Ray-Space Epipolar Geometry for Light Field Cameras

Qi Zhang, Qing Wang, *Senior Member, IEEE,* Hongdong Li, and Jingyi Yu,

Abstract—Light field essentially represents rays in space. The epipolar geometry between two light fields is an important relationship that captures ray-ray correspondences and relative configuration of two views. Unfortunately, so far little work has been done in deriving a formal epipolar geometry model that is specifically tailored for light field cameras. This is primarily due to the high-dimensional nature of the ray sampling process with a light field camera. This paper fills in this gap by developing a novel *ray-space epipolar geometry* which intrinsically encapsulates the complete projective relationship between two light fields, while the generalized epipolar geometry which describes relationship of normalized light fields is the specialization of the proposed model to calibrated cameras. With Plücker parameterization, we propose the ray-space projection model involving a 6×6 *ray-space intrinsic matrix* for ray sampling of light field camera. Ray-space fundamental matrix and its properties are then derived to constrain ray-ray correspondences for general and special motions. Finally, based on ray-space epipolar geometry, we present two novel algorithms, one for fundamental matrix estimation, and the other for calibration. Experiments on synthetic and real data have validated the effectiveness of ray-space epipolar geometry in solving 3D computer vision tasks with light field cameras.

Index Terms—Ray-Space Epipolar Geometry; Ray-Space Fundamental Matrix; Light Field Camera; Plücker Parameterization.

1 Introduction

Light field camera (LFC) such as Lytro [1] and Raytrix [2] can record spatial and angular information of rays in 3D space. Based on the angular sampling of light rays, advanced multiple-view 3D vision problems such as structure-from-motion (SfM), light field stitching, and more robust SLAM have been investigated (e.g., [3], [4], [5], [6] [7], [8], [9], [10], [11], [12], [13]). To facilitate these multi-view light fields based 3D applications, it is desirable to have a unified geometric framework that encapsulates two view ray-ray correspondences, just like the conventional two-view epipolar geometry for the conventional pinhole camera. However, so far, little work has been done along this line of research for light field cameras. Although an LFC can be treated as an array of pinhole cameras, repeatedly applying the traditional epipolor geometry to each individual sub-aperture images remains a unduly onerous task. Moreover, given that the viewpoints of an LFC are regularly arranged on a planar grid, treating each sub-aperture as a pinhole camera is unable to capture the more natural ray-ray relations among different LFCs. It is much desirable to have a dedicated "epipolar geometry" theory between two light fields that can uniformly constrain ray-ray correspondence and compute projection matrices for light fields, and this is the central motivation of this paper.

- *Qi Zhang and Qing Wang (corresponding author) are with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: qwang@nwpu.edu.cn).*
- *Hongdong Li is with the ANU and ACRV, Australian National University, Australia (e-mail: hongdong.li@anu.edu.au)*
- *Jingyi Yu is with the ShanghaiTech University, Shanghai 200031, China (e-mail: jingyi.udel@gmail.com)*

- *The work was supported by NSFC under Grant 61531014, 61801396, 62031023.*

Manuscript received April 19, 2005; revised August 26, 2015.

Existing LFC models [14], [15] mostly define the projection from an arbitrary point in 3D space (passing through micro-lens) to the corresponding pixel on the sensor, but only focused on monocular LFC. A reliable mathematical mechanism is necessary to uniformly describe ray transformations with intrinsic and extrinsic parameters. The Plücker coordinate explicitly provides a homogeneous parameterization for rays to effectively formulate ray-ray correspondence, whose performance has been verified in generalized epipolar geometry [3], [16]. However, generalized epipolar geometry [16] only defines the relationship of normalized light fields (calibrated LFCs). It is crucial to generalize this model to comprehensively describe complete projective geometry of light fields including intrinsic parameters. To intrinsically explore the complete projective geometry of light fields, LFC projection model and intrinsic parameters suited for Plücker parameterization are indispensable. Dansereau *et al.* [17] describe pixel-ray correspondences and present a 4D intrinsic matrix. However, their model has redundancy and dependency, which makes the Plücker representation impossible. Zhang *et al.* [15] propose anther state-of-the-art projection model, which provides independent and effective intrinsic parameters for the Plücker representation (as also verified in the shorter version of our work [18]).

To our knowledge, the shorter version of our work [18] is the first to intrinsically explore ray-ray correspondence and generalize a ray-space fundamental matrix from [16] for multi-view light fields instead of normalized light fields. This paper significantly extends [18] with the ray-space epipolar geometry based on the proposed ray-space projection model, along with properties and corollaries of the proposed fundamental matrix, just as what has been developed for traditional pinhole cameras. We also present an efficient fundamental matrix estimation algorithm.

Our main contributions are:

1) The ray-space epipolar geometry among light field cameras is exploited based on ray-space projection model.
2) The properties of ray-space fundamental matrix are elucidated with Plücker parameterization, both for general and special motions.
3) Two novel algorithms, including fundamental matrix estimation and light field camera calibration, are proposed to verify the proposed epipolar geometry.

2 Related Work

2.1 LFC Projection Model

Since the hand-held LFC is put forwarded by Ng [19], many research groups [14], [15], [17], [18], [20], [21] have extensively explored various projection models for LFCs. In general, the LFC models can be roughly divided into three categories.

Dansereau *et al.* [17] first propose a 12-free-parameter LFC model, corresponding the recorded pixels to the rays outside the camera. They derive a 4D decoding matrix for ray sampling. However, the 4D intrinsic matrix has redundancy and dependency, which results in irregular rays sampling during the calibration and rectification. It is difficult to uniformly describe ray sampling with intrinsic parameters and ray transformation with extrinsic parameters (as also verified in [7]).

Different from the calibration based on corner features of sub-aperture images, Bok *et al.* [14] utilize line features which are directly extracted from raw data to calibrate an LFC. They formulate a 6-parameter coupling geometric projection model with clear physical meaning for an LFC to relate the scene point to raw data. Taken the point projection into consideration, the coupling of intrinsic parameters confronts a significant challenge to linearly express the ray sampling and transformation.

More recently, Zhang *et al.* [15] propose a 6-parameter multi-projection-center (MPC) model with clear physical meaning for LFCs, including traditional and focused LFCs. A 3D projective transformation is deduced to describe the relationship of geometric structure between light filed and camera coordinate frames. The projections of an LFC on planes and conics are also explored under MPC model [22], [23]. Considering the independence of intrinsic parameters and advantages of Plücker parameterization, it is convincing to uniformly describe ray sampling and transformation with the Plücker parameterization. Consequently, based on the MPC model, a ray-space projection model is proposed to correspond the ray recorded by an LFC to the ray in space in a shorter version of our work [18], which also verifies the convenience and effectiveness of Plücker parameterization.

2.2 Generalized Epipolar Geometry

Epipolar geometry is proposed to constrain image points correspondence and reconstruct camera geometry for traditional cameras over the decades [24]. In order to conveniently represent image point and estimate relation among cameras, Grossberg and Nayar [25] first define the image pixel as the light from a cone around a ray and propose a generalized camera model. Pless [16] then simplifies this model so that it only includes the definition of ray that the pixel samples. A general linear framework is proposed to describe any cameras as an unordered collection of normalized rays which are obtained from sensor elements via calibration. The correspondences between normalized rays need to be established with the assumption that these rays intersect at a single scene point. Then, the generalized epipolar geometry is proposed to constrain normalized ray-ray correspondences with Plücker parameterization and prior calibration. The generalized epipolar geometry is the projective geometry for calibrated cameras which only relies on the relative pose. Sturm [26] introduces a hierarchy of general camera model. In this framework, 17 corresponding normalized rays are sufficient to solve linearly for pose estimation. Li *et al.* [27] carry out a pose estimation based on the generalized epipolar constraint. This can also be applied to estimate the motion of calibrated LFCs.

Guo *et al.* [8] propose a ray-space motion matrix to establish normalized ray-ray transformation for motion estimation. Moreover, Johannsen *et al.* [3] extend the generalized epipolar constraint into point-ray constraint. A linear mathematical framework is built from the relationships between scene geometry and normalized rays for motion estimation. In summary, the existing methods utilize generalized epipolar geometry to describe the projective geometry of normalized light fields (calibrated LFCs), which *only* relies on the relative pose. However, an LFC essentially records the scenes via light fields instead of normalized light fields. It is incomprehensive to describe the complete projective geometry of LFCs without considering intrinsic parameters. Consequently, similar to the traditional epipolar geometry for image point-point correspondence, it is essential to uniformly define the ray-space epipolar geometry for light field ray-ray correspondences through generalizing from [16]. It depends on intrinsic parameters and relative pose, and widely extends applications for multi-view light fields, whereas generalized epipolar geometry is specialized to estimate motion with pre-calibration.

2.3 Fundamental Matrix

Fundamental matrix has gained increasing attention since the seminal work presented by Higgins [28]. Hartley [29] presents a fundamental matrix estimation algorithm from arbitrary seven correspondences according to the rank-2 constraint and solves the cubic polynomial equation. Barath [30] estimates the fundamental matrix in two views from five correspondences with some assumptions, *i.e.* co-planar three correspondences and arbitrary two correspondences. In order to improve numerical stability for fundamental matrix estimation, a simple normalized transformation of corresponding point is involved [31]. Zhou *et al.* [32] propose a normalization algorithm to estimate the fundamental matrix from at least three plane homographies. Moreover, many existing methods utilize fewer correspondences to estimate fundamental matrices with strict constraints. (*e.g.* known principle points [33], calibrated camera [34], [35], and special camera motion [24], [36]).

A seemingly straightforward choice for fundamental matrix of two light fields is to consider it as a direct extension of that for a monocular pinhole camera. Treating

sub-aperture images independently, however, creates larger image sets, and defies the aim that light field essentially represents rays which are regularly recorded by an LFC. Therefore, it is essential to exploit the ray-space epipolar geometry of two light fields and propose a ray-space fundamental matrix specifically designed for LFCs.

3 Ray-Space Projection Model

3.1 The Multi-Projection-Center Model

LFCs, especially those micro-lens array based, represent an innovative departure from the traditional pinhole camera. With the shifted views, an LFC maps 3D space to many sub-aperture images, which produces 4D light field. The ray in 4D light field is parameterized in a *relative* two-parallel-plane coordinates [37], where $Z=0$ denotes the view plane and $Z=f$ for the image plane. In this parameterization, the normalized physical ray is described as $\boldsymbol{r}=(s,t,x,y)^\top$ in term of specific (*e.g.*, meter) dimension. The ray $\boldsymbol{r}$ intersects with the view plane at projection center $(s,t,0)^\top$. The pair (x,y) is the intersection of the ray $\boldsymbol{r}$ with the image plane, but it is relative to $(s,t,f)^\top$ which is the origin of image plane. The $(x,y,f)^\top$ describes the direction of ray. Then, according to the MPC model [15], a 3D point $\boldsymbol{X}=(X,Y,Z)^\top$ is mapped to the pixel $(x,y)^\top$ in the image plane,

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

This is analogous to classical projective camera model with projection center at $(s,t,0)^\top$ and principal axis paralleling to the Z-axis.

The ray $\boldsymbol{l}=(i,j,u,v)^\top$ captured by an LFC in light field coordinate frame is transformed into a normalized undistorted physical ray $\boldsymbol{r}$ in camera coordinate frame by a homogeneous decoding matrix $\boldsymbol{D}\in\mathbb{R}^{5\times5}$ [15],

$$\begin{bmatrix}s\\t\\x\\y\\1\end{bmatrix}=\begin{bmatrix}k_i&0&0&0&0\\0&k_j&0&0&0\\0&0&k_u&0&u_0\\0&0&0&k_v&v_0\\0&0&0&0&1\end{bmatrix}\begin{bmatrix}i\\j\\u\\v\\1\end{bmatrix}, \tag{2}$$

where $(k_i,k_j,k_u,k_v,u_0,v_0)$ are intrinsic parameters of an LFC. (k_i,k_j) are scale factors for s and t axes in the view plane and (k_u,k_v) for x and y axes in the image plane respectively. In addition, $(-u_0/k_u,-v_0/k_v)^\top$ represents the principal point in the sub-aperture image.

