbf{0}. \tag{11}
$$

where $\mathbf{M}$ is a $2n \times 4$ measurement matrix of n rays and $n \geq 2$. These rays are derived from the indexed pixels $L(i, j, u, v)$ as mentioned in Eq. (1).

Suppose that the calibration board is on the plane of $Z = 0$ in the world coordinates, thus $Z_w = 0$. To solve the unknown parameters, we simplify Eq. (11) as,

$$
\mathbf{M} \otimes [X_w \quad Y_w \quad 1]\vec{\mathbf{H}} = \mathbf{0}, \tag{12}
$$

where $\vec{\mathbf{H}}$ is a 12×1 matrix stretched on row from $\mathbf{H}$. $\otimes$ is a direct product operator. $\mathbf{H}$ is a 4×3 matrix only consisting of intrinsic and extrinsic parameters, defined as

$$
\mathbf{H} = \mathbf{P} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \\ 0 & 0 & 1 \end{bmatrix}. \tag{13}
$$

In addition, $\mathbf{M}$ is a matrix containing at least 2 rays from light field $L(i, j, u, v)$, according to Eq. (2). By stacking measurements from at least 3 non-collinear points $\mathbf{X}_w$, the homography $\mathbf{H}$ can be estimated by Eq. (12).

In order to derive intrinsic parameters from $\mathbf{H}$, we can partition $\mathbf{P}$ to extract a 3×3 upper triangle matrix $\mathbf{A}$. Let h_{ij} denote the element on the ith row and jth column of $\mathbf{H}$, we rewrite Eq. (13) as follows,

$$
\begin{bmatrix} \mathbf{G} & \begin{matrix} h_{13} \\ h_{23} \\ h_{33} \end{matrix} \\ \hline \mathbf{0}_{1 \times 2} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \hline \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \\ 0 & 0 & 1 \end{bmatrix}, \tag{14}
$$

where $\mathbf{G}$ is a 3×2 matrix, i.e., top-left 3×2 of $\mathbf{H}$.

Let $\mathbf{g}_j = (g_{1j}, g_{2j}, g_{3j})^\top, j = 1, 2$ denote the jth column vector of $\mathbf{G}$. Utilizing the orthogonality and identity of $\mathbf{R}$, we have

$$\mathbf{g}_1^\top \mathbf{A}^{-\top} \mathbf{A}^{-1} \mathbf{g}_2 = 0,$$
$$\mathbf{g}_1^\top \mathbf{A}^{-\top} \mathbf{A}^{-1} \mathbf{g}_1 = \mathbf{g}_2^\top \mathbf{A}^{-\top} \mathbf{A}^{-1} \mathbf{g}_2, \tag{15}$$

where $\mathbf{A}^{-1} = \begin{bmatrix} k_i & 0 & u_0 k_i/k_u \\ 0 & k_j & v_0 k_j/k_v \\ 0 & 0 & k_i/k_u \end{bmatrix}$.

Let a symmetric matrix $\mathbf{B}$ denote $\mathbf{A}^{-\top}\mathbf{A}^{-1}$. The analytical form of $\mathbf{B}$ is

$$\mathbf{B} = \begin{bmatrix} k_i^2 & 0 & k_i^2 u_0/k_u \\ 0 & k_j^2 & k_j^2 v_0/k_v \\ k_i^2 u_0/k_u & k_j^2 v_0/k_v & k_i^2/k_u^2 (1 + u_0^2 + v_0^2) \end{bmatrix}. \tag{16}$$

Note that there are only 5 distinct non-zero elements in $\mathbf{B}$, denoted by $\mathbf{b} := (b_{11}, b_{13}, b_{22}, b_{23}, b_{33})^\top$. To solve $\mathbf{B}$, we rewrite Eq. (15) as follows,

$$\begin{bmatrix} g_{11}g_{12} & g_{11}^2 - g_{12}^2 \\ g_{11}g_{32} + g_{12}g_{31} & 2(g_{11}g_{31} - g_{12}g_{32}) \\ g_{21}g_{22} & g_{21}^2 - g_{22}^2 \\ g_{21}g_{32} + g_{31}g_{22} & 2(g_{21}g_{31} - g_{22}g_{32}) \\ g_{31}g_{32} & g_{31}^2 - g_{32}^2 \end{bmatrix}^\top \mathbf{b} = \mathbf{0}. \tag{17}$$

By stacking at least two such equations (from two poses) as Eq. (17), we can obtain a unique general non-zeros solution for $\mathbf{b}$, which is defined up to an unknown scale factor.

Once $\mathbf{B} = \mathbf{A}^{-\top}\mathbf{A}^{-1}$ is determined, it is an easy matter to solve $\mathbf{A}^{-1}$ using Cholesky factorization [27]. Let $\hat{\mathbf{A}}$ denote the estimation of $\mathbf{A}$, i.e., $\lambda\hat{\mathbf{A}} = \mathbf{A}$. Let $\hat{a}_{ij}$ denote the element on the ith row and jth column of $\hat{\mathbf{A}}$, intrinsic parameters except k_i and k_j are estimated by the ratio of elements

$$k_u = \hat{a}_{11}/\hat{a}_{33}, \quad k_v = \hat{a}_{22}/\hat{a}_{33},$$
$$u_0 = \hat{a}_{13}/\hat{a}_{33}, \quad v_0 = \hat{a}_{23}/\hat{a}_{33}. \tag{18}$$

Apart from intrinsic parameters, extrinsic parameters in different poses can be extracted as follows,

$$\lambda = \alpha(\|\hat{\mathbf{A}}^{-1}\mathbf{g}_1\| + \|\hat{\mathbf{A}}^{-1}\mathbf{g}_2\|)/2,$$
$$\mathbf{r}_1 = \hat{\mathbf{A}}^{-1}\mathbf{g}_1/\lambda,$$
$$\mathbf{r}_2 = \hat{\mathbf{A}}^{-1}\mathbf{g}_2/\lambda, \tag{19}$$
$$\mathbf{t} = \hat{\mathbf{A}}^{-1}[h_{13}\ h_{23}\ h_{33}]^\top/\lambda,$$
$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2,$$

where $\|\cdot\|$ denotes L_2 norm. α values 1 or -1 and it is decided by image formation. In conventional light field camera and the focused one with shorter light path (as shown in Figs. 3 and 4b), α makes $t_z > 0$. Otherwise, in the focused light field camera with longer light path (see Fig. 4a), α makes $t_z < 0$.

To obtain other two intrinsic parameters k_i and k_j, we substitute the results in Eq. (19) for Eq. (11) and obtain $\mathbf{X}_c = \mathbf{R}\mathbf{X}_w + \mathbf{t}$ using the estimated extrinsic parameters. Then, Eq. (2) is rewritten as,

$$\begin{bmatrix} i & 0 \\ 0 & j \end{bmatrix}\begin{bmatrix} k_i \\ k_j \end{bmatrix} = \begin{bmatrix} X_c - xZ_c \\ Y_c - yZ_c \end{bmatrix}. \tag{20}$$

Stacking the measurements in different poses, we can obtain a unique non-zeros solution for k_i and k_j.

4.2 Nonlinear Optimization

The most common distortion of traditional camera is radial distortion. The optical property of main lens and physical machining error of the MLA might lead to the distortion of rays in light field camera. Theoretically, due to two level imaging design with main lens and micro-lens array, there should exist radial distortion on the image plane $(x, y)^\top$ and sampling distortion on the view plane $(s, t)^\top$ simultaneously. In the paper, we only consider the distortion on the image plane and omit sampling distortion on the view plane (i.e., angular sampling grid is ideal without distortion).

$$\begin{cases} x^u = (1 + k_1 r_{xy}^2 + k_2 r_{xy}^4)x + k_3 s \\ y^u = (1 + k_1 r_{xy}^2 + k_2 r_{xy}^4)y + k_4 t, \end{cases} \tag{21}$$

where $r_{xy}^2 = x^2 + y^2$ and $\mathbf{r} = (s, t, x, y)^\top$ is the ray transformed from the measurement $L(i, j, u, v)$ by intrinsic parameter $\mathcal{P}$ according to Eq. (9). $\mathbf{d} = (k_1, k_2, k_3, k_4)^\top$ denotes distortion vector and $\mathbf{x}^u = (x^u, y^u)^\top$ is undistorted projection from the distorted one $\mathbf{x} = (x, y)^\top$ in the local image coordinates under the $(s, t)^\top$ view. In the distortion vector $\mathbf{d}$, k_1 and k_2 regulate radial distortion on the image plane. k_3 and k_4 represent the distortion of image plane affected by the sampling view $(s, t)^\top$, which is caused by non-paraxial rays of the main lens.

