Supplementary Material: Ray-Space Projection Model for Light Field Camera*

Qi Zhang[1], Jinbo Ling[1], Qing Wang[1], and Jingyi Yu[2]

[1]School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, P.R. China,
qwang@nwpu.edu.cn
[2]ShanghaiTech University, Shanghai 200031, P.R. China

As the supplementary material, we provide detailed derivation of the ray-space projection model in Sec. 3, the linear form and cost function in Sec. 4.

1. Equation

Before presenting the material, it helps to go over some equations involving cross product. If $\boldsymbol{a} = (a_1, a_2, a_3)^\top$ is a 3D column vector, one defines a corresponding skew-symmetric matrix as follows [2]

$$[\boldsymbol{a}]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}. \tag{1}$$

The cross product of two vectors $\boldsymbol{a} \times \boldsymbol{b}$ is related to skew-symmetric matrices according to

$$\boldsymbol{a} \times \boldsymbol{b} = [\boldsymbol{a}]_\times \boldsymbol{b} = -[\boldsymbol{b}]_\times \boldsymbol{a}. \tag{2}$$

Let $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ be three vectors in $\mathbb{R}^3$. The following associative law holds,

$$\boldsymbol{a}^\top(\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \times \boldsymbol{b})^\top \boldsymbol{c}. \tag{3}$$

1.1. Ray-Space Intrinsic Matrix (RSIM)

According to the homogeneous decoding matrix mentioned in the submission, we rewrite the decoding matrix as,

$$\begin{cases} s = k_i\, i, \\ t = k_j\, j, \\ x = k_u\, u + u_0, \\ y = k_v\, v + v_0, \end{cases} \tag{4}$$

where $(k_i, k_j, k_u, k_v, u_0, v_0)$ are intrinsic parameters of a light field camera. Eq. (4) represents the relationship between the light field $L(i, j, u, v)$ recorded by the camera and the undistorted physical light field $L(s, t, x, y)$. As mentioned in the submission, the moment vector and direction vector of the undistorted physical ray $\boldsymbol{r} = (s, t, x, y)$ are defined as,

$$\begin{cases} \boldsymbol{m} = (s, t, 0)^\top \times (x, y, 1)^\top = (t, -s, sy - tx)^\top \\ \boldsymbol{q} = (x, y, 1)^\top \end{cases}, \tag{5}$$

where $(\boldsymbol{m}^\top, \boldsymbol{q}^\top)^\top$ are the Plücker coordinates of the ray. Substituting s, t, x, y by Eq. (4), Eq. (5) becomes,

$$\begin{aligned} \boldsymbol{m} &= (k_j\, j, -k_i\, i, k_i\, i(k_v\, v + v_0) - k_j\, j(k_u\, u + u_0))^\top \\ &= (k_j\, j, -k_i\, i, k_i k_v (iv - ju) + k_i v_0\, i - k_j u_0\, j)^\top \end{aligned}, \tag{6}$$

which needs to satisfy the condition $k_s/k_t = k_i/k_j$. Meanwhile, the moment vector $\boldsymbol{n}$ and direction vector $\boldsymbol{p}$ of the ray $\boldsymbol{l} = (i, j, u, v)$ recorded by the light field camera are represented as $(i, j, 0)^\top \times (u, v, 1)^\top = (j, -i, iv - ju)^\top$ and $(u, v, 1)$ respectively. Then, the RSIM $\boldsymbol{K}$ is formulated as,

$$\begin{bmatrix} \boldsymbol{m} \\ \boldsymbol{q} \end{bmatrix} = \underbrace{\begin{bmatrix} k_j & 0 & 0 & 0 & 0 & 0 \\ 0 & k_i & 0 & 0 & 0 & 0 \\ -k_j u_0 & -k_i v_0 & k_i k_v & 0 & 0 & 0 \\ 0 & 0 & 0 & k_u & 0 & u_0 \\ 0 & 0 & 0 & 0 & k_v & v_0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{=:\boldsymbol{K}} \begin{bmatrix} \boldsymbol{n} \\ \boldsymbol{p} \end{bmatrix}. \tag{7}$$

Fig. 1 illustrates the ray-space intrinsic transformations $\mathcal{L}_c = \boldsymbol{K}\mathcal{L}$ and $\mathcal{L}'_c = \boldsymbol{K}'\mathcal{L}'$.