3.2 Ray-Space Intrinsic Matrix

According to the MPC model, an LFC is assumed as an array of pinhole cameras. In this framework, a light field is described as a set of sub-aperture images recorded by a collection of perspective cameras. In order to simplify the discussion of geometric analysis in multiple light fields, the pixel captured by an LFC is generalized and simplified to a ray [16], [19]. The light field essentially represents all rays as a whole. Consequently, we need a new mechanism to describe arbitrary rays in 3D projective space. The Plücker parameterization provides convenience to mathematically formulate concise and efficient correspondence equations (*e.g.*, rotation and translation). In addition, the Plücker coordinate is also a homogeneous parameterization to unambiguously represent a ray in 3D projective geometry. We will briefly review the core theory leading the equations for ray-space projection model.

With the introduction of Plücker parameterization, the ray is mathematically represented by a pair of vectors $(\boldsymbol{m}^\top,\boldsymbol{q}^\top)^\top\in\mathbb{R}^6$, named moment and direction vectors respectively. Moreover, the moment vector denotes $\boldsymbol{m}=\boldsymbol{X}\times\boldsymbol{q}$, for an arbitrary point $\boldsymbol{X}$ on the ray. Further, as mentioned above, the physical ray $\boldsymbol{r}=(s,t,x,y)^\top$ in 3D space contains directional sampling $(x,y)^\top$ and positional sampling $(s,t)^\top$ of the ray. Therefore, with the simplicity that the interval between two-parallel-plane f is set to unit, the moment vector and direction vector of $\boldsymbol{r}$ are defined as [38],

$$\begin{cases}\boldsymbol{m}=(s,t,0)^\top\times(x,y,1)^\top=(t,-s,sy-tx)^\top\\ \boldsymbol{q}=(x,y,1)^\top\end{cases}, \tag{3}$$

where $\mathcal{R}=(\boldsymbol{m}^\top,\boldsymbol{q}^\top)^\top$ is a Plücker coordinate.

Substituting Eq. (2) into Eq. (3), there is a transformation caused by the intrinsic parameters $(k_i,k_j,k_u,k_v,u_0,v_0)$. Then the ray-space intrinsic matrix (RSIM) $\boldsymbol{K}\in\mathbb{R}^{6\times6}$ is established to describe the relationship between the ray $\mathcal{L}=(\boldsymbol{n}^\top,\boldsymbol{p}^\top)^\top$ in light field coordinate frame and the normalized undistorted physical ray $\mathcal{R}=(\boldsymbol{m}^\top,\boldsymbol{q}^\top)^\top$ in camera coordinate frame with the Plücker parameterization,

$$\begin{bmatrix}\boldsymbol{m}\\\boldsymbol{q}\end{bmatrix}=\underbrace{\begin{bmatrix}k_j&0&0&0&0&0\\0&k_i&0&0&0&0\\-k_ju_0&-k_iv_0&k_ik_v&0&0&0\\0&0&0&k_u&0&u_0\\0&0&0&0&k_v&v_0\\0&0&0&0&0&1\end{bmatrix}}_{=:\boldsymbol{K}}\begin{bmatrix}\boldsymbol{n}\\\boldsymbol{p}\end{bmatrix}, \tag{4}$$

which needs to satisfy the condition $k_u/k_v=k_i/k_j$. $(u,v)^\top$ is pixel coordinate extracted from sub-aperture image at the view coordinate $(i,j)^\top$. Then, $\boldsymbol{n}=(i,j,0)^\top\times(u,v,1)^\top=(j,-i,iv-ju)^\top$ and $\boldsymbol{p}=(u,v,1)^\top$ represent the moment and direction vectors respectively.

Besides, the RSIM is abbreviated to a 3×3 lower triangular matrix $\boldsymbol{K}_{ij}$ and a 3×3 upper triangular matrix $\boldsymbol{K}_{uv}$. Given that moment vector also implies the 3D points lying on the ray, two corollaries of RSIM are derived.

Corollary 1. *Block intrinsic matrices $\boldsymbol{K}_{ij}$ and $\boldsymbol{K}_{uv}$ are orthogonality, i.e. $\boldsymbol{K}_{ij}\boldsymbol{K}_{uv}^\top=\boldsymbol{K}_{ij}^\top\boldsymbol{K}_{uv}=k_ik_v\boldsymbol{I}$.*

Corollary 2. *Suppose two corresponding 3D points $\boldsymbol{X}_d$ and $\boldsymbol{X}$ lying on $\mathcal{L}$ and $\mathcal{R}$ respectively, $\boldsymbol{X}_d$ and $\boldsymbol{X}$ are related by $\boldsymbol{X}_d=\frac{1}{k_ik_j}\boldsymbol{K}_{ij}^\top\boldsymbol{X}=\frac{k_u}{k_i}\boldsymbol{K}_{uv}^{-1}\boldsymbol{X}$, which is the same as 3D perspective transformation in MPC model [15].*

3.3 Ray-Space Projection Matrix

In general, considering $\boldsymbol{X}_w$ denotes a point in the world coordinate frame, the transformation between the world and camera coordinate frames is described by a rotation matrix $\boldsymbol{R}\in SO(3)$ and a translation vector $\boldsymbol{t}=(t_x,t_y,t_z)^\top\in\mathbb{R}^3$, formulated as $\boldsymbol{X}=\boldsymbol{R}\boldsymbol{X}_w+\boldsymbol{t}$. Consequently, the Plücker

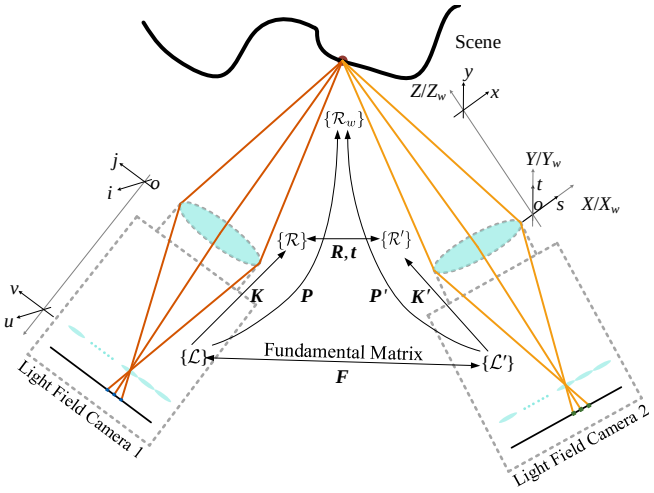


Fig. 1. RSP model and fundamental matrix among LFCs.

transformation can be formulated according to generalized epipolar geometry [16],

$$\mathcal{R}_w = \begin{bmatrix} \boldsymbol{R}^\top & \boldsymbol{E}^\top \\ \boldsymbol{O}_{3\times3} & \boldsymbol{R}^\top \end{bmatrix} \mathcal{R}, \tag{5}$$

where $\boldsymbol{E} = [\boldsymbol{t}]_\times \boldsymbol{R}$ is the essential matrix and $[\cdot]_\times$ refers to the vector cross product [24]. $\mathcal{R} = (\boldsymbol{m}^\top, \boldsymbol{q}^\top)^\top$ and $\mathcal{R}_w = (\boldsymbol{m}_w^\top, \boldsymbol{q}_w^\top)^\top$ are expressed the rays in the camera and world coordinate frames respectively. Subsequently, according to Eqs. (3) and (4), the homogeneous ray-space projection matrix (RSP) $\boldsymbol{P}$ can be written as,

$$\mathcal{R}_w = \underbrace{\begin{bmatrix} \boldsymbol{R}^\top & \boldsymbol{E}^\top \\ \boldsymbol{O}_{3\times3} & \boldsymbol{R}^\top \end{bmatrix} \boldsymbol{K}}_{=:\boldsymbol{P}} \mathcal{L}, \tag{6}$$

which refers to the transformation between $\mathcal{L}$ in light field coordinate frame and $\mathcal{R}_w$ in the world coordinate frame as shown in Fig. 1.

Furthermore, for arbitrary Plücker ray $\mathcal{R}$, it satisfies the self-constraint $\mathcal{R}^\top \boldsymbol{\Omega} \mathcal{R} = 0$ [39],

$$\mathcal{R}^\top \begin{bmatrix} \boldsymbol{O}_{3\times3} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{O}_{3\times3} \end{bmatrix} \mathcal{R} = 0, \tag{7}$$

where $\boldsymbol{\Omega}$ refers to the *Klein quadric* that contains all rays with Plücker parameterization in 5D projective space $\mathbb{P}^5$. As shown in Eq. (7), a Plücker ray indicates a point on Klein quadric in $\mathbb{P}^5$. Let $\mathcal{L}$, $\mathcal{R}$ and $\mathcal{R}_w$ denote the same ray in different coordinate frames and relate to each other under RSIM $\boldsymbol{K}$ and RSP $\boldsymbol{P}$, as shown in Fig. 1. According to Eq. (7) and *Corollary* 1, we thereafter obtain a corollary of RSP,

Corollary 3. *The projections of Klein quadric $\boldsymbol{\Omega}$ under RSIM $\boldsymbol{K}$ and RSP $\boldsymbol{P}$ are equal to Klein quadric up to a scale, i.e. $\boldsymbol{P}^\top \boldsymbol{\Omega} \boldsymbol{P} = \boldsymbol{K}^\top \boldsymbol{\Omega} \boldsymbol{K} = k_i k_v \boldsymbol{\Omega}$.*

Remarks. The presences of $\boldsymbol{K}$ and $\boldsymbol{P}$ may be expressed by saying that a quaric transforms invariantly. Moreover, the scale can be ignored due to the homogeneity of Plücker rays.

4 Ray-Space Epipolar Geometry

In essence, RSP model is a unified framework that considers all rays collected by the LFC as a whole. We derive the dedicated multi-view geometry relationship for LFCs, just as what has been developed for the traditional perspective cameras. The ray-space epipolar geometry is the 5D intrinsically complete projective geometry between two light fields. It is independent of scene structure and only depends on the RSIM and relative pose. The ray-space fundamental matrix $\boldsymbol{F}$ encapsulates this intrinsically complete projective geometry and is unchanged by projective transformation.

4.1 Ray-Space Fundamental Matrix

In order to constrain the ray-ray correspondences between two LFCs, the ray-space fundamental matrix $\boldsymbol{F}$ is proposed,

Corollary 4. *The ray-space fundamental matrix satisfies the constraint $\mathcal{L}^\top \boldsymbol{F} \mathcal{L}' = 0$ for any ray-ray correspondences $\mathcal{L} \leftrightarrow \mathcal{L}'$ in two light fields.*

Proof. Given RSP matrices for two LFCs, the second camera coordinate frame is assumed as the world coordinate frame as shown in Fig. 1,

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{R}^\top & \boldsymbol{E}^\top \\ \boldsymbol{O}_{3\times3} & \boldsymbol{R}^\top \end{bmatrix} \boldsymbol{K}, \ \boldsymbol{P}' = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{O}_{3\times3} \\ \boldsymbol{O}_{3\times3} & \boldsymbol{I} \end{bmatrix} \boldsymbol{K}'. \tag{8}$$

Suppose two intersecting rays $\mathcal{R}_w = (\boldsymbol{m}_w^\top, \boldsymbol{q}_w^\top)^\top$ and $\mathcal{R}'_w = (\boldsymbol{m}'^\top_w, \boldsymbol{q}'^\top_w)^\top$ are captured by two LFCs respectively. $\mathcal{R}_w$ and $\mathcal{R}'_w$ satisfy $\boldsymbol{q}_w^\top \boldsymbol{m}'_w + \boldsymbol{m}_w^\top \boldsymbol{q}'_w = 0$ [39], or in bilinear form $\mathcal{R}_w^\top \boldsymbol{\Omega} \mathcal{R}'_w = 0$. Let $\mathcal{L}$ and $\mathcal{L}'$ denote the rays captured by the first and second LFCs respectively. We then derive a geometry constraint between $\mathcal{L}$ and $\mathcal{L}'$ under projections $\boldsymbol{P}$ and $\boldsymbol{P}'$, *i.e.* $\mathcal{L}^\top \boldsymbol{F} \mathcal{L}' = 0$,

$$\mathcal{L}^\top \boldsymbol{P}^\top \begin{bmatrix} \boldsymbol{O}_{3\times3} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{O}_{3\times3} \end{bmatrix} \boldsymbol{P}' \mathcal{L}' = 0, \tag{9}$$

where $\boldsymbol{F}$ is uniquely generated by a pair of RSP matrices $(\boldsymbol{P}, \boldsymbol{P}')$, and $\mathcal{L}$ is conjugate to $\mathcal{L}'$ with respect to $\boldsymbol{F}$. □

Besides, substituting Eq. (8) into Eq. (9), $\boldsymbol{F}$ is partitioned into 2×2 block matrices,

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{O} & \boldsymbol{F}_{12} \\ \boldsymbol{F}_{21} & \boldsymbol{F}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{O} & \boldsymbol{K}_{ij}^\top \boldsymbol{R} \boldsymbol{K}'_{uv} \\ \boldsymbol{K}_{uv}^\top \boldsymbol{R} \boldsymbol{K}'_{ij} & \boldsymbol{K}_{uv}^\top \boldsymbol{E} \boldsymbol{K}'_{uv} \end{bmatrix}, \tag{10}$$

where $\boldsymbol{F}_{ij}$ is a 3×3 block matrix.