We minimize the following cost function with the initialization solved in Section 4.1 to refine the parameters, including intrinsic parameter $\mathcal{P}$, distortion vector $\mathbf{d}$, and extrinsic parameters $\mathbf{R}_p$ and $\mathbf{t}_p$, $p = 1, \ldots, P$, P is the number of poses.

$$\sum_{p=1}^{\#pose} \sum_{n=1}^{\#point} \sum_{i=1}^{\#view} \left\| \mathbf{x}_i^u(\mathcal{P}, \mathbf{d}) - \hat{\mathbf{x}}_i(\mathbf{R}_p, \mathbf{t}_p, \mathbf{X}_{w,n}) \right\|, \tag{22}$$

where $\mathbf{x}^u$ is the image point from $L(i, j, u, v)$ according to Eq. (9) and followed by distortion rectification according to Eq. (21). $\hat{\mathbf{x}}$ is the projection of 3D point $\mathbf{X}_{w,n}$ in the world coordinates according to Eq. (1).

In Eq. (22), $\mathbf{R}$ is parameterized by Rodrigues formula [28]. In addition, the Jacobian matrix of cost function is simple and sparse. This nonlinear minimization problem can be solved with the Levenberg-Marquardt algorithm based on trust region method [29]. We adopt MATLAB's lsqnonlin function to complete the optimization.

4.3 Computational Complexity

The calibration algorithm of light field camera is summarized in Algorithm 1. Let (S, T) denote sampling number on the view plane, (M, N) be the number of prior points on the calibration board, and P be the number of poses, respectively. For the measurement of each pose, there are $(2 \times S \times T) \times 12$ linear equations to solve $\mathbf{H}$. Then $(2 \times P) \times 5$ linear equations and $(2 \times P \times M \times N) \times 2$ equations are solved to obtain intrinsic parameters. The main complexity is spent on the solution of $\mathbf{H}$ from different poses, i.e., $O(P)$.

By contrast, the algorithm in Dansereau et al. [11] calculates a homography for every sub-aperture image in different view, i.e., $O(P \times S \times T)$. It suffers from a higher

TABLE 2
Intrinsic Parameter Configuration of the
Simulated Light Field Camera

k_i	k_j	k_u	k_v	u_0	v_0
2.4000e-04	2.5000e-04	2.0000e-03	1.9000e-03	−0.3200	−0.3300

complexity and a lower accuracy on parameter initialization. The algorithm in Bok et al. [13] solves a linear equation on every pose and its computational complexity is $O(P)$. However, there are intrinsic and extrinsic parameters in the equations, which causes inaccuracy on the solution.

Algorithm 1. Light Field Camera Calibration Algorithm

Input: 3D prior points $\mathbf{X}_w$ and corresponding rays $L(i,j,u,v)$.
Output: Intrinsic parameters $\mathcal{P} = (k_i, k_j, k_u, k_v, u_0, v_0)$;
    Extrinsic parameters $\mathbf{R}_p, \mathbf{t}_p (1 \le p \le P)$;
    Distortion vector $\mathbf{d} = (k_1, k_2, k_3, k_4)^\top$.
1: **for** $p = 1$ to P **do**
2:   **for** each 3D point $\mathbf{X}_w$ **do**
3:     Generate the measurement matrix $\mathbf{M}$ from indexed
      pixel $L(i,j,u,v)$ ▷ Eq. (2)
4:   **end for**
5:   Calculate the homography matrix $\mathbf{H}_p$ according to $\mathbf{X}_w$
    and $\mathbf{M}$ ▷ Eq. (12)
6: **end for**
7: Calculate the matrix $\hat{\mathbf{B}}$ ▷ Eq. (17)
8: Calculate projection matrix $\hat{\mathbf{A}}$ from $\hat{\mathbf{B}}$ using Cholesky
  factorization
9: Obtain four intrinsic parameters (k_u, k_v, u_0, v_0) ▷ Eq. (18)
10: **for** $p = 1$ to P **do**
11:   Get extrinsic parameters $\mathbf{R}_p$ and $\mathbf{t}_p$ ▷ Eq. (19)
12: **end for**
13: Obtain other two intrinsic parameters (k_i, k_j) ▷ Eq. (20)
14: Initialize distortion coefficient $\mathbf{d} = (0,0,0,0)^\top$
15: Create the cost function according to intrinsic parameters,
  extrinsic parameters and distortion coefficient ▷ Eq. (22)
16: Obtain optimized results using nonlinear LM algorithm

5 EXPERIMENTAL RESULTS

In this section, we verify our light field camera model by the calibration of intrinsic parameters. We present various experimental results both on simulated and real datasets. The performance is analyzed by comparing with the ground truth or baseline algorithms [11] and [13].

5.1 Simulated Data

In this section we verify our calibration method on simulated data. The simulated light field camera has the following property referred to Eq. (9), as shown in Table 2. These parameters are close to the setting of Lytro camera so that we obtain plausible input close to real-world scenarios. The checkerboard is a pattern with 12×12 points with $3.51\,\text{mm} \times 3.51\,\text{mm}$ cells.

5.1.1 Performance w.r.t. the Number of Poses and Views

First, we test the performance with respect to the number of poses and the number of views. We vary the number of
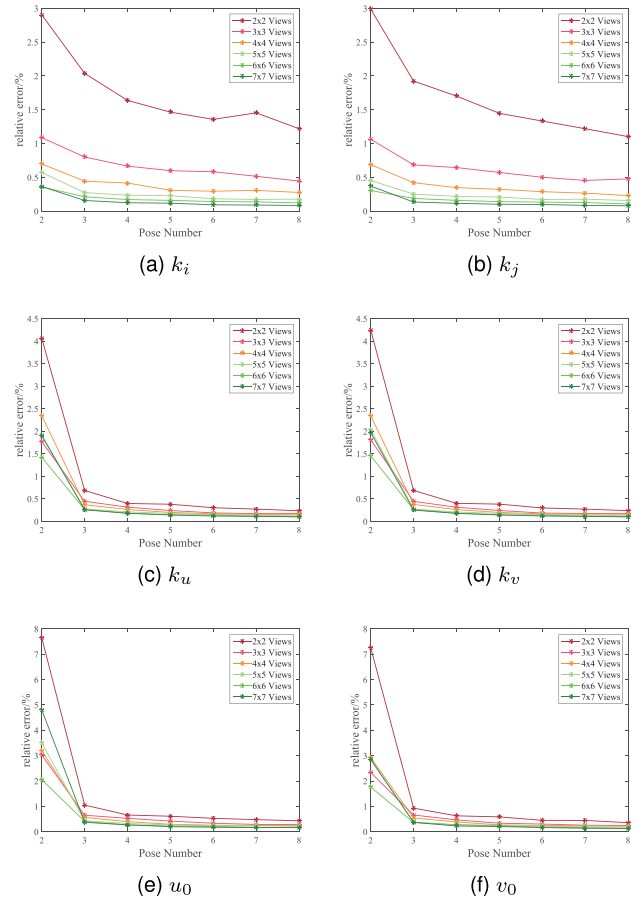

Fig. 6. Relative errors of intrinsic parameters on the simulated data with different number of poses and views.

poses from 2 to 8 and the number of views from 2×2 to 7×7. For each combination of pose and view, 200 trails of independent calibration board poses are generated. The rotation angles are randomly generated from $-30°$ to $30°$, and the measurements are all added with Gaussian noise with zero mean and standard deviation 0.5 pixels.