1.2. Fundamental Matrix

There are two constraints of the Plücker coordinates in 3D projective space. The one is $\boldsymbol{m}^\top \cdot \boldsymbol{q} = 0$, the other is *generalized epipolar constraint* [4],

$$\begin{gathered} \boldsymbol{q}'^\top \boldsymbol{E} \boldsymbol{q} + \boldsymbol{q}'^\top \boldsymbol{R} \boldsymbol{m} + \boldsymbol{m}'^\top \boldsymbol{R} \boldsymbol{q} = 0, \\ \begin{bmatrix} \boldsymbol{m}' \\ \boldsymbol{q}' \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{0}_{3\times3} & \boldsymbol{R}^\top \\ \boldsymbol{R}^\top & \boldsymbol{E}^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{m} \\ \boldsymbol{q} \end{bmatrix} = 0 \end{gathered} \tag{8}$$

which is obtained from the theorem that *two lines* $(\boldsymbol{m}, \boldsymbol{q}_1)^\top$ *and* $(\boldsymbol{m}_2, \boldsymbol{q}_2)^\top$ *are coplanar if and only if* $\boldsymbol{m}_1^\top \boldsymbol{q}_2 + \boldsymbol{q}_1^\top \boldsymbol{m}_2 =$

*The work was supported by NSFC under Grant 61531014.

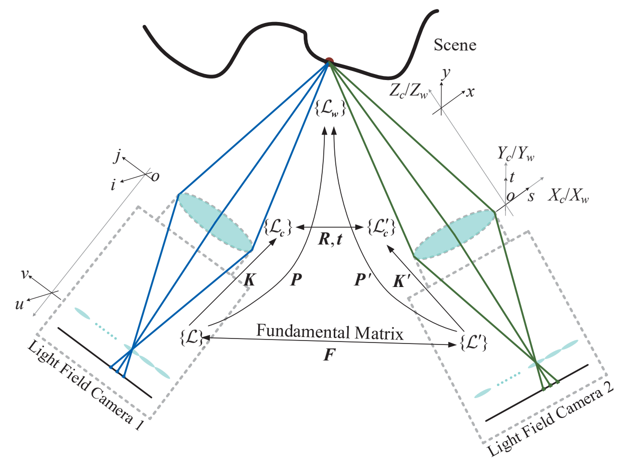

Figure 1. Ray-space projection model and ray-ray transformation among two light field cameras.

0 . $\boldsymbol{R}$, $\boldsymbol{t}$ denote the rotation and translation between two light field cameras coordinates.

According to ray-space intrinsic matrix Eq. (7) and generalized epipolar constraint Eq. (8), we formulate the ray-ray transformation $\{\mathcal{L}'\} \leftrightarrow \{\mathcal{L}\}$ as shown in Fig. 1,

$$\mathcal{L}'^{\top} \underbrace{\boldsymbol{K}'^{\top} \begin{bmatrix} \boldsymbol{0}_{3\times 3} & \boldsymbol{R}^{\top} \\ \boldsymbol{R}^{\top} & \boldsymbol{E}^{\top} \end{bmatrix} \boldsymbol{K}}_{=:\boldsymbol{F}} \mathcal{L} = 0. \tag{9}$$

$\boldsymbol{F}$ is the fundamental matrix. $\mathcal{L}'^{\top}\boldsymbol{F}\mathcal{L} = 0$ represents the ray-ray correspondences among two light fields.

1.3. The Relationship between Rays and Planes

According to the theorem [3], a Plücker line $(\boldsymbol{m}^{\top}, \boldsymbol{q}^{\top})^{\top}$ insects with a plane in the point with homogeneous coordinate,

$$\boldsymbol{X} = (\boldsymbol{\pi} \times \boldsymbol{m} - d\,\boldsymbol{q}) / \boldsymbol{\pi}^{\top}\boldsymbol{q}, \tag{10}$$

where the plane in 3D space can be expressed by a homogeneous vector $(\boldsymbol{\pi}^{\top}, d)^{\top}$, $\boldsymbol{\pi} \in \mathbb{R}^3$, $d \in \mathbb{R}$. Therefore, a point $\boldsymbol{X}_w$ in the world coordinates can be described as the intersection of the ray $\mathcal{L}_w = (\boldsymbol{m}_w^{\top}, \boldsymbol{q}_w^{\top})^{\top}$ with the plane $Z = Z_w$(*i.e.* $\boldsymbol{\pi}_w = (0,0,1)^{\top}$, $d_w = -Z_w$),

$$\boldsymbol{X}_w = ([\boldsymbol{\pi}_w]_{\times}\boldsymbol{m}_w + Z_w\boldsymbol{q}_w)/\boldsymbol{\pi}_w^{\top}\boldsymbol{q}. \tag{11}$$