Consequently, Eq. (9) is restated,

$$\boldsymbol{p}^\top \boldsymbol{F}_{21} \boldsymbol{n}' + \boldsymbol{p}^\top \boldsymbol{F}_{22} \boldsymbol{p}' + \boldsymbol{n}^\top \boldsymbol{F}_{12} \boldsymbol{p}' = 0. \tag{11}$$

For a valid correspondence, all rays in both light fields must come from the same scene point, as shown in Fig. 1.

Subsequently, according to Eq. (10) and *Corollary* 1, two important constraints of $\boldsymbol{F}$ are deduced for the sake of $\boldsymbol{F}$ computation as follows,

1) **Orthogonal Constraint.** $\boldsymbol{F}_{21}$ and $\boldsymbol{F}_{12}$ are orthogonal, *i.e.* $\boldsymbol{F}_{12}^\top \boldsymbol{F}_{21} = \boldsymbol{F}_{21}^\top \boldsymbol{F}_{12} = \lambda \boldsymbol{I}$.
2) **Singular Constraint.** $\boldsymbol{F}_{22}$ is singular, in fact of rank 2.

Remarks. $\boldsymbol{F}$ has 17 degrees of freedom: $\boldsymbol{F}$ has 26 independent non-zero elements (27–1, one for global scale); however $\boldsymbol{F}$ also satisfies the orthogonal and singular constraints which decreases the degrees of freedom to 17, (26−9−1+1, 9 for orthogonal constraint, 1 for singular constraint, the last +1 for unknown factor λ).

Considering that $\boldsymbol{n}$ is the moment vector, Eq. (11) can also be derived to indicate the fundamental matrix between arbitrary sub-aperture images of two light fields,

$$\boldsymbol{p}^\top \left(\boldsymbol{F}_{21} \left[(i', j', 0)^\top \right]_\times + \boldsymbol{F}_{22} - \left[(i, j, 0)^\top \right]_\times \boldsymbol{F}_{12} \right) \boldsymbol{p}' = 0, \tag{12}$$

where $\boldsymbol{F}_{22}$ indicates the fundamental matrix between central views of two light fields.

Moreover, according to Eq. (9) and *Corollary* 3, we now turn to another crucial corollary of $\boldsymbol{F}$, that the matrix may be used to determine the RSP matrices of two light fields.

Corollary 5. *Suppose $\boldsymbol{P}_T$ is a RSP matrix representing a homogeneous ray-ray transformation, the ray-space fundamental matrices corresponding to the pairs of RSP matrices $(\boldsymbol{P}, \boldsymbol{P}')$ and $(\boldsymbol{P}_T\boldsymbol{P}, \boldsymbol{P}_T\boldsymbol{P}')$ are the same.*

Remarks. Despite that a pair of RSP matrices uniquely determines $\boldsymbol{F}$ from Eq. (9), the converse is not true. It is easy to observe that RSP matrices can be computed from the fundamental matrix up to a projective ambiguity. It can also be applied for scene reconstruction and projective rectification. In addition, if $\boldsymbol{F}$ is the ray-space fundamental matrix of a pair of LFCs $(\boldsymbol{P}, \boldsymbol{P}')$, then $\boldsymbol{F}^\top$ is the fundamental matrix of the pair in the opposite order $(\boldsymbol{P}', \boldsymbol{P})$.

4.2 Ray-Space Epipolar Geometry

The ray-space epipolar geometry between two light fields is essentially the geometry of two ray bundles in $\mathbb{P}^3$. According to Eq. (7), all rays with Plücker parameterization satisfy self-constraint of Klein quadric $\boldsymbol{\Omega}$ in $\mathbb{P}^5$, which is also a special case of Grassmann manifold. Subsequently, the relation between the 4-dimensional ray in $\mathbb{P}^3$ and the point of $\boldsymbol{\Omega}$ in $\mathbb{P}^5$ is bijective correspondence, as shown in Fig. 2. The set of rays $\{\mathcal{L}\}$ intersecting at the single point is called a ray bundle. A Klein quadric carries a 3-parameter family of 2-dimensional subspaces (*epipolar hyperplanes*) in $\mathbb{P}^5$, which corresponds to ray bundles in $\mathbb{P}^3$.

It is interesting to note that the "epipolar line" for an LFC is epipolar hyperplane, symbolized by $\boldsymbol{\Pi}$. The ray-space fundamental matrix $\boldsymbol{F}$ is the algebraic representation of epipolar geometry for LFCs. Consequently, we have considered the mapping $\mathcal{L}' \mapsto \boldsymbol{\Pi}$ defined by $\boldsymbol{F}$. In other words, the ray $\mathcal{L}$ which is equivalent to the points on the Klein quadric $\boldsymbol{\Omega}$ lies on the epipolar hyperplane $\boldsymbol{\Pi}$ mapped from $\mathcal{L}'$ by $\boldsymbol{F}$ in $\mathbb{P}^5$ (*i.e.* $\mathcal{L}^\top \boldsymbol{\Pi} = 0$), as shown in Fig. 2. $\boldsymbol{F}$ is a *perspective correlation* which maps points to hyperplane on Klein quadric in $\mathbb{P}^5$. $\boldsymbol{F}$ represents a mapping between two 2-dimensional subspaces, and hence is a full rank matrix. The full rank $\boldsymbol{F}$ means that there is inverse mapping which also relates ray to hyperplane.

Geometrically, for any ray $\mathcal{L}'$, $\boldsymbol{\Pi} = \boldsymbol{F}\mathcal{L}'$ is the corresponding epipolar hyperplane, as shown in Fig. 2. $\mathcal{L}$ also lies on the same epipolar hyperplane, which refers to $\mathcal{L}^\top \boldsymbol{F} \mathcal{L}' = 0$. This ensures that $\boldsymbol{F}$ can be estimated from ray-ray correspondences. Besides, Eq. (9) also proves that $\boldsymbol{F}$ is independent of scene structure, and can be determined uniquely from RSP matrices.

In summary, the definition and properties of ray-space fundamental matrix $\boldsymbol{F}$ are briefly summarized as follows,

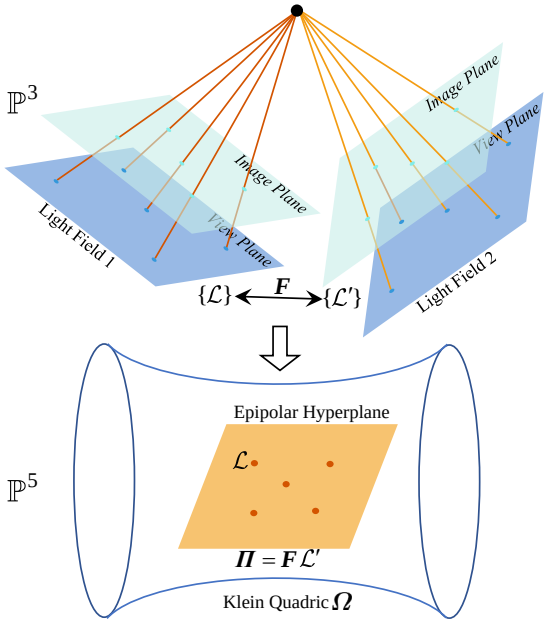


Fig. 2. Ray-space epipolar geometry. The ray bundles $\{\mathcal{L}\}$ and $\{\mathcal{L}'\}$ of point $\boldsymbol{X}$ in light field 1 and 2 are marked by red and orange respectively. They are also points on Klein quadric in $\mathbb{P}^5$. The epipolar hyperplane $\boldsymbol{\Pi} = \boldsymbol{F}\mathcal{L}'$ refers to a hyperplane on Klein quadric in $\mathbb{P}^5$. It is mapped from $\mathcal{L}'$ by $\boldsymbol{F}$. The corresponding ray $\mathcal{L}$ lies on $\boldsymbol{\Pi}$, that is $\mathcal{L}^\top \boldsymbol{F} \mathcal{L}' = 0$.

Definition 1. *The* ray-space fundamental matrix $\boldsymbol{F}$ *between two light fields is a unique* 6×6 *full rank homogeneous matrix which satisfies,*

$$\mathcal{L}^\top \underbrace{\boldsymbol{K}^\top \begin{bmatrix} \boldsymbol{O}_{3\times3} & \boldsymbol{R} \\ \boldsymbol{R} & \boldsymbol{E} \end{bmatrix} \boldsymbol{K}'}_{=:\boldsymbol{F}} \mathcal{L}' = 0, \tag{13}$$

for all ray-ray correspondences $\mathcal{L} \leftrightarrow \mathcal{L}'$.

1) $\boldsymbol{F}$ is a full rank matrix with 17 degrees of freedom.
2) $\boldsymbol{F}$ is estimated from ray-ray correspondences $\mathcal{L}^\top \boldsymbol{F} \mathcal{L}' = 0$.
3) $\boldsymbol{\Pi} = \boldsymbol{F}\mathcal{L}'$ is the epipolar hyperplane on Klein quadric corresponding to $\mathcal{L}'$ in $\mathbb{P}^5$. Similarly, $\boldsymbol{\Pi}' = \boldsymbol{F}^\top \mathcal{L}$ is the epipolar hyperplane corresponding to $\mathcal{L}$.
4) $\boldsymbol{F}$ is uniquely computed from RSP matrices $\boldsymbol{F} = \boldsymbol{P}^\top \boldsymbol{\Omega} \boldsymbol{P}'$.

Remarks. Eq. (13) could be thought of as the generalization of previous work [16] in which the assumption of calibrated LFC is removed. Specifically, Eq. (13) intrinsically constrains the ray-ray correspondences among light fields and encapsulates the ray-space epipolar geometry, whereas the generalized epipolar constraint [16] *only* defines the relationship of motion among normalized light fields (calibrated LFCs). Consequently, the ray-space fundamental matrix is a basic algebraic entity of multi-view light fields. The properties of the ray-space fundamental matrix also provide the theoretical basis for developing multi-view applications.

4.3 Special Cases of Fundamental Matrix

Certain special motions, or the constant intrinsic parameters, allow the ray-space fundamental matrix to be simplified. We will discuss two cases: *pure translation* and *pure rotation*. The 'pure' indicates that there is no change in the intrinsic parameters.

Pure translation. Suppose the motion of the LFC is a pure translation with no rotation ($\boldsymbol{R}=\boldsymbol{I}$) and no change in the intrinsic parameters ($\boldsymbol{K}=\boldsymbol{K}'$). According to Eq. (13) and *Corollary* 1, we formulate

$$\boldsymbol{F_t}=\begin{bmatrix} \boldsymbol{O}_{3\times3} & \boldsymbol{I} \\ \boldsymbol{I} & \frac{k_u}{k_i}\left[\boldsymbol{K}_{uv}^{-1}\boldsymbol{t}\right]_\times \end{bmatrix}, \tag{14}$$

where $\boldsymbol{F_t}$ has only 3 degrees of freedom. Moreover, a special property of $\boldsymbol{F_t}$ is deduced, namely, $\boldsymbol{F_t}^\top\boldsymbol{\Omega}\boldsymbol{F_t}=\boldsymbol{F_t}\boldsymbol{\Omega}\boldsymbol{F_t}^\top=\boldsymbol{\Omega}$.