The calibration results with increasing measurements are shown in Fig. 6. We find that the relative errors decrease with the increase in the number of poses. When the number of pose is greater than 2, all the relative errors are within an acceptable level, as summarized in Table 3. Meanwhile, the errors reduce as the number of views grows once the number of poses is fixed. In particular, when $\#pose \ge 3$ and $\#view \ge 4 \times 4$, all the relative errors are less than 0.5 percent. Furthermore, the standard deviations of relative errors of Fig. 6 are shown in Fig. 7, from which we can see that standard deviations decrease significantly when the number of pose is greater than 2. Particularly, when $\#pose \ge 5$ and $\#view \ge 3 \times 3$, standard deviations keep at a low level

TABLE 3
Min and Max Relative Errors of Intrinsic Parameters (Unit: %) on the Simulated Data When the Number of Poses Is Great Than 2

	k_i	k_j	k_u	k_v	u_0	v_0
Min	0.0842	0.0795	0.1019	0.1020	0.1633	0.1295
Max	2.0376	1.9238	0.6871	0.6881	1.0511	0.9298

(a) k_i
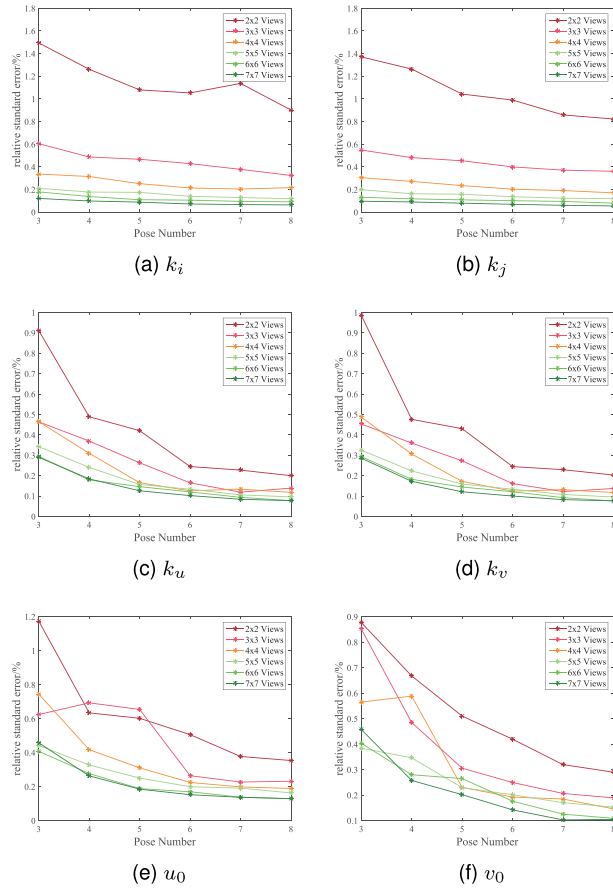
(b) k_j

(c) k_u

(d) k_v

(e) u_0

(f) v_0

Fig. 7. Standard deviations of relative errors of intrinsic parameters on the simulated data with different number of poses and views.

TABLE 4
RMS Ray Re-Projection Errors of Initial Parameter Estimation and Optimization with Distortion Rectification (Unit: mm)

		A	A(6)	B	C	D	E
Initial	DPW [11]-1	3.2000	-	5.0600	8.6300	5.9200	13.8000
	DPW [11]-2	**0.5190**	**0.4229**	**0.5403**	0.8832	1.1021	5.9567
	Ours	15.3753	0.5400	0.5952	**0.5837**	**0.7473**	**2.6235**
Optimized	DPW [11]-1	0.0835	-	0.0628	**0.1060**	0.1050	**0.3630**
	DPW [11]-2	0.0822	0.0903	0.0598	0.1300	0.1149	0.3843
	Ours	**0.0810**	**0.0810**	**0.0572**	0.1123	**0.1046**	0.5390

The datasets are from [11]. The (N) indicates the number of light fields used for calibration among 10 light fields of dataset A. The errors of DPW [11]-1 are provided by the paper directly, and the errors of DPW [11]-2 are obtained by running the latest released code.

rotation angles of 3 poses are $(6°, 28°, -8°)$, $(12°, -10°, 15°)$ and $(-5°, 5°, -27°)$ respectively. Gaussian noise with zero mean and a standard deviation σ is added to the projected image points. We vary σ from 0.1 to 1.5 pixels with a 0.1 step. For each noise level, we performed 150 independent trials. The mean results compared with ground truth are shown in Fig. 8. It demonstrates that the errors increase almost linearly with the noise level. For $\sigma = 0.5$ pixels which is larger than normal noise in practical calibration, the errors of (k_i, k_j) and (k_u, k_v) are less than 0.13 percent. Although the relative error of (u_0, v_0) is 0.24 percent, the absolute error of $(-u_0/k_u, -v_0/k_v)$ is less than 0.23 pixel (In Eq. (9), $u = (x - u_0)/k_u$ and $v = (y - v_0)/k_v$, where $(-u_0/k_u, -v_0/k_v)^\top$ is the principal point of sub-aperture imaging), which further exhibits that the proposed algorithm is robust to higher noise level.

5.2 Physical Camera

We also verify the calibration method on real scene light fields captured by conventional and focused light filed cameras. For the conventional light field camera, we use Lytro and Illum to obtain measurements. For the focused one, we use a self-assembly camera according to optical design in Fig. 4a.

5.2.1 Conventional Light Field Camera

The sub-aperture images are obtained by the method of Dansereau et al. [11]. We compare the proposed method in ray re-projection error with state-of-the-arts, including DPW by Dansereau et al. [11] and BJW by Bok et al. [13].

First, we carry out calibration on the datasets collected with [11]. For every different pose, the middle 7×7 sub-apertures are utilized similar to DPW. Table 4 summarizes the root mean square (RMS) ray re-projection errors of our method and DPW[11]. In Table 4, the errors of DPW [11]-1 are taking from the paper directly. The errors of DPW [11]-2 are obtained by running their latest released code. On the item of initial, the proposed method provides a smaller ray re-projection error than DPW except on datasets A and B. The result on dataset A performs worse because of bad corner extraction from several poses (i.e., 7th, 8th, 9th and 10th light field). On the item of optimized, compared with DPW [11] which employs 12 intrinsic parameters, the proposed MPC model only employs a half of parameters but achieves similar performance on ray re-projection errors

stably. The results in Figs. 6 and 7 have verified the effectiveness of the proposed calibration algorithm.

5.1.2 Performance w.r.t. the Measurement Noise

Second, we employ the measurements of 3 poses and 7×7 views to verify the robustness of calibration algorithm. The
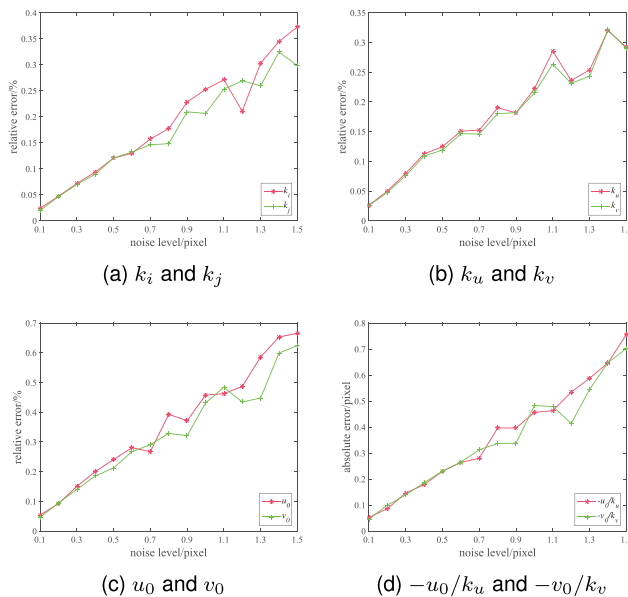

(a) k_i and k_j

(b) k_u and k_v

(c) u_0 and v_0

(d) $-u_0/k_u$ and $-v_0/k_v$

Fig. 8. Relative errors of intrinsic parameters on the simulated data with different noise levels from 0.1 to 1.5 pixels.