Pure rotation. In this case, the motion is a pure rotation with no translation ($\boldsymbol{t}=0$) and constant intrinsic parameters ($\boldsymbol{K}=\boldsymbol{K}'$). According to Eq. (13) and *Corollary* 1, we simplify the fundamental matrix,

$$\boldsymbol{F_R}=\begin{bmatrix} \boldsymbol{O}_{3\times3} & \boldsymbol{K}_{uv}^{-1}\boldsymbol{R}\boldsymbol{K}_{uv} \\ \boldsymbol{K}_{uv}^{\top}\boldsymbol{R}\boldsymbol{K}_{uv}^{-\top} & \boldsymbol{O}_{3\times3} \end{bmatrix}, \tag{15}$$

where $\boldsymbol{F_R}$ has only 9 degrees of freedom. Similarly, orthogonality of $\boldsymbol{R}$ and *Corollary* 1 derive a special property of $\boldsymbol{F_R}$, that is $\boldsymbol{F_R}^\top\boldsymbol{\Omega}\boldsymbol{F_R}=\boldsymbol{F_R}\boldsymbol{\Omega}\boldsymbol{F_R}^\top=\boldsymbol{\Omega}$.

General motion. The pure translation and pure rotation give additional insight into the general motion. The general motion is decomposed into pure rotation and pure translation. According to Eqs. (7), (14) and (15), we have,

$$\boldsymbol{F}\sim\boldsymbol{F_t}\boldsymbol{\Omega}\boldsymbol{F_R}, \tag{16}$$

where $\sim$ refers to equality up to a scale. The analysis of such cases is important, firstly because special motions are frequently occurring in practice, secondly because the fundamental matrix has a simplified form for convenient computation.

5 Applications

We implement two different applications based on the proposed ray-space epipolar geometry: ray-space fundamental matrix estimation and LFC calibration.

5.1 Ray-Space Fundamental Matrix Estimation

The ray-space fundamental matrix is independent of scene structure. However, it can be estimated by ray-ray correspondences from same scene point alone, without requiring knowledge of intrinsic and extrinsic parameters.

5.1.1 Linear Initialization

Using Kronecker product operator $\otimes$, Eq. (11) can be simplified,

$$\left(\boldsymbol{p}^\top\otimes\boldsymbol{n}'^\top,\boldsymbol{p}^\top\otimes\boldsymbol{p}'^\top,\boldsymbol{n}^\top\otimes\boldsymbol{p}'^\top\right)\boldsymbol{f}=0, \tag{17}$$

where $\boldsymbol{f}=\left(\vec{\boldsymbol{F}}_{21},\vec{\boldsymbol{F}}_{22},\vec{\boldsymbol{F}}_{12}\right)^\top$ refers to a 27-vector. $\vec{\boldsymbol{F}}_{ij}$ is a 9-vector made up of the entries of block matrix $\boldsymbol{F}_{ij}$ in row-major order.

Given a set of $n\times m$ ray-ray correspondences with Plücker parameterization,

$$\{\mathcal{L}_i\}_{i=1,\dots,n}\longleftrightarrow\{\mathcal{L}'_j\}_{j=1,\dots,m}, \tag{18}$$

where $\mathcal{L}_i$ and $\mathcal{L}'_j$ are from the same scene point but are recorded in different light fields. Eq. (17) is stacked as a homogeneous set of linear equations $\boldsymbol{A}\boldsymbol{f}=0$. Hence, $\boldsymbol{f}$ can only be linearly solved with a scale, only if there are at least 26 ray-ray correspondences. The solution is the generator of the right null-space of $\boldsymbol{A}$.

However, $\boldsymbol{F}$ computed by Eq. (17) is not satisfied with two important constraints of the ray-space fundamental matrix, that is, orthogonal and singular constraints. Taken singularity and orthogonality into consideration, the most useful method to correct $\boldsymbol{F}$ is the singular value decomposition (SVD) [24]. We then take two independent stages to enforce these constraints.

First, $\boldsymbol{F}_{21}$ and $\boldsymbol{F}_{12}$ generate the ray-space fundamental matrix $\boldsymbol{F}$ if the translation $\boldsymbol{t}$ equals to zero according to Eq. (10). The singular values of $\boldsymbol{F}_{21}$ and $\boldsymbol{F}_{12}$ are same. Specifically, consider $\boldsymbol{U}_{12}\boldsymbol{D}_{12}\boldsymbol{V}_{12}^\top$ and $\boldsymbol{U}_{21}\boldsymbol{D}_{21}\boldsymbol{V}_{21}^\top$ denoting the SVD factorization of $\boldsymbol{F}_{12}$ and $\boldsymbol{F}_{21}$ respectively, where the main diagonal elements of $\boldsymbol{D}_{21}$ and $\boldsymbol{D}_{12}$ represent the singular values of $\boldsymbol{F}_{12}$ and $\boldsymbol{F}_{21}$. Consequently, we use the average of singular values to refine $\boldsymbol{F}_{21}$ and $\boldsymbol{F}_{12}$ to satisfy orthogonal constraint.

Second, $\boldsymbol{F}_{22}$ is the fundamental matrix between central sub-apertures based on Eq. (12). Similarly, we decompose $\boldsymbol{F}_{22}$ into $\boldsymbol{U}_{22}\boldsymbol{D}_{22}\boldsymbol{V}_{22}^\top$ by SVD factorization. Hence, we use zero to replace the minimal singular value in $\boldsymbol{D}_{22}$ so that $\boldsymbol{F}_{22}$ is corrected to enforce singular constraint.

In order to accurately and robustly estimate $\boldsymbol{F}$, it is essential to implement a proper normalization algorithm specifically designed for ray-space. Similarly, according to Eq. (4), we use translations and scale factors to normalize rays so that the centroidal axis of the reference rays is at the principal axis and the RMS geometric distance of these rays from the principal axis is equal to $\sqrt{2}$.

Overall, the algorithm just described is the essence of a method called the normalized 26-ray algorithm for the initialization of $\boldsymbol{F}$, as shown in Alg. 1.

5.1.2 Non-Linear Optimization

The initial solution is then refined via nonlinear optimization. Similar to symmetric epipolar error of traditional fundamental matrix, we minimize the geometrically more meaningful symmetric epipolar error,

$$\sum^{\#point}\sum_{i}^{\#rayL}\sum_{j}^{\#rayR}\left|\mathcal{L}_i^\top\boldsymbol{F}\mathcal{L}'_j\right|\left(\frac{1}{\left\|\boldsymbol{F}\mathcal{L}'_j\right\|}+\frac{1}{\left\|\boldsymbol{F}^\top\mathcal{L}_i\right\|}\right), \tag{19}$$

where $\|\cdot\|$ denotes L_2 norm, $\{\mathcal{L}_i\}_{\#rayL}\leftrightarrow\{\mathcal{L}'_j\}_{\#rayR}$ is the ray-ray correspondences of each point. Compared with algebraic distance Eq. (13), Eq. (19) minimizes the distance of a ray from its projected epipolar hyperplane with clear geometry definition. According to Eqs. (12) and (19), the traditional symmetric epipolar distance between two images is a special case of the proposed symmetric epipolar distance. More importantly, once the fundamental matrix is estimated by the normalized 26-ray algorithm, Eq. (19) can be the standard for outliers detection within a RANSAC framework. In order to minimize the above nonlinear function Eq. (19), we utilize Levenberg-Marquardt algorithm based on the trust region reflective method [40] with the help of MATLAB's lsqnonlin function. The proposed fundamental matrix estimation algorithm within a RANSAC framework is illustrated in Alg. 1.

Algorithm 1 Ray-Space Fundamental Matrix Estimation.

Input: Ray-ray correspondences $\{\mathcal{L}\} \leftrightarrow \{\mathcal{L}'\}$.
Output: Ray-space fundamental matrix $\boldsymbol{F}$.

while $n_c \leq \#Count$ **do**
  $(\widetilde{\mathcal{L}}, \widetilde{\mathcal{L}}') = \mathit{NormalizeRays}(\mathcal{L}, \mathcal{L}', \boldsymbol{T}, \boldsymbol{T}')$ ▷ Eqs. (4),(27)
  $\widetilde{\boldsymbol{F}} = \mathit{EstimateFundamentalMatrix}(\widetilde{\mathcal{L}}, \widetilde{\mathcal{L}}')$ ▷ Eq. (17)
  $\mathit{EnforceOrthogonalConstraint}(\widetilde{\boldsymbol{F}})$
  $\mathit{EnforceSingularConstraint}(\widetilde{\boldsymbol{F}})$
  $\boldsymbol{F} = \mathit{DenormalizeFundamentalMatrix}(\widetilde{\boldsymbol{F}}, \boldsymbol{T}, \boldsymbol{T}')$
  $d = \mathit{ComputeSymEpipolarError}(\boldsymbol{F}, \mathcal{L}, \mathcal{L}')$ ▷ Eq. (19)
  $\mathit{DetectInliers}(d \leqslant t)$
end while
$\mathit{Optimization}(\boldsymbol{F})$ ▷ Eq. (19)

5.2 Light Field Camera Calibration

To further verify the effectiveness of ray-space epipolar geometry, we propose an LFC calibration algorithm based on the point-ray constraint established by the ray-space fundamental matrix.

5.2.1 Point-Ray Constraint

In 3D projective geometry, a point $\boldsymbol{X}_w$ in the world coordinate frame can be described as the intersection of $\mathcal{R}_w = (\boldsymbol{m}_w^\top, \boldsymbol{q}_w^\top)^\top$ with the plane $Z = Z_w$. $\mathcal{R}_w$ is captured by an LFC $\boldsymbol{P}$. Given two rays intersecting at $\boldsymbol{X}_w$ and paralleling to X_w-axis and the Y_w-axis respectively. Therefore, according to the ray-space epipolar geometry, we establish a constraint between $\boldsymbol{X}_w$ and the corresponding ray $\mathcal{L}$ captured by an LFC $\boldsymbol{P}$,

$$\underbrace{\begin{bmatrix} 0 & Z_w & -Y_w & 1 & 0 & 0 \\ -Z_w & 0 & X_w & 0 & 1 & 0 \end{bmatrix}}_{=:\boldsymbol{M}(\boldsymbol{X}_w)} \boldsymbol{\Omega}\boldsymbol{P} \begin{bmatrix} \boldsymbol{n} \\ \boldsymbol{p} \end{bmatrix} = 0, \tag{20}$$

where $\boldsymbol{M}$ is a 2×6 measurement matrix, representing the ray bundle of $\boldsymbol{X}_w$ on the plane $Z = Z_w$. That means Eq. (20) is also the ray-space fundamental matrix which is determined by a pair of RSP matrices $(\boldsymbol{I}, \boldsymbol{P})$. Consequently, Eq. (20) can be used to linearly estimate RSP matrix $\boldsymbol{P}$ of an LFC.

Remarks. The point-ray constraint is first provided by Johannsen *et al.* [3] for pose estimation with calibrated LFCs. Compared with them, Eq. (20) introduces the proposed RSIM to intrinsically describe point-ray constraints of light fields. Meanwhile, points of $\boldsymbol{M}$ in Eq. (20) are accurate checkerboard corners in the world coordinate frame for calibration. Inversely, reconstructed points in [3] for pose estimation are sensitive to small noises, due to the LFC ultra-small baseline. LFC pose estimation via point-ray constraint turns out to be unstable. Consequently, point-ray constraint is suitable to linearly calibrate an LFC instead of pose estimation.