(a) A   (b) A(6)

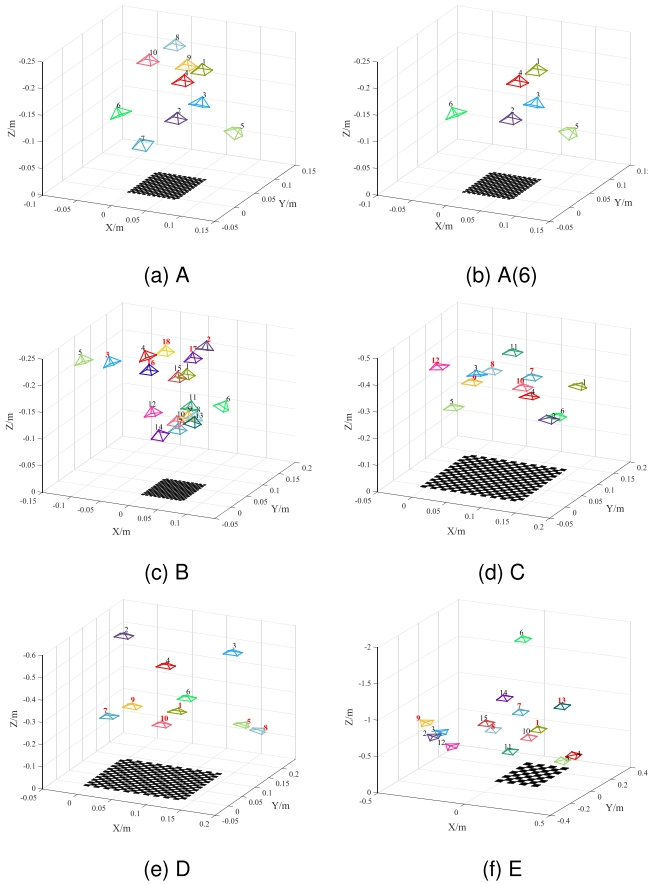(c) B   (d) C

(e) D   (f) E

Fig. 9. Pose estimation results of datasets captured by [11]. Light fields used for calibration in Table 6 are indicated with bold red indexes of corresponding camera poses in Figs. (c-f).

TABLE 6
RMS Errors of Optimization with Distortion Rectification Using Fewer Poses

	Ray re-projection error unit: mm		Re-projection error unit: pixel	
	DPW [11]	Ours	DPW [11]	Ours
B(5)	0.0643	**0.0622**	0.2380	**0.1458**
C(5)	0.1260	**0.1250**	0.2323	**0.1705**
D(6)	0.0941	**0.0622**	0.2024	**0.1458**
E(5)	0.2967	**0.2888**	0.3525	**0.2049**

The (N) indicates the number of light fields used for calibration.

TABLE 7
Intrinsic Parameter Estimation Results of Datasets Captured by [11]

	A	B	C	D	E
k_i	2.6998e-04	2.7937e-04	2.4569e-04	2.6833e-04	2.3004e-04
k_j	2.7608e-04	2.8874e-04	2.5359e-04	2.6930e-04	2.3073e-04
k_u	1.8572e-03	1.8357e-03	1.8122e-03	1.8342e-03	1.7585e-03
k_v	1.8692e-03	1.8323e-03	1.8133e-03	1.8352e-03	1.7634e-03
u_0	−0.3417	−0.3415	−0.3550	−0.3343	−0.3520
v_0	−0.3449	−0.3344	−0.3382	−0.3275	−0.3615
k_1	0.2288	0.1829	0.1639	0.1719	0.1612
k_2	−0.0928	0.0875	0.0174	0.0213	−0.0483
k_3	−4.5308	−3.6330	−3.3591	−3.5122	2.7747
k_4	−4.4428	−3.6064	−3.3394	−3.4662	2.8320

(the results on datasets A, B and D are better but the results on datasets C and E are worse). Light fields within each dataset are taken over a range of depths and orientations, as shown in Fig. 9. The ranges of datasets A, B are $0.25\,\mathrm{m}$ whilst the ranges of datasets C and D do not exceed $0.5\,\mathrm{m}$. Meanwhile, the ranges do not exceed $2\,\mathrm{m}$ in dataset E. Large ranges are reasonable in all datasets only deducing the accuracy in light of distortion model considering the shifted view. This is the main reason why the performance of the proposed method is worse than that of DPW on dataset E. From the dataset A, we select 6 light fields from which the corners are exactly extracted for the proposed method. The ray re-projection error decreases obviously in Table 4. Considering the fact that the errors exhibited in DPW are minimized in its own optimization (i.e., ray re-projection error), we additionally evaluate the performance in mean re-projection error of DPW and BJW. As exhibited in Table 5,

the errors of the proposed method are obviously smaller than those of DPW and BJW. In addition, the calibration with fewer number of poses on the datasets [11] is conducted. For dataset D, we randomly select 6 light fields, and for datasets B, C and E, 5 light fields are randomly selected for calibration. Table 6 summarizes RMS ray re-projection errors and RMS re-projection errors of the proposed method and DPW [11] respectively. In Table 6, the proposed method achieves smaller errors than DPW. Besides, the calibration results on datasets D and E are obviously improved by reducing the number of poses. We find that smaller range of poses contributes to a performance improvement on datasets D and E, which is shown in Fig. 9. Table 7 lists intrinsic parameter estimation results. The re-projection errors of 7×7 sub-aperture images of B are summarized in Table 8. The distribution of errors is almost homogeneous. All results have verified the effectiveness of the proposed method.

Unlike the core idea of DPW, BJW directly utilizes raw data instead of sub-apertures. However it has a stricter requirement on the acquisition of the calibration board. The data for calibration must be unfocused in order to make the measurements detectable, thus some datasets provided by DPW are incalculable for BJW, just as shown in Table 5 (i.e., datasets C and D). In order to directly compare with DPW and BJW, we collect other 4 datasets[1] using Lytro and Illum cameras. The dataset Illum-1 shoots 9×13 corners with $15.0\,\mathrm{mm} \times 15.0\,\mathrm{mm}$ cells, including 9 poses. The dataset Lytro-1 shoots 8×11 corners with $11.3\,\mathrm{mm} \times 11.3\,\mathrm{mm}$ cells, including 8 poses. The datasets Illum-2 and Lytro-2 shoot 8×11 corners with $8.9\,\mathrm{mm} \times 8.9\,\mathrm{mm}$ cells, including 10

TABLE 5
Mean Re-Projection Errors of Optimization with Distortion Rectification (Unit: pixel)

	A	A(6)	B	C	D	E
DPW [11]	0.2284	0.3338	0.1582	0.1948	0.1674	0.3360
BJW [23]	0.3736	-	0.2589	-	-	0.2742
Ours	**0.2200**	**0.2375**	**0.1568**	**0.1752**	**0.1475**	**0.2731**

The results of DPW [11] are obtained by running their latest released code. The results of BJW [23] are from their latest paper.

1. http://www.npu-cvpg.org/opensource

TABLE 8
RMS Re-Projection Error of Sub-Apertures
in Dataset B (Uint: pixel)

i	j						
	-3	-2	-1	0	1	2	3
-3	0.1930	0.1820	0.1781	0.1759	0.1759	0.1812	0.2372
-2	0.1836	0.1763	0.1700	0.1687	0.1718	0.1786	0.1813
-1	0.1815	0.1724	0.1669	0.1658	0.1692	0.1761	0.1826
0	0.1783	0.1731	0.1683	0.1662	0.1713	0.1798	0.1897
1	0.1772	0.1733	0.1706	0.1705	0.1748	0.1837	0.1851
2	0.1769	0.1761	0.1757	0.1768	0.1809	0.1836	0.1815
3	0.2039	0.1746	0.1728	0.1730	0.1798	0.1833	0.2755

TABLE 9
RMS Ray Re-Projection Errors of Initial Parameter
Estimation Optimizations without and with
Distortion Rectification (Unit: mm)

		Illum-1	Illum-2	Lytro-1	Lytro-2
Initial	DPW [11]	0.9355	0.6274	0.6201	0.5057
	BJW [13]	1.0765	0.8330	1.6676	1.0201
	Ours	**0.7104**	**0.4899**	**0.3538**	**0.2364**
Optimized without Rectification	DPW [11]	0.5909	0.4866	0.1711	**0.1287**
	BJW [13]	-	-	-	-
	Ours	**0.5654**	**0.4139**	**0.1703**	0.1316
Optimized with Rectification	DPW [11]	0.2461	0.2497	0.1459	0.1228
	BJW [13]	0.3966	0.3199	0.4411	0.2673
	Ours	**0.1404**	**0.0936**	**0.1400**	**0.1124**

TABLE 10
Intrinsic Parameter Estimation Results of Our Collected Datasets

	Illum-1	Illum-2	Lytro-1	Lytro-2
k_i	3.5721e-04	2.2464e-04	5.9386e-04	3.8915e-04
k_j	3.5455e-04	2.3299e-04	5.7870e-04	3.8247e-04
k_u	1.4309e-03	1.6670e-03	9.5083e-04	1.3195e-03
k_v	1.4303e-03	1.6657e-03	9.4794e-04	1.3261e-03
u_0	-0.4565	-0.5178	-0.1964	-0.2775
v_0	-0.2827	-0.3557	-0.1865	-0.2521
k_1	0.3001	0.3562	-0.4559	0.0254
k_2	0.2779	0.2595	6.8221	0.8469
k_3	-1.4109	-0.6185	-1.3060	-2.2441
k_4	-1.4204	-0.8879	-1.3234	-2.2684

poses. For Illum-1 and Illum-2 datasets, the middle 13×13 views are used (15×15 views in total). For Lytro-1 and Lytro-2, the middle 7×7 views are used (9×9 views in total). Table 9 summarizes the RMS ray re-projection errors compared with DPW and BJW at three calibration stages. As exhibited in Table 9, the proposed method obtains smaller ray re-projection errors on the item of initial solution which verified the effectiveness of linear initial solution for both intrinsic and extrinsic parameters. Besides, the proposed method provides similar or even smaller ray re-projection errors on the item of optimization without rectification compared with DPW. It is noticed that the result on dataset Lytro-2 is relatively larger than that of DPW. The main reason is that distortion coefficients k_3 and k_4 in our model are similar to the elements of the decoding matrix in [11]. Considering the fact that MPC model
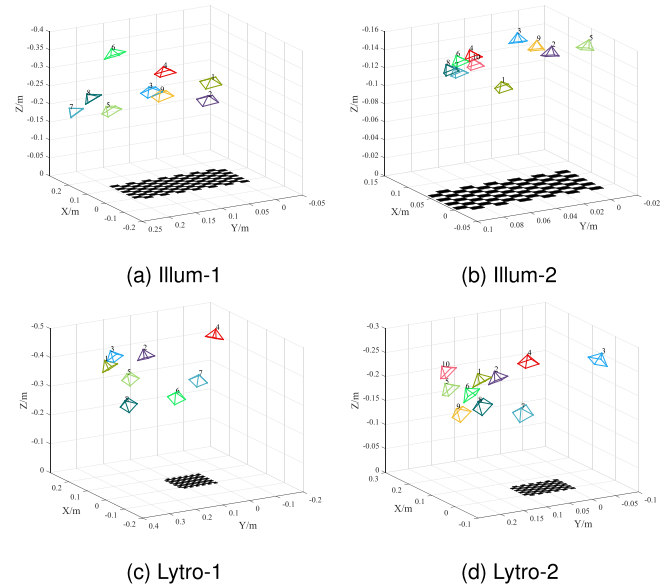

(a) Illum-1   (b) Illum-2

(c) Lytro-1   (d) Lytro-2

Fig. 10. Pose estimation results of our collected datasets.

TABLE 11
RMS Re-Projection Error of Sub-Apertures in
Dataset Illum-1 (Uint: pixel)

i	j						
	-5	-3	-1	0	1	3	5
-5	0.7880	0.2997	0.3008	0.3041	0.3058	0.3178	0.8070
-3	0.2992	0.3003	0.3033	0.3025	0.3015	0.2930	0.3077
-1	0.2988	0.3086	0.3176	0.3182	0.3141	0.2996	0.2827
0	0.2942	0.3139	0.3115	0.3058	0.3064	0.3024	0.2772
1	0.2934	0.3178	0.3118	0.2963	0.3077	0.3057	0.2784
3	0.3002	0.2966	0.3170	0.3093	0.3115	0.2856	0.2843
5	0.3283	0.2961	0.2851	0.2854	0.2841	0.2852	0.3102

employs less parameters (i.e., 6-parameter) than DPW (i.e., 12-parameter), the proposed method is competitive with acceptable calibration performance. Further, it is more important that we achieve smaller ray re-projection errors if distortion rectification is introduced in optimization. The ray re-projection errors are encouraging that the proposed method outperforms DPW and BJW. Consequently, the 6-parameter MPC model and 4-parameter distortion model are effective to represent light field cameras.

The reason why we compare ray re-projection errors here is to eliminate differences in camera models. The decoding matrix in [11] is similar to Eq. (9), except for the non-diagonal elements. The non-zero elements $H_{1,3}$ and $H_{2,4}$ indicate that pixels on the same sub-aperture image have specific relationships among different views. If we calculate the estimated rays by $\mathbf{X}_c$ for the measurement $(s, t, x, y)^\top$, the views may be different. It indicates that there are errors both on the view plane and image plane in [11]. As a result, it is not reasonable to compare re-projection error only.

Moreover, the results of intrinsic parameter estimation and pose estimation on our datasets are demonstrated in Table 10 and Fig. 10 respectively. After the calibration process, we measure the RMS re-projection errors of sub-aperture images by utilizing estimated parameters, as shown in Table 11. In order to further verify the accuracy of intrinsic
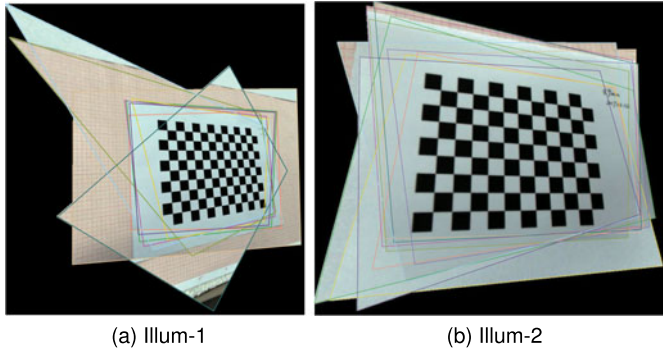
(a) Illum-1   (b) Illum-2

Fig. 11. The stitching results of Illum-1 and Illum-2 datasets (the first pose is regarded as the reference view).

(a) Original central view image   (b) Proposed (0.3159)
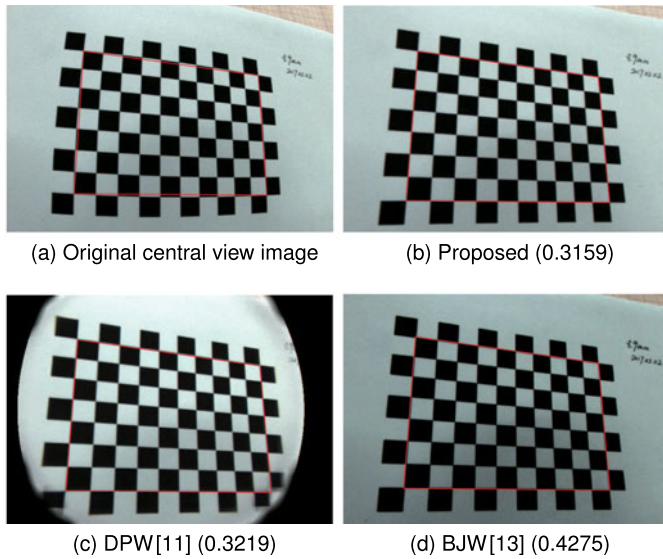
(c) DPW[11] (0.3219)   (d) BJW[13] (0.4275)

Fig. 12. The central view sub-aperture and distortion rectification results of first pose light field in Illum-2 dataset. The re-projection error (unit: pixel) of central view sub-aperture image is represented in parentheses.