5.2.2 Linear Initialization

Without loss of generality, there is an assumption that the checkerboard is on the plane $Z_w = 0$ in the world coordinate frame, which leads to a simplified form of Eq. (20),

$$\begin{bmatrix} 1 & 0 & -Y_w \\ 0 & 1 & X_w \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{n}^\top & \boldsymbol{p}^\top \end{bmatrix} \vec{\boldsymbol{H}}_s = 0, \tag{21}$$

where $\vec{\boldsymbol{H}}_s$ is an 18×1 matrix stretched on row from the simplified homogeneous ray-space fundamental matrix $\boldsymbol{H}_s$. Subsequently, $\boldsymbol{H}_s$ denotes a 3×6 matrix only using intrinsic and extrinsic parameters,

$$\boldsymbol{H}_s = \begin{bmatrix} \boldsymbol{r}_1^\top & -\boldsymbol{r}_1^\top[\boldsymbol{t}]_\times \\ \boldsymbol{r}_2^\top & -\boldsymbol{r}_2^\top[\boldsymbol{t}]_\times \\ \boldsymbol{O}_{1\times3} & \boldsymbol{r}_3^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{K}_{ij} & \boldsymbol{O}_{3\times3} \\ \boldsymbol{O}_{3\times3} & \boldsymbol{K}_{uv} \end{bmatrix}, \tag{22}$$

where $\boldsymbol{r}_i$ is the i-th column vector of rotation matrix $\boldsymbol{R}$.

In order to derive intrinsic parameters, we abbreviate with $[\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{0}_{1\times3}]^\top$ the first three and with $[\boldsymbol{h}_3, \boldsymbol{h}_4, \boldsymbol{h}_5]^\top$ the second three columns of $\boldsymbol{h}_s$ respectively. $\boldsymbol{h}_i$ denotes the row vector (h_{i1}, h_{i2}, h_{i3}). Utilizing the orthogonality of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, we have

$$\begin{aligned} &\boldsymbol{h}_1 \boldsymbol{K}_{ij}^{-1} \boldsymbol{K}_{ij}^{-\top} \boldsymbol{h}_2^\top = 0 \\ &\boldsymbol{h}_1 \boldsymbol{K}_{ij}^{-1} \boldsymbol{K}_{ij}^{-\top} \boldsymbol{h}_1^\top = \boldsymbol{h}_2 \boldsymbol{K}_{ij}^{-1} \boldsymbol{K}_{ij}^{-\top} \boldsymbol{h}_2^\top . \end{aligned} \tag{23}$$

It is noting that $\boldsymbol{K}_{ij}^{-1}\boldsymbol{K}_{ij}^{-\top}$ is a symmetric matrix which contains only 5 distinct non-zero elements. Consequently, Eq. (23) is rewritten as two homogeneous equations. It can be solved only if there are at least four such equations (from two positions). Once $\boldsymbol{K}_{ij}^{-1}\boldsymbol{K}_{ij}^{-\top}$ is obtained up to a scale, $\hat{\boldsymbol{K}}_{uv}$ is then linearly computed based on *Corollary* 1 and Cholesky factorization [41]. Furthermore, the rest intrinsic parameters and extrinsic parameters of different poses can be obtained as follows,

$$\begin{aligned} &\lambda = \frac{1}{2}\left(\left\|\hat{\boldsymbol{K}}_{uv}\boldsymbol{h}_1^\top\right\| + \left\|\hat{\boldsymbol{K}}_{uv}\boldsymbol{h}_2^\top\right\|\right), \\ &\tau = 1/\left\|\hat{\boldsymbol{K}}_{uv}^{-\top}\boldsymbol{h}_5^\top\right\|, \\ &\boldsymbol{r}_1 = \frac{\alpha}{\lambda}\hat{\boldsymbol{K}}_{uv}\boldsymbol{h}_1^\top,\ \boldsymbol{r}_2 = \frac{\alpha}{\lambda}\hat{\boldsymbol{K}}_{uv}\boldsymbol{h}_2^\top,\ \boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2, \\ &\boldsymbol{t} = (\boldsymbol{G}^\top\boldsymbol{G})^{-1}(\boldsymbol{G}^\top\boldsymbol{g}), \\ &\boldsymbol{G} = (-[\boldsymbol{r}_1]_\times, -[\boldsymbol{r}_2]_\times)^\top,\ \boldsymbol{g} = (\tau\hat{\boldsymbol{K}}_{uv}\boldsymbol{h}_3^\top, \tau\hat{\boldsymbol{K}}_{uv}\boldsymbol{h}_4^\top)^\top, \\ &\boldsymbol{K}_{ij} = \lambda\tau\hat{\boldsymbol{K}}_{uv}^{-\top}, \end{aligned} \tag{24}$$

where $\|\cdot\|$ denotes L_2 norm, α denotes a sign function which is determined by t_z because it must be positive (*i.e.* the checkerboard is put in front of the LFC).

5.2.3 Non-Linear Optimization

Similar to [15], only radial distortion is considered. The undistorted coordinate $(\widetilde{x}, \widetilde{y})^\top$ is rectified by the distorted coordinate $(x, y)^\top$ under the view $(s, t)^\top$,

$$\begin{cases} \widetilde{x} = x + (k_1 r_{xy}^2 + k_2 r_{xy}^4)(x - b_1) + k_3 s \\ \widetilde{y} = y + (k_1 r_{xy}^2 + k_2 r_{xy}^4)(y - b_2) + k_4 t \end{cases}, \tag{25}$$

where $r_{xy}^2 = (x - b_1)^2 + (y - b_2)^2$.

The initial solution computed by the linear method is refined via nonlinear optimization. Instead of minimizing the distance between checkerboard corners and rays [17] and the re-projection error in traditional multi-view geometry [24], we define a ray-ray cost function to acquire the nonlinear solution,

$$\sum_{p=1}^{\#pose} \sum_{n=1}^{\#point} \sum_{i=1}^{\#view} \left\| d\left(\widetilde{\mathcal{R}}'_{w,i}(\mathcal{P}, \boldsymbol{k}^d, \boldsymbol{R}_p, \boldsymbol{t}_p), \mathcal{R}_w(\boldsymbol{X}_{w,n})\right)\right\|, \tag{26}$$

where $\widetilde{\mathcal{R}}'_{w,i}$ is the projected ray from $\mathcal{L}_i$ based on Eq. (6), followed by the distortion model Eq. (25). $\mathcal{R}_w$ denotes the rays of $\boldsymbol{X}_w$ on the checkerboard as shown in Eq. (20). $\mathcal{P}$ represents intrinsic parameters, $\boldsymbol{k}^d$ is distortion vector and $\boldsymbol{R}_p$, $\boldsymbol{t}_p$ are extrinsic parameter at each position, $1 \leq p \leq P$.

Moreover, the ray-ray distance $d(\mathcal{R}, \mathcal{R}')$ is geometrically described as the point-point distance on Klien quadric according to ray-space epipolar geometry,

$$d(\mathcal{R}, \mathcal{R}') = \frac{\left|\mathcal{R}^\top \boldsymbol{\Omega} \mathcal{R}'\right|}{\|\boldsymbol{q} \times \boldsymbol{q}'\|} = \frac{\left|\boldsymbol{m}^\top \boldsymbol{q}' + \boldsymbol{q}^\top \boldsymbol{m}'\right|}{\|\boldsymbol{q} \times \boldsymbol{q}'\|}. \quad (27)$$

Eq. (26) is a nonlinear objective function which can be solved using Levenberg-Marquardt algorithm based on the trust region reflective method [40]. In addition, $\boldsymbol{R}$ is parameterized by Rodrigues formula [42]. MATLAB's lsqnonlin function is utilized to implement the optimization. The LFC calibration algorithm is summarized in Alg. 2.

Algorithm 2 Light Field Camera Calibration.

Input: Checkerboard corners $\boldsymbol{X}_w$,
Corresponding rays $\mathcal{L}$.
Output: Intrinsic parameter $\mathcal{P} = (k_i, k_j, k_u, k_v, u_0, v_0)$,
Distortion vector $\boldsymbol{k}^d = (k_1, k_2, k_3, k_4, b_1, b_2)$,
Extrinsic parameters $\boldsymbol{R}_p, \boldsymbol{t}_p$, $(1 \leqslant p \leqslant P)$.

1: **for** $p = 1$ to P **do**
2: $\quad \boldsymbol{P}_s = \mathit{EstimateProjectionMatrix}(\boldsymbol{X}_w, \mathcal{L})$ ▷ Eq. (21)
3: **end for**
4: $\boldsymbol{B} = \mathit{EstimateMatrix}(\boldsymbol{P}_s)$ ▷ Eq. (23)
5: $(k_u, k_v, u_0, v_0) = \mathit{CalculateK}_{uv}(\boldsymbol{B})$
6: **for** $p = 1$ to P **do**
7: $\quad (\boldsymbol{R}_p, \boldsymbol{t}_p) = \mathit{CalculateRT}(\boldsymbol{H}_s, k_u, k_v, u_0, v_0)$ ▷ Eq. (24)
8: **end for**
9: $(k_i, k_j) = \mathit{CalculateK}_{ij}(\hat{\boldsymbol{K}}_{ij})$ ▷ Eq. (24)
10: $\mathit{Optimization}(\mathcal{P}, \boldsymbol{k}^d, \bigcup_{p=1}^{P}(\boldsymbol{R}_p, \boldsymbol{t}_p))$ ▷ Eq. (26)

6 Experiments

We evaluate the performance of ray-space fundamental matrix estimation and light field camera calibration on both simulated and real light fields.

6.1 Experiments on Fundamental Matrix Estimation

6.1.1 Simulated Data

In order to evaluate the performance of the proposed fundamental matrix estimation method, two experiments have been undertaken. We simulate a realistic LFC close to Lytro Illum, whose intrinsic parameters are listed as $k_i = k_j = 3.6e^{-4}$, $k_u = k_v = 2.0e^{-3}$, $u_0 = 0.54$, and $v_0 = 0.36$. The rotation angles between a light field pair are randomly generated from $-30°$ to $30°$, while the translation is randomly chosen in the box $[0, 0.2]^3$. The depth of scene points ranges from $0.2m$ to $0.8m$.

Performance w.r.t. the noise level. In the first experiment, we generate a pair of light fields with Gaussian noise to examine the noise resilience on the proposed algorithm. We add Gaussian noise varying from 0.1 to 1.5 pixels with a 0.1 pixels step to light fields. For each noise level, we carry out 150 independent trials with different combinations, including input scenarios, linear methods and motion types.

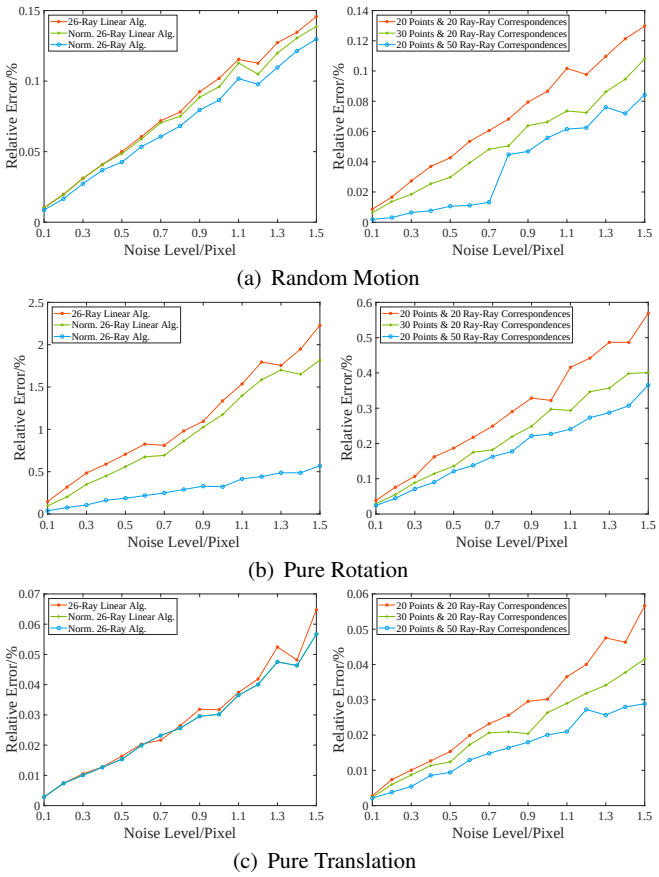


Fig. 3. Relative errors of ray-space fundamental matrix on simulated data with different levels of noise.