(a) Ruler

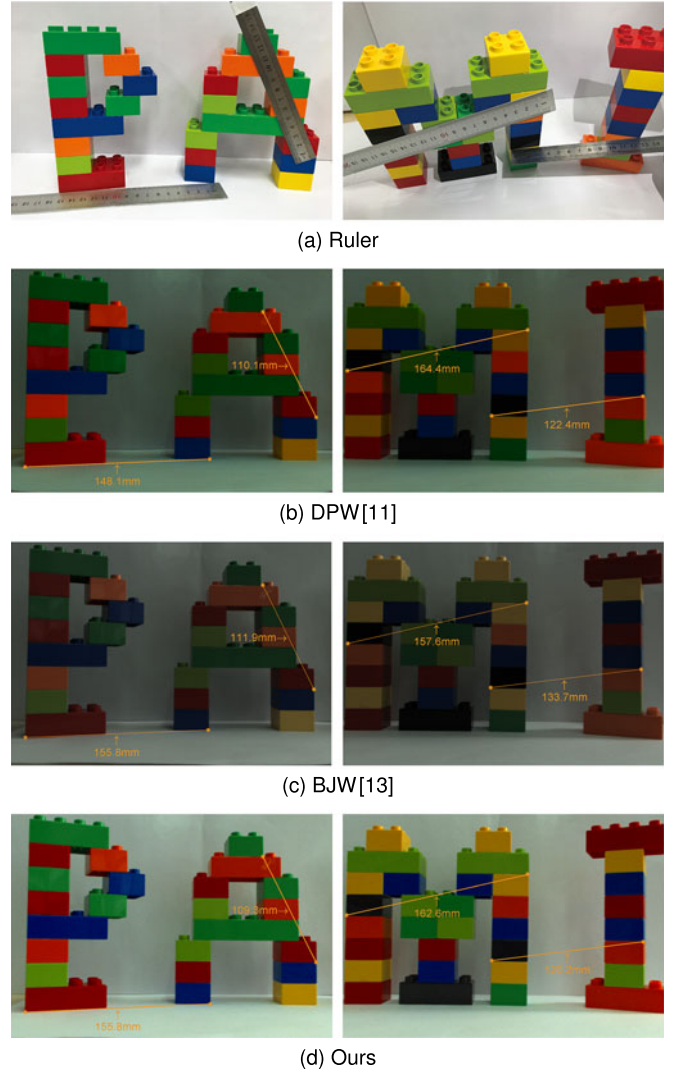(b) DPW[11]

(c) BJW[13]

(d) Ours

Fig. 13. The evaluations of light field measurements. (a) shows distances between 3D points measured by rulers. (b), (c) and (d) demonstrate the estimated distances between the reconstructed points after light filed camera calibration with different methods.

TABLE 12
Quantitative Comparison of Different Calibration Methods (Unit: mm)

	'P'	'A'	'M'	'I'
Ruler	155.0	108.5	163.0	128.5
DPW[11]	148.1 (4.45%)	110.1 (1.47%)	164.4 (0.86%)	122.4 (4.75%)
BJW[13]	155.8 (**0.52**%)	111.9 (3.13%)	157.6 (3.31%)	133.7 (4.05%)
Ours	155.8 (**0.52**%)	109.3 (**0.74**%)	162.6 (**0.25**%)	126.2 (**1.79**%)

The relative error is indicated in parentheses.

and extrinsic parameter estimation, we stitch all other light fields on the first pose, as shown in Fig. 11, from which we can see all view light fields are registered and stitched very well. Eventually, it is worthy noting that there exists distinct distortion in the Illum camera. In Fig. 12, we show original central view sub-aperture image and rectification results using distortion models of the proposed method, DPW [11] and BJW [13] respectively. Since the re-projection error indicates the image distance between a projected point and a rectified one, it can be used to quantify the error of distortion rectification results. In Fig. 12, we list the RMS re-projection error of central view sub-aperture image using different methods in parentheses, which further verifies that the rectification results of the proposed method are better than those of baseline algorithms.

High-precision calibration is essential in early stages of light field processing pipeline. In order to verify the accuracy of geometric reconstruction of the proposed method compared with baseline methods, we capture two real scene light fields, then reconstruct several typical corner points and estimate the distances between them. Fig. 13 shows reconstruction results on the central view sub-aperture images. As exhibited in Fig. 13, the estimated distances between points reconstructed by the proposed method are nearly equal to those measured lengths from real objects by rulers. In addition, Table 12 lists the comparisons of reconstruction results with state-of-the-art methods. The relative errors of reconstruction results demonstrate the performance of our method.

As mentioned above, since BJW has a stricter requirement on the image of calibration board, some datasets (i.e., C and D) [11] are incalculable for BJW. In order to directly compare with DPW and BJW, we utilize our collected datasets to analyze the running time of initial

TABLE 13
The Running Time of Initial Parameter Estimation (Unit: s)

	Illum-1	Illum-2	Lytro-1	Lytro-2
DPW [11]	10.7718	12.1735	4.3390	5.3729
BJW [13]	20.0266	24.2377	14.0859	15.2629
Ours	**0.2359**	**0.1878**	**0.1263**	**0.1428**

TABLE 14
Calibration Results of Intrinsic Parameters of Our Physical Camera with Different Number of Poses

#Pose		k_i	k_j	k_u	k_v	u_0	v_0
5	Init.	1.4497e-02	1.3755e-02	5.3071e-03	5.3068e-03	−0.1239	−0.2913
	$Opt_{w/o}$	1.4538e-02	1.3973e-02	4.9733e-03	4.9741e-03	−0.1719	−0.2965
	Opt_w	1.3710e-02	1.2990e-02	4.9732e-03	4.9740e-03	−0.1719	−0.2965
7	Init.	1.4160e-02	1.1321e-02	4.4666e-03	4.4723e-03	−0.1333	−0.2720
	$Opt_{w/o}$	1.4321e-02	1.1900e-02	4.6850e-03	4.6521e-03	−0.1732	−0.2901
	Opt_w	1.3363e-02	1.2466e-02	4.6850e-03	4.6521e-03	−0.1792	−0.2901
9	Init.	1.4066e-02	1.0478e-02	5.2007e-03	5.1898e-03	−0.1104	−0.2675
	$Opt_{w/o}$	1.4046e-02	1.0927e-02	4.7477e-03	4.4723e-03	−0.1645	−0.2812
	Opt_w	1.2917e-02	1.1741e-02	4.7476e-03	4.4722e-03	−0.1645	−0.2812

TABLE 15
The Distortion Vectors and RMS Re-Projection Errors (Unit: pixel) of Our Physical Camera with Different Number of Poses

#Pose	k_1	k_2	k_3	k_4	Error
5	5.4550e-04	6.0268e-05	−1.9239e-03	−2.3802e-03	0.6225
7	−2.4021e-04	−4.1706e-06	−2.1438e-03	1.5318e-03	0.6904
9	−3.7344e-04	−7.0466e-06	−2.5751e-03	2.3966e-03	0.7363

Fig. 14. The self-assembly focused light field camera and the MLA inside the camera. The light field design pattern is shown in Fig. 4a.

parameter estimation, as illustrated in Table 13. All algorithms are executed in MATLAB 2014 on a desktop computer (2.8 GHz CPU and 16G RAM). The proposed method is most efficient comparing to other methods. Since the computing of SVD in initial linear solution estimation of BJW is time-consuming (i.e., the running time of SVD on four datasets is 18.5277, 22.2501, 13.1004 and 14.2971 s respectively), the running time of BJW is shorter than that of DPW if the execution time of SVD is excluded. The results of running time conform well to our complexity analysis in Section 4.3.

5.2.2 Focused Light Field Camera

We capture data using a self-assembly focused light field camera. The camera and the MLA are shown in Fig. 14. The camera consists of a GigE camera with a CCD image sensor whose resolution is 4008×2672 with $9\mu m$ pixel width, a Nikon AF Nikkor f/1.4D F-mount lens with 50 mm focal length, and a MLA with $300\mu m$ diameter and 2.726 mm focal length in hexagon layout. The light field design pattern is shown in Fig. 14. The calibration board is 8×10 points with 20 mm $\times 20$ mm cells. We shoot 10 raw images with different poses of calibration board. Besides, we also collect 9 raw images with real scene and calibration board. All the images are stitched together to generate an enhanced light field with wider FOV.
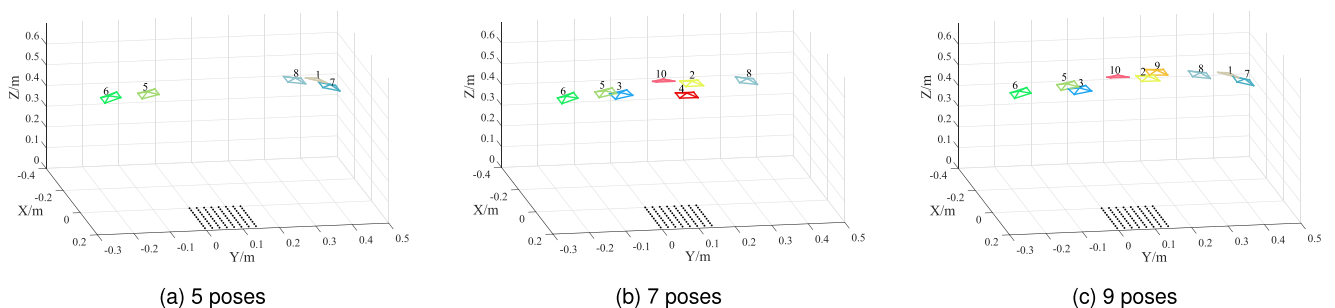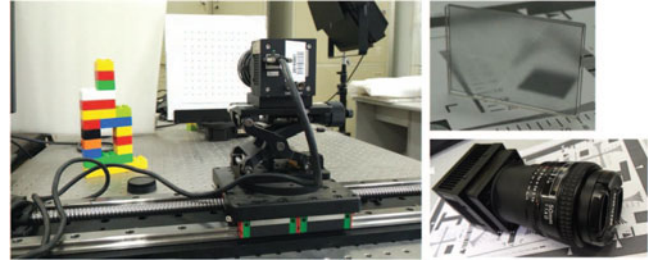
We carry out calibration using different numbers of light fields from 10 poses (i.e., 5 poses, 7 poses and 9 poses). The estimated intrinsic parameters of our physical camera are listed in Table 14. The top row shows the estimated results by direct linear initialization. The middle and bottom rows show the optimized results without and with distortion rectification. As shown in Table 15, the RMS re-projection errors are less than 0.75 pixels after the optimization with distortion rectification. Fig. 15 shows the estimated poses of our physical camera.

After the calibration, we render the images by ray tracing using 9 raw images with real scene and calibration board in different poses. The refocus rendering pipeline of a focused light field camera is shown in Fig. 16, which includes four key steps. The estimated poses and rendered results are shown in Fig. 17. On the stitched image in Fig. 17a, the dot-array calibration board on the background is focused and the blocks on the foreground are unfocused. The smooth boundary from different poses indicates that our calibration algorithm can estimate exact parameters.

6 CONCLUSION

In the paper, we present a multi-projection-center model to parameterize light field and describe light field imaging formation. We deduce the transformations to describe the relationship between 4D rays and 3D scene structure by a projective transformation. Then we verify our light field
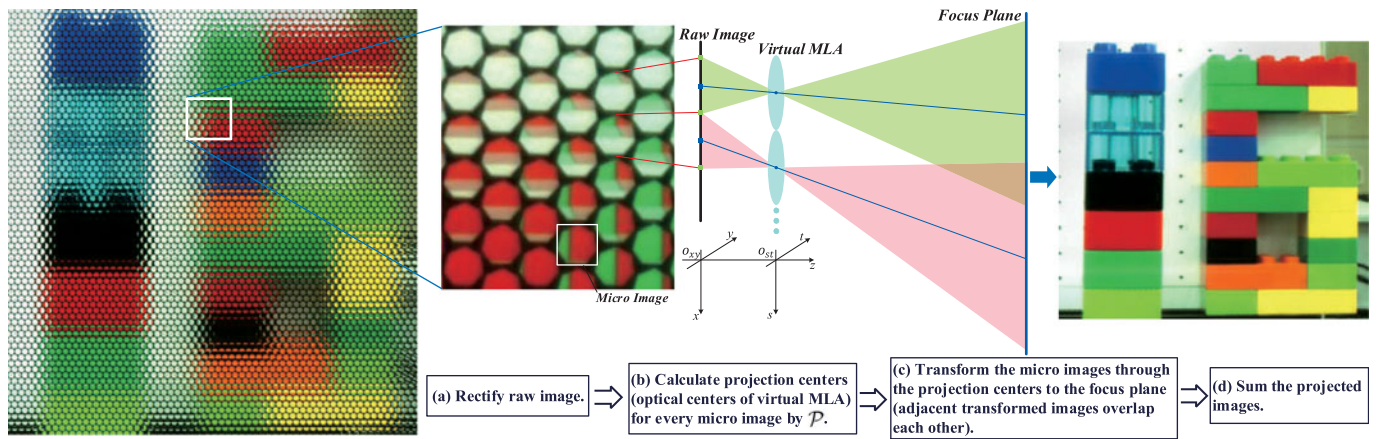

(a) 5 poses   (b) 7 poses   (c) 9 poses

Fig. 15. Pose estimation results of our self-assembly focused light field camera with different number of poses.

Fig. 16. The refocus rendering pipeline of a focused light field camera.
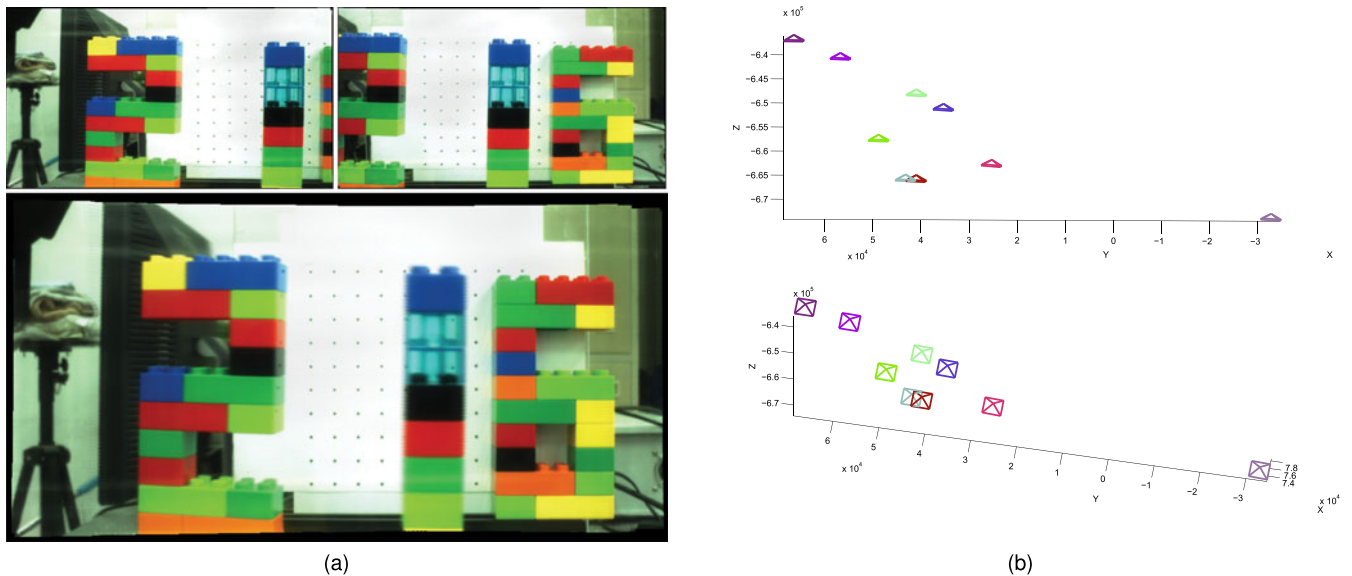

(a)

(b)

Fig. 17. Experimental results of light field stitching. (a) shows the rendered images. The top row shows leftmost and rightmost light fields captured by our focused light field camera. The bottom row is refocused result of the stitched light field from 9 light fields. (b) shows the estimated poses of 9 light fields.

camera model by intrinsic parameter calibration. We first derive a closed-form solution for initial estimation of intrinsic and extrinsic parameters and then propose a parameter refinement by minimizing the re-projection error. Experiments on conventional light field camera Lytro and Illum and a self-assembly focused light field camera are performed and analyzed extensively. The comparisons with ground truth and state-of-the-art calibration methods have verified the robustness and validity of the proposed model and our calibration method, especially the initialization and optimization with distortion rectification.