Fig. 3 illustrates the mean relative errors of $\boldsymbol{F}$ which are computed by the Frobenius norm on three camera motions. It verifies the noise resilience of our algorithm that errors increase almost linearly with the levels of noise.

The left column of Fig. 3 summaries the relative errors with three different initialization methods on the input scenario of 20 points and 20 ray-ray correspondences. It can be seen that the introduction of normalization leads the smaller relative errors, which verifies the effectiveness of the proposed ray normalization. The orthogonal and singular constraints also improve the accuracy of the initial solution. Especially, the errors reduce obviously on the pure rotation due to the enforcement of orthogonal constraint. Yet, the errors on pure translation are almost constant with the singular correction. It also indicates that the orthogonal constraint effectively improves the performance of our algorithm compared with singular constraint.

Meanwhile, the right column of Fig. 3 shows the relative errors of normalized 26-ray algorithm with different input scenarios. When the noise is fixed, the errors decline with the number of correspondences (*i.e.* points×ray-ray correspondences), which exhibits the numerical stability of our algorithm.

In addition, generalized epipolar geometry is a special case of the proposed model to calibrated cameras. To further investigate the noise resilience of the proposed algorithm, another experiment is conducted on relative pose estimation compared with state-of-the-art method proposed by

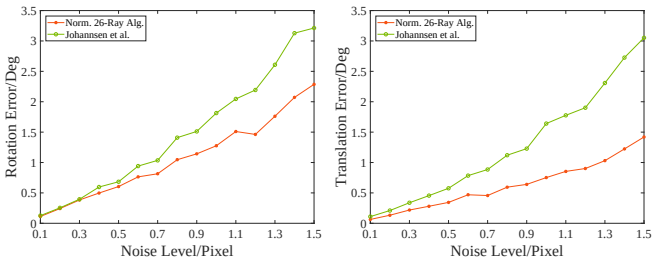


Fig. 4. Comparisons of relative pose estimation using ray-space fundamental matrix and state-of-the-art method by Johannsen *et al.* [3].

Johannsen *et al.* [3]. Similarly, a pair of light fields with Gaussian noise (varying from 0.1 to 1.5 pixels) is generated with random motion. To illustrate the influence of rays for point reconstruction, 150 independent trials with the input scenario of 30 points and 20 ray-ray correspondences are performed. The normalized 26-ray algorithm is first used to estimate ray-space fundamental matrix. The relative pose is then decomposed from $\boldsymbol{F}$ with the help of ground-truth intrinsic parameters according to Eq. (13). Meanwhile, the calibrated ray-ray correspondences is computed according to Eq. (2) for the relative pose estimation of [3]. Error metrics are angular differences from the ground truth in degrees for the estimated rotation and translation. Fig. 4 shows the mean rotation and translation errors of the proposed method and state-of-the-art [3] respectively.

Note that all algorithms are initialization methods. The proposed method provides better results with fewer errors, but state-of-the-art [3] is difficult to obtain. On one hand, as discussed in Sec. 5.2.1, the point-ray constraint is used to estimate LFC pose in [3]. However, considering the small baseline of an LFC, 3D point reconstruction in a light field is very sensitive to noise. Thus, [3] results in larger errors, especially translation errors, with the increasing noise levels. It is evident that point-ray constraint is suitable for calibration with the introduction of RSIM, that is Eq. (20). On the other hand, ray-space epipolar geometry intrinsically constrains ray-ray correspondences of light fields instead of calibrated point-ray constraints. It verifies the robustness of the proposed method on relative pose estimation.

6.1.2 Real Scenes

To further substantiate the proposed fundamental matrix estimation method, experiments on real scene light fields are performed. We utilize 4 datasets of multiple light fields captured by Lytro Illum, including public datasets "Flowers" and "Trees" [43], and self-captured datasets "Toys" and "Books". Each of the public datasets contains different scenes captured by different cameras, with between 5 to 7 light fields of each scene. In contrast, we also utilize a Lytro Illum to collect two datasets, each of which includes 17 and 30 multi-view light fields respectively. The datasets include indoor and outdoor complex environments. The set of "Books" is also a special case of pure translation.

Besides, since there is no calibration data provided by the public datasets, we directly use Lytro Power Tools [1] instead of the proposed calibration method to rectify the raw data to an undistorted light field with $11\times11\times541\times376$ samples. For the ray feature extraction, we extract sparse images point from every sub-aperture image within LF by DoG (Difference of Gaussian) and match the central sub-aperture image with other sub-aperture images by SIFT [44]. Taken the regular and planar arrangement of sub-aperture images into consideration, the matched features are filtered according to the invariant depth. We then generate a ray bundle within the light field. We obtain the ray-ray correspondences between two light fields through matching center views of each light field.

After extracting ray-ray correspondences of light fields, we estimate their ray-space fundamental matrix. In a sense, the proposed method is, to our knowledge, the first attempt to generalize and estimate a ray-space fundamental matrix between two light fields. Consequently, the proposed algorithm is quantitatively compared with 8-point, 7-point and 5-point algorithms through treating the LFC as an array of pinhole cameras. The 8-point algorithm [24] is provided by MATLAB's estimateFundamentalMatrix function. The 7-point algorithm [24] is coded by ourselves. The 5-point algorithm [30] is run by their latest released code.

Given that the considerable number of light fields in each dataset, we randomly choose two light fields and estimate the ray-space fundamental matrix for 80 instances. Tab. 1 summarizes the mean and maximum of the root mean square (RMS) symmetric epipolar errors for 80 instances. As mentioned in Sec. 5.1, the traditional symmetric epipolar distance is equivalent to the proposed symmetric epipolar distance of central views. We hence compare with traditional baseline methods applied to ray-ray correspondences

TABLE 1
Mean and Maximum RMS Symmetric Epipolar Errors with 80 Random Pairs of Light Fields (Unit: $pixels$).

		Trees (399)	Flowers (719)	Toys (17)	Books (30)
8-point^C	mean	0.3011	0.4250	**0.1526**	**0.1059**
	max	1.1729	8.0772	0.2442	**0.1574**
7-point^C	mean	0.2898	0.2808	0.1663	0.1181
	max	0.4692	0.5107	0.2161	0.1811
5-point^C	mean	2.6567	2.1223	1.7706	0.4803
	max	13.2227	17.3187	4.7503	0.8772
Ours^C	mean	**0.2791**	**0.2741**	0.1567	0.1808
	max	**0.4037**	**0.4729**	**0.1906**	0.2275
Ours	mean	0.5739	0.3197	0.2450	0.3848
	max	0.9729	0.5332	0.2805	0.4347
Inlier (%)	mean	57.45	57.33	59.46	61.70

The superscript C indicates the errors between central views.
The N in the parentheses indicates the number of light fields.

TABLE 2
RMS Symmetric Epipolar Errors between Sub-Aperture Images of Dataset "Toys" with 80 iterations (Unit: *pixels*).

1st LF	2nd LF								
	(-3, -3)	(0, -3)	(3, -3)	(-3, 0)	(0, 0)	(3, 0)	(-3, 3)	(0, 3)	(3, 3)
(-3, -3)	0.1707	0.2077	0.2303	0.1718	0.2246	0.2087	0.1822	0.1844	0.2216
(0, -3)	0.1691	0.1986	0.1781	0.1748	0.2068	0.2023	0.1844	0.2110	0.2294
(3, -3)	0.2110	0.1821	0.1754	0.2065	0.2265	0.2095	0.1833	0.2511	0.2231
(-3, 0)	0.1632	0.1917	0.2178	0.1536	0.1739	0.2076	0.1658	0.1530	0.1664
(0, 0)	0.1969	0.2021	0.2474	0.1738	0.1567	0.1979	0.1868	0.1804	0.1812
(3, 0)	0.1790	0.1828	0.1745	0.1869	0.1969	0.1917	0.1855	0.1827	0.1673
(-3, 3)	0.1966	0.2188	0.2348	0.1618	0.1985	0.2231	0.1468	0.1600	0.1649
(0, 3)	0.1962	0.2176	0.2487	0.1561	0.1665	0.2093	0.1695	0.1708	0.1615
(3, 3)	0.1912	0.2141	0.2142	0.1743	0.1786	0.1874	0.1583	0.1532	0.1558

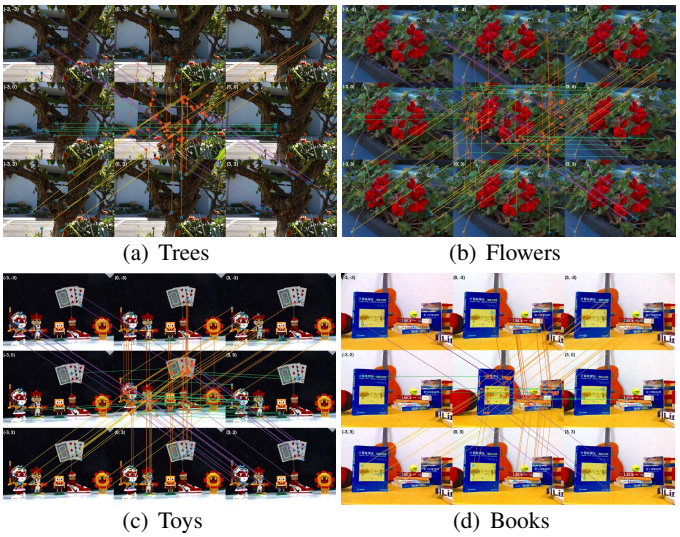


Fig. 5. The results of the proposed meothod combined with a RANSAC framework. Each pair of light field contains 50 random inliers of ray-ray correspondences.

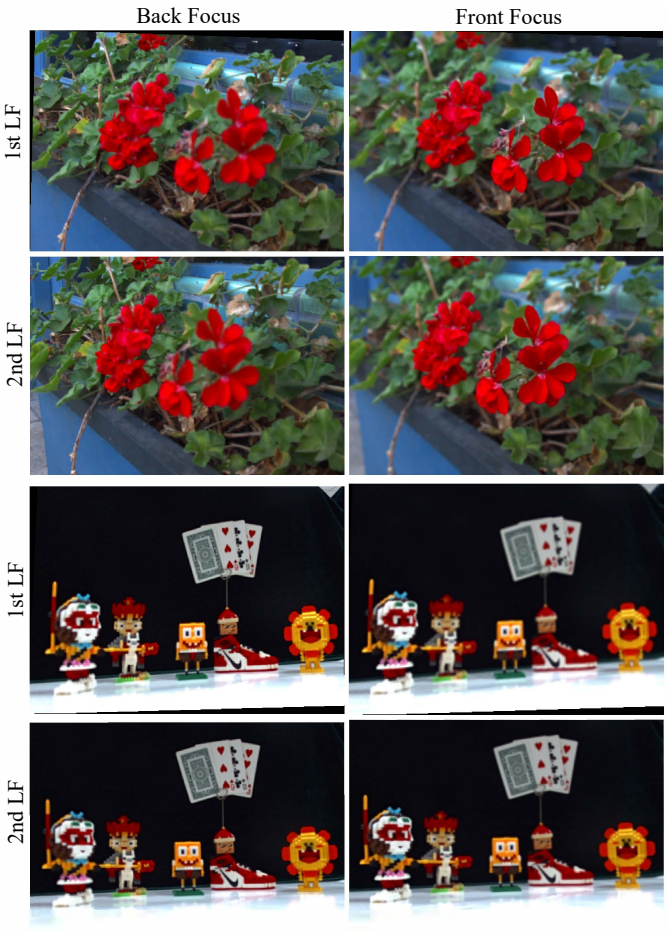


Fig. 6. The refocus results of projective rectified light field.

of the central sub-aperture images, as shown in Tab. 1. For the mean errors of central views, the proposed method provides a similar even smaller results compared with those of baseline methods except on dataset "Books". Those results show the effectiveness of the proposed method. The result on dataset "Books" performs worse because of the special motion. As mentioned in Sec. 4.3, the degrees of freedom of $\boldsymbol{F}$ reduces to 3 for pure translation motion. The high dimensional ray features with noise will make the solution complex compared with traditional baseline methods. Since the proposed method uses the only central sub-aperture images for internal ray feature extraction, taken errors introduced by the feature extraction into consideration, the results are acceptable. Moreover, the performance of the maximum errors demonstrates the robustness of the proposed method.