In future, we tend to focus on sampling distortion model on the view plane, light field registration and enhancement from un-calibrated cameras for arbitrary scenes, and re-parameterization of 4D light field from arbitrary poses.

REFERENCES

[1] R. Ng, "Digital light field photography," Ph.D. dissertation, Dept. Comput. Sci. Committee Graduate Stud., Stanford Univ., Stanford, CA, USA, 2006.
[2] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in Proc. IEEE Int. Conf. Comput. Photography, 2009, pp. 1–8.
[3] M. Levoy and P. Hanrahan, "Light field rendering," in Proc. ACM 23rd Annu. Conf. Comput. Graph. Interactive Tech., 1996, pp. 31–42.
[4] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in Proc. ACM 23rd Annu. Conf. Comput. Graph. Interactive Tech., 1996, pp. 43–54.
[5] R. Ng, "Fourier slice photography," ACM Trans. Graph., vol. 24, no. 3, pp. 735–744, 2005.
[6] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1547–1555.
[7] S. Wanner, C. Straehle, and B. Goldluecke, "Globally consistent multi-label assignment on the ray space of 4d light fields," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 1011–1018.
[8] O. Johannsen, A. Sulc, and B. Goldluecke, "On linear structure from motion for light field cameras," in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 720–728.
[9] C. Birklbauer and O. Bimber, "Panorama light-field imaging," Comput. Graph. Forum, vol. 33, no. 2, pp. 43–52, 2014.

[10] X. Guo, Z. Yu, S. B. Kang, H. Lin, and J. Yu, "Enhancing light fields through ray-space stitching," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 7, pp. 1852–1861, Jul. 2016.

[11] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1027–1034.

[12] O. Johannsen, C. Heinze, B. Goldluecke, and C. Perwaß, "On the calibration of focused plenoptic cameras," in *Proc. Seminar Time-Flight Depth Imag. Sens., Algorithms, Appl.*, 2013, pp. 302–317.

[13] Y. Bok, H.-G. Jeon, and I. S. Kweon, "Geometric calibration of micro-lens-based light-field cameras using line features," in *Proc. 13th Eur. Conf. Comput. Vis.*, 2014, pp. 47–61.

[14] C. Hahne, A. Aggoun, S. Haxha, V. Velisavljevic, and J. C. J. Fernández, "Light field geometry of a standard plenoptic camera," *Opt. Express*, vol. 22, no. 22, pp. 26 659–26 673, 2014.

[15] C. Thomason, B. Thurow, and T. Fahringer, "Calibration of a microlens array for a plenoptic camera," in *Proc. 52nd Aerosp. Sci. Meeting: Am. Inst. Aeronautics Astronautics*, 2014, pp. 1456–1460.

[16] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, 2005.

[17] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy, "Using plane+ parallax for calibrating dense camera arrays," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 2–9, 2004.

[18] R. Pless, "Using many cameras as one," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, pp. II–587, 2003.

[19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[20] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 972–986, May 2012.

[21] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 3280–3287.

[22] S. Wanner, J. Fehr, and B. Jähne, "Generating epi representations of 4d light fields with a single lens focused plenoptic camera," in *Proc. 7th Int. Conf. Adv. Visual Comput.*, 2011, pp. 90–101.

[23] Y. Bok, H.-G. Jeon, and I. S. Kweon, "Geometric calibration of micro-lens-based light field cameras using line features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 287–300, Feb. 2017.

[24] C. Hahne, A. Aggoun, and V. Velisavljevic, "The refocusing distance of a standard plenoptic photograph," in *Proc. Conf. 3DTV*, 2015, pp. 1–4.

[25] C. Heinze, S. Spyropoulos, S. Hussmann, and C. Perwaß, "Automated robust metric calibration of multi-focus plenoptic cameras," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, 2015, pp. 2038–2043.

[26] C. Zhang, Z. Ji, and Q. Wang, "Decoding and calibration method on focused plenoptic camera," *Comput. Visual Media*, vol. 2, no. 1, pp. 57–69, 2016.

[27] R. Hartley, "Self-calibration of stationary cameras," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 5–23, 1997.

[28] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA, USA: MIT Press, 1993.

[29] K. Madsen, H. B. Nielsen, and O. Tingleff, *Methods for Non-Linear Least Squares Problems*, 2nd ed. Kongens Lyngby, Denmark: Informatics and Mathematical Modelling, Technical University of Denmark, April 2004.

Qi Zhang received the BE degree in electronic and information engineering from Xi'an University of Architecture and Technology, in 2013, and the master's degree in electrical engineering from Northwestern Polytechnical University, in 2015. He is working toward the PhD degree in the School of Computer Science, Northwestern Polytechnical University. His research interests include computational photography, light field imaging and processing, multiview geometry and applications.

Chunping Zhang received the BE and master's degrees in computer science from the School of Computer Science, Northwestern Polytechnical University, in 2014 and 2017, respectively. Her research interests include computational photography, light field camera model and calibration, multiview stereo reconstruction.
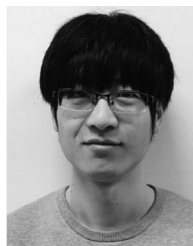
Jinbo Ling received the BE degree in software engineering from the School of Information Engineering, Chang'an University, in 2016. He is working toward the postgraduate degree in the School of Computer Science, Northwestern Polytechnical University. His research interests include 3D reconstruction, multiview light field computing and processing.

Qing Wang (M'05) received the BS degree from Peking University, in 1991, and the master's and PhD degrees from Northwestern Polytechnical University, in 1997 and 2000, respectively. He is now a professor with the School of Computer Science, Northwestern Polytechnical University. After receiving of BS degree, he joined Northwestern Polytechnical University. In 2006, he was awarded as outstanding talent program of new century by Ministry of Education, China. He worked as research assistant and research scientist with the Department of Electronic and Information Engineering, the Hong Kong Polytechnic University from 1999 to 2002. He also worked as a visiting scholar with the School of Information Engineering, The University of Sydney, Australia, in 2003 and 2004. In 2009 and 2012, he visited Human Computer Interaction Institute, Carnegie Mellon University, for six months and Department of Computer Science, University of Delaware, for one month. His research interests include computer vision and computational photography, such as 3D reconstruction, object detection, tracking and recognition in dynamic environment, light field imaging and processing. He has published more than 100 papers in the international journals and conferences. He is a member of the IEEE and ACM. He is also a senior member of China Computer Federation (CCF).

Jingyi Yu received the BS degree from Caltech, in 2000 and the PhD degree from MIT, in 2005. He is the director of Virtual Reality and Visual Computing Center with the School of Information Science and Technology at ShanghaiTech University. He is also a full professor with the University of Delaware. His research interests include computer vision and computer graphics, especially on computational photography and nonconventional optics and camera designs. He has published more than 100 papers at highly refereed conferences and journals including more than 50 papers at the premiere conferences and journals CVPR, ICCV, ECCV, the *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He has been granted 10 US patents. His research has been generously supported by the National Science Foundation (NSF), the National Institute of Health (NIH), the Army Research Office (ARO), and the Air Force Office of Scientific Research (AFOSR). He is a recipient of the NSF CAREER Award, the AFOSR YIP Award, and the Outstanding Junior Faculty Award at the University of Delaware. He has previously served as general chair, program chair, and area chair of many international conferences such as ICCV, ICCP and NIPS. He is currently an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, the *Computer Vision and Image Understanding* (Elsevier), *The Visual Computer* (Springer) and the *Journal of Machine Vision and Applications* (Springer).

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.