Tab. 1 also illustrates the errors of the proposed method applied to light fields. The errors of the proposed method are slightly higher compared with baseline methods applied to central views. This discrepancy can also be observed in [14], [15] and relates to astigmatism and field curvature that affect micro-lens based LFCs. Once we compute the ray-space fundamental matrix, we measure the root mean square (RMS) symmetric epipolar errors between arbitrary sub-aperture images on dataset "Toys", as shown in Tab. 2. We can observe the distribution of error is homogeneous and similar to the distribution of field curvature.

As mentioned in Sec. 5.1, the symmetric epipolar distance can be used to discard the outliers based on a RANSAC framework, as shown in Tab. 1. Fig. 5 shows light field pairs from each dataset with ray-ray correspondences of 50 random inliers. Fig. 5 illustrates the ray-ray correspondence on the central view of the first light field and arbitrary views of the second light field. It can be seen that the results seem good: the ray features within the light field maintain the depth invariance and the ray-ray correspondences lie on the same position in the first (central) and second (surround) light fields. As shown in Fig. 5, the reason why the performance of the proposed method is worse on dataset "Books" is too much co-planar points.

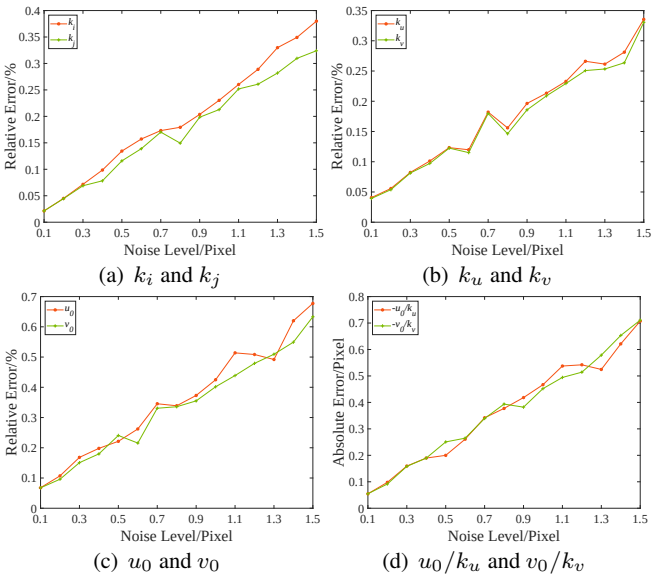
In order to further verify the performance of the proposed fundamental matrix, we propose another application of the fundamental matrix. Suppose the RSP matrix of the second LFC equals an identity matrix, we recovery the RSP matrix of the first LFC according to *Corollary* 5 and the estimated fundamental matrix. Then multi-view light fields may be resampled and rectified in the same coordinate frame up to a projective ambiguity. Fig. 6 shows the refocus results of the projective rectified light field. All results have verified the effectiveness and robustness of the proposed fundamental matrix.

6.2 Experiments on Light Field Camera Calibration

6.2.1 Simulated Data

In order to evaluate the performance of the proposed LFC calibration algorithm, we also simulate an LFC, whose intrinsic parameters are summarised as $k_i = 2.4e^{-4}$, $k_j = 2.5e^{-4}$, $k_u = 2.0e^{-3}$, $k_v = 1.9e^{-3}$, $u_0 = -3.2$ and $v_0 = -0.32$, similar to [15]. The checkerboard is a pattern with a 12×12 grid of $3.51mm$ cells.

Performance w.r.t. the noise level. In this experiment, we employ the measurements of 3 poses and 7×7 views to verify the robustness of calibration algorithm. The rotation



(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

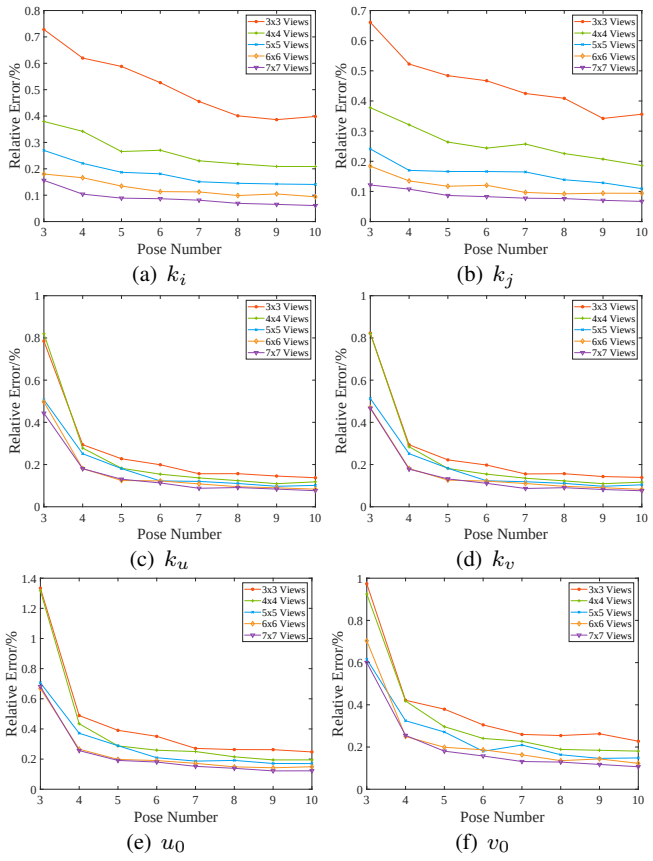
Fig. 7. Performance evaluation of intrinsic parameters on the simulated data with different levels of noise.

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 pixel step. For each noise level, we perform 150 independent trials. The estimated intrinsic parameters are evaluated by the average of relative errors with ground truth. As shown in Fig. 7, the errors almost linearly increase with noise level. For $\sigma = 0.5$ pixels which is larger than normal noise in practical calibration, the relative errors of intrinsic parameters and absolute errors of principle points are less than 0.25% and 0.24 pixel respectively, which demonstrates the robustness of the proposed method to high noise level.

Performance w.r.t. the number of poses and views. This experiment investigates the performance with respect to the number of poses and views. We vary the number of poses from 3 to 10 and the number of views from 3×3 to 7×7. For each combination of pose and view, by adding a Gaussian noise with zero mean and a standard deviation of 0.5 pixel, 200 trails with independent checkerboard poses are conducted. The rotation angles are randomly generated from $-30°$ to $30°$. The average relative errors of calibration results with increasing measurements are shown in Fig. 8. The relative errors decrease with the number of poses. Meanwhile, when the number of poses is fixed, the errors reduce with the number of views. In particular, when $\#pose \geq 4$ and $\#view \geq 4 \times 4$, all relative errors are less than 0.5%, which further exhibits the effectiveness of the proposed calibration method.

6.2.2 Real Data

To further substantiate the proposed light field calibration method, we compare the proposed method in ray re-projection error and re-projection error with state-of-the-arts on real light fields, including DPW by Dansereau *et al.* [17], BJW by Bok *et al.* [14] and MPC by Zhang *et al.* [15]. The datasets include light field datasets (Lytro) released by DPW



(a) k_i (b) k_j (c) k_u (d) k_v (e) u_0 (f) v_0

Fig. 8. Relative errors of intrinsic parameters on simulated data with different numbers of poses and views.

TABLE 3
RMS Ray Re-Projection Errors (Unit: mm).

	A	B	C	D	E
DPW [17]	0.0835	0.0628	**0.1060**	0.1050	0.3630
MPC [15]	0.0810	0.0572	0.1123	0.1046	0.5390
Ours	**0.0705**	**0.0438**	0.1199	**0.0740**	**0.2907**

and light field datasets[1] (Lytro and Illum) released by MPC.

The sub-aperture images are easy to be decoded by raw data. We improve the preprocessing of raw data described in [17] to obtain sub-aperture images. The preprocess of raw data begins with demosaicing after alignment of micro-lens array. Then, the vignetting raw data is refined in accordance with white image. Finally, normalized cross-correlation (NCC) of the white images is used to locate the centers of micro-lens images and estimate the average size of micro-lens images. It can be utilized for resampling and sub-aperture image extraction.

We firstly conduct calibration on the datasets collected with DPW [17]. For a fair comparison, the middle 7×7 sub-apertures are utilized. Tab. 3 summarizes the root mean square (RMS) ray re-projection error. Compared with DPW which employs 12 intrinsic parameters, the proposed ray-space projection model only employs a half of parameters but achieves smaller ray re-projection error except on dataset C. Given that the errors exhibited in DPW are

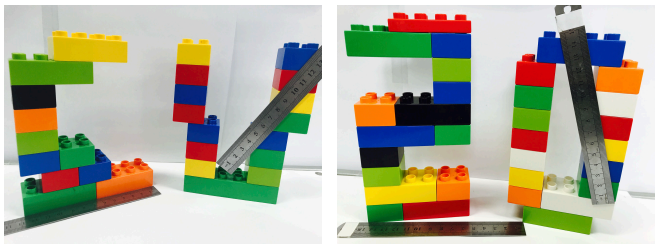
1. http://www.npu-cvpg.org/opensource

TABLE 4
Mean Re-Projection Errors (Unit: $pixel$).

	A	B	C	D	E
DPW [17]	0.2284	0.1582	0.1948	0.1674	0.3360
BJW [14]	0.3736	0.2589	-	-	0.2742
MPC [15]	0.2200	0.1568	0.1752	0.1475	0.2731
Ours	**0.1843**	**0.1245**	**0.1678**	**0.1069**	**0.1383**

TABLE 5
RMS Ray Re-Projection Errors of Optimizations without and with Distortion Rectification (Unit: mm).

		Illum-1	Illum-2	Lytro-1	Lytro-2
Optimized without Rectification	DPW [17]	0.5909	0.4866	0.1711	0.1287
	BJW [14]	-	-	-	-
	MPC [15]	0.5654	0.4139	0.1703	0.1316
	Ours	**0.5641**	**0.4132**	**0.1572**	**0.1237**
Optimized with Rectification	DPW [17]	0.2461	0.2497	0.1459	0.1228
	BJW [14]	0.3966	0.3199	0.4411	0.2673
	MPC [15]	0.1404	0.0936	0.1400	0.1124
	Ours	**0.1294**	**0.0837**	**0.1142**	**0.0980**



(a) 'C': 128.0mm, 'V': 97.5mm (b) '2': 147.5mm, '0': 102.0mm

Fig. 10. Measurements between specific points by rulers.

TABLE 6
Quantitative Comparison of Different Calibration Methods (Unit: mm).

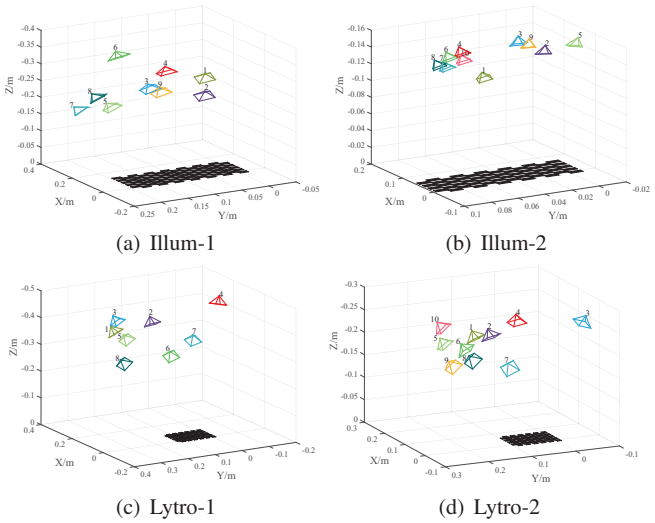
	'C'	'V'	'2'	'0'
Ruler	128.0	97.5	147.5	102.0
DPW [17]	124.3 (2.89%)	100.2 (2.77%)	145.7 (1.22%)	106.6 (4.51%)
BJW [14]	120.6 (5.78%)	106.8 (9.54%)	151.9 (2.98%)	103.4 (**1.37**%)
MPC [15]	127.5 (**0.39**%)	97.0 (**0.51**%)	144.9 (1.76%)	103.6 (1.57%)
Ours	127.4 (0.47%)	97.0 (**0.51**%)	145.9 (**1.08**%)	103.6 (1.57%)

The relative error is indicated in parentheses.

minimized in its own optimization (*i.e.*, ray re-projection error), we additionally evaluate the performance in mean re-projection error with DPW and BJW. As exhibited in Tab. 4, the errors of the proposed method are obviously smaller than those of DPW and BJW, which further verifies the effectiveness of nonlinear optimization (*i.e.* the cost function in Eq. (26)).

Unlike the core idea of DPW, BJW conducts the calibration on the raw data directly instead of sub-aperture images. It poses a significant challenge to obtain line feature accurately which is extracted from raw data to estimate an initial solution of intrinsic parameters. The light field data for calibration must be out of focus to make the measurement detectable. Therefore, as shown in Tab. 4, several datasets, *i.e.* C and D by [17], can not be estimated by BJW.

In order to comprehensively compare with DPW, BJW and MPC, we also carry out calibration on the datasets captured by MPC [15]. Tab. 5 lists the RMS ray re-projection errors compared with DPW, BJW and MPC at two calibration stages. As exhibited in Tab. 5, the proposed method obtains smaller ray re-projection errors on the item of optimization without rectification compared with DPW and MPC. Furthermore, it is more important we achieve small errors once the distortion is introduced in the optimization. According to the item of optimization with rectification, the proposed method outperforms DPW, BJW and MPC. Consequently, such optimization results substantiate that our 6-parameter ray-space projection model is effective to describe sampling of an LFC. Fig. 9 demonstrates the results of pose estimation on datasets of MPC.



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

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

Fig. 9. Pose estimation results of the datasets captured by MPC.

In order to verify the effectiveness of geometric reconstruction of the proposed method compared with state-of-the-art methods, we capture four light fields in real scenes and reconstruct several specific corner points and estimate the distances between them. As shown in Tab. 6, the estimated distances between the reconstructed points are nearly equal to those measured lengths from real objects by rulers (see Fig. 10). For these four measurement examples, the relative errors of distance between reconstructed points demonstrate the performance of the proposed model compared with state-of-the-art methods.

7 Conclusion

The paper has presented a unified framework for intrinsic epipolar geometry for LFCs. Specifically, we have derived ray-space projection matrix and ray-space epipolar geometry by using Plücker parameterization. We have reached a novel 6×6 ray-space fundamental matrix, which generalizes the conventional 3×3 fundamental matrix for pinhole cameras and the generalized epipolar constraint for calibrated LFCs. The ray-space fundamental matrix is a basic algebraic entity of multi-view light fields, of which the properties have been derived. We have provided effective algorithms

to compute this ray-space epipolar geometry as well as LFC calibration, and demonstrated the benefits of applying such ray-space epipolar geometry for various light field based multi-view geometry computations.

Extensive experiments are conducted on synthetic and real light field data, which confirm the effectiveness and robustness of the proposed framework. In the future, we will research projective rectification from the ray-space fundamental matrix. Future work may also include developing minimal solvers for computing ray-space camera pose and metric reconstruction using light field cameras.

Acknowledgments

The work was supported by NSFC under Grant 61531014, 61801396, 62031023. Qi Zhang was also supported by Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under CX201919 and China Scholarship Council (CSC).

References

[1] Lytro, "Lytro redefines photography with light field cameras," http://www.lytro.com, 2011.
[2] Raytrix, "3d light field camera technology," http://www.raytrix.de, 2013.
[3] 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.
[4] Y. Zhang, Z. Li, W. Yang, P. Yu, H. Lin, and J. Yu, "The light field 3d scanner," in *Proc. IEEE Int. Conf. Comput. Photography*, 2017, pp. 1–9.
[5] Y. Zhang, P. Yu, W. Yang, Y. Ma, and J. Yu, "Ray space features for plenoptic structure-from-motion," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4631–4639.
[6] S. Nousias, M. Lourakis, and C. Bergeles, "Large-scale, metric structure from motion for unordered light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3292–3301.
[7] C. Birklbauer and O. Bimber, "Panorama light-field imaging," in *Comput. Graph. Forum*, vol. 33, no. 2, 2014, pp. 43–52.
[8] X. Guo, Z. Yu, S. B. Kang, H. Lin, and J. Yu, "Enhancing light fields through ray-space stitching," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 7, pp. 1852–1861, 2016.
[9] Z. Ren, Q. Zhang, H. Zhu, and Q. Wang, "Extending the FOV from disparity and color consistencies in multiview light fields," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 1157–1161.
[10] M. Guo, H. Zhu, G. Zhou, and Q. Wang, "Dense light field reconstruction from sparse sampling using residual network," in *Proc. Asian Conf. Comput. Vis.*, 2018.
[11] D. G. Dansereau, I. Mahon, O. Pizarro, and S. B. Williams, "Plenoptic flow: Closed-form visual odometry for light field cameras," in *Proc. IEEE Int. Conf. Intell. Robot. and Syst.*, 2011, pp. 4455–4462.
[12] F. Dong, S.-H. Ieng, X. Savatier, R. Etienne-Cummings, and R. Benosman, "Plenoptic cameras in real-time robotics," *Int. J. Robot. Res.*, vol. 32, no. 2, pp. 206–217, 2013.
[13] Y. Li, Q. Zhang, X. Wang, and Q. Wang, "Light field slam based on ray-space projection model," in *Proc. SPIE Optoelectron. Imaging Multimed. Technol. VI*, vol. 11187, 2019, p. 1118706.
[14] 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, 2017.
[15] Q. Zhang, C. Zhang, J. Ling, Q. Wang, and J. Yu, "A generic multi-projection-center model and calibration method for light field cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2539–2552, 2019.
[16] R. Pless, "Using many cameras as one," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2003, pp. 587–593.
[17] 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.
[18] Q. Zhang, J. Ling, Q. Wang, and J. Yu, "Ray-space projection model for light field camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10 121–10 129.
[19] R. Ng, "Digital light field photography," Ph.D. dissertation, Stanford University, 2006.
[20] S. Nousias, F. Chadebecq, J. Pichat, P. Keane, S. Ourselin, and C. Bergeles, "Corner-based geometric calibration of multi-focus plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 957–965.
[21] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Baseline and triangulation geometry in a standard plenoptic camera," *Int. J. Comput. Vis.*, vol. 126, no. 1, pp. 21–35, 2018.
[22] Q. Zhang and Q. Wang, "Common self-polar triangle of concentric conics for light field camera calibration," in *Proc. Asian Conf. Comput. Vis.*, 2018.
[23] Q. Zhang, X. Wang, and Q. Wang, "Light field planar homography and its application," in *Proc. SPIE Optoelectron. Imaging Multimed. Technol. VI*, vol. 11187, 2019, p. 111870S.
[24] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.
[25] M. D. Grossberg and S. K. Nayar, "A general imaging model and a method for finding its parameters," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, 2001, pp. 108–115.
[26] P. Sturm, "Multi-view geometry for general camera models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2005, pp. 206–212.
[27] H. Li, R. Hartley, and J.-H. Kim, "A linear approach to motion estimation using generalized camera models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[28] H. C. Longuet-Higgins, "The reconstruction of a plane surface from two perspective projections," *Proc. of the Royal Society of London. Series B. Biological Sciences*, vol. 227, no. 1249, pp. 399–410, 1986.
[29] R. Hartley, "Projective reconstruction and invariants from multiple images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 10, pp. 1036–1041, 1994.
[30] D. Barath, "Five-point fundamental matrix estimation for uncalibrated cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 235–243.
[31] R. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 6, pp. 580–593, 1997.
[32] Y. Zhou, L. Kneip, and H. Li, "A revisit of methods for determining the fundamental matrix with planes," in *Proc. Int. Conf. Digit. Image Comput. Tech. Appl.* IEEE, 2015, pp. 1–7.
[33] H. Li, "A simple solution to the six-point two-view focal-length problem," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 200–213.
[34] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 0756–777, 2004.
[35] R. Hartley and H. Li, "An efficient hidden variable approach to minimal-case camera motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2303–2314, 2012.
[36] D. Scaramuzza, "1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints," *Int. J. Comput. Vis.*, vol. 95, no. 1, pp. 74–85, 2011.
[37] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. ACM Annu. Conf. Comput. Graph. Interactive Techn.*, 1996, pp. 31–42.
[38] W. V. D. Hodge, W. Hodge, and D. Pedoe, *Methods of algebraic geometry*. Cambridge University Press, 1994, vol. 2.
[39] H. Pottmann and J. Wallner, *Computational line geometry*. Springer Science & Business Media, 2009.
[40] K. Madsen, H. B. Nielsen, and O. Tingleff, *Methods for non-linear least squares problems, 2nd Edition*. Informatics and Mathematical Modelling, Technical University of Denmark, 2004.
[41] R. Hartley, "Self-calibration of stationary cameras," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 5–23, 1997.
[42] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.
[43] D. G. Dansereau, B. Girod, and G. Wetzstein, "Liff: Light field features in scale and depth," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8042–8051.
[44] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

Qi Zhang received the B.E. degree in Electronic and Information Engineering from Xi'an University of Architecture and Technology in 2013, and Master degree in Electrical Engineering from Northwestern Polytechnical University in 2015. He is now a Ph.D. candidate at School of Computer Science, Northwestern Polytechnical University. His research interests include computational photography, light field imaging and processing, multi-view geometry and applications.

Qing Wang (M'05, SM'19) is now a Professor in the School of Computer Science, Northwestern Polytechnical University. He graduated from the Department of Mathematics, Peking University, in 1991. He then joined Northwestern Polytechnical University. In 2000 he obtained Ph.D degree in the Department of Computer Science and Engineering, Northwestern Polytechnical University. In 2006, he was awarded as outstanding talent program of new century by Ministry of Education, China. He is now a Senior Member of IEEE and China Computer Federation (CCF). He is also a Member of ACM. He worked as research assistant and research scientist in the Department of Electronic and Information Engineering, the Hong Kong Polytechnic University from 1999 to 2002. He also worked as a visiting scholar in 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. Prof. Wang's research interests include computer vision and computational photography, such as 3D reconstruction, multi-view geometry, light field imaging and processing. He has published more than 100 papers in the international journals and conferences.

Dr. Hongdong Li is currently a Professor with the Computer Vision Group of ANU (Australian National University). He is also a Chief Investigator for the Australia ARC Centre of Excellence for Robotic Vision (ACRV). His research interests include 3D vision reconstruction, structure from motion, multi-view geometry, as well as applications of optimization methods in computer vision. Prior to 2010, he was with NICTA Canberra Labs working on the Australia Bionic Eyes project. He is an Associate Editor for IEEE T-PAMI, and served as Area Chair in recent year ICCV, ECCV and CVPR. He was a Program Chair for ACRA 2015 Australia Conference on Robotics and Automation, and a Program Co-Chair for ACCV 2018 Asian Conference on Computer Vision. He won a number of best paper awards in computer vision and pattern recognition, and was the receipt for the CVPR 2012 Best Paper Award, the ICCV Marr Prize Honorable Mention in 2017, and a shortlist of the CVPR 2020 best paper award.

Jingyi Yu is Director of Virtual Reality and Visual Computing Center in the School of Information Science and Technology at ShanghaiTech University. He received B.S. from Caltech in 2000 and Ph.D. from MIT in 2005. He is also a full professor at the University of Delaware. His research interests span a range of topics in computer vision and computer graphics, especially on computational photography and non-conventional optics and camera designs. He has published over 100 papers at highly refereed conferences and journals including over 50 papers at the premiere conferences and journals CVPR/ICCV/ECCV/TPAMI. 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 IEEE TPAMI, Elsevier CVIU, Springer TVCJ and Springer MVA.