Then, Eq. (11) is extended by Eq. (1),

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{m}_w + \begin{bmatrix} Z_w & 0 & 0 \\ 0 & Z_w & 0 \\ 0 & 0 & Z_w \end{bmatrix} \boldsymbol{q}_w. \tag{12}$$

Being substituted by

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} 0 & 0 & X_w \\ 0 & 0 & Y_w \\ 0 & 0 & Z_w \end{bmatrix} \boldsymbol{q}_w, \tag{13}$$

Eq. (12) becomes

$$\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{m}_w + \begin{bmatrix} Z_w & 0 & -X_w \\ 0 & Z_w & -Y_w \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{q}_w = \boldsymbol{0}. \tag{14}$$

Then, Eq. (14) is simplified as

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & Z_w & -Y_w \\ 0 & 1 & 0 & -Z_w & 0 & X_w \end{bmatrix}}_{=:\boldsymbol{M}(\boldsymbol{X}_w)} \begin{bmatrix} \boldsymbol{m}_w \\ \boldsymbol{q}_w \end{bmatrix} = \boldsymbol{0}. \tag{15}$$

1.4. Linear Form for the Calibration

As we have mentioned in the submission, we use the R-SIM, ray-space extrinsic matrix and Eq. (15) to represent the relationship between a world point and its rays, *i.e.*,

$$\boldsymbol{M}(\boldsymbol{X}_w) \begin{bmatrix} \boldsymbol{R}^{\top} & \boldsymbol{E}^{\top} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{R}^{\top} \end{bmatrix} \boldsymbol{K} \begin{bmatrix} \boldsymbol{n} \\ \boldsymbol{p} \end{bmatrix} = 0, \tag{16}$$

where $\boldsymbol{K}$ is the RSIM which is abbreviated to a lower triangle matrix $\boldsymbol{K}_{ij}$ and a upper trinangle matrix $\boldsymbol{K}_{uv}$. According to an essential assumption that the checkerboard is on the plane $Z_w = 0$ in the world coordinates, Eq. (16) is simplified as,

$$\underbrace{\begin{bmatrix} 1 & 0 & -Y_w \\ 0 & 1 & X_w \end{bmatrix}}_{=:\boldsymbol{M}_s} \boldsymbol{H}_s \begin{bmatrix} \boldsymbol{n} \\ \boldsymbol{p} \end{bmatrix} = 0, \tag{17}$$

where $\boldsymbol{H}_s$ is the simplified ray-sapce projection matrix. We utilize the direct product operator to compute $\boldsymbol{H}_s$. Subsequently, $\boldsymbol{H}_s$ denotes a 3×6 matrix only using intrinsic and extrinsic parameters,

$$\boldsymbol{H}_s = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_3 \\ \boldsymbol{h}_2 & \boldsymbol{h}_4 \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{h}_5 \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_1^{\top} & -\boldsymbol{r}_1^{\top}[\boldsymbol{t}]_{\times} \\ \boldsymbol{r}_2^{\top} & -\boldsymbol{r}_2^{\top}[\boldsymbol{t}]_{\times} \\ \boldsymbol{0}_{1\times 3} & \boldsymbol{r}_3^{\top} \end{bmatrix} \begin{bmatrix} \boldsymbol{K}_{ij} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{K}_{uv} \end{bmatrix}, \tag{18}$$

where $\boldsymbol{h}_i$ denotes the row vector (h_{i1}, h_{i2}, h_{i3}).

$$\begin{cases} \boldsymbol{h}_1 = \boldsymbol{r}_1^{\top}\boldsymbol{K}_{ij}, \\ \boldsymbol{h}_2 = \boldsymbol{r}_2^{\top}\boldsymbol{K}_{ij}, \\ \boldsymbol{h}_3 = ([\boldsymbol{t}]_{\times}\boldsymbol{r}_1)^{\top}\boldsymbol{K}_{uv}, \\ \boldsymbol{h}_4 = ([\boldsymbol{t}]_{\times}\boldsymbol{r}_2)^{\top}\boldsymbol{K}_{uv}, \\ \boldsymbol{h}_5 = \boldsymbol{r}_3\boldsymbol{K}_{uv}. \end{cases} \tag{19}$$

Then, we utilize the orthogonality of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ and the Cholesky factorization [1] to obtain the estimated intrinsic matrix $\hat{\boldsymbol{K}}_{ij}$ which is determined up to an unknown scale factor. The effect of the scale factor is eliminated by calculating the ratio of elements. Therefore, we compute the intrinsic matrix $\hat{\boldsymbol{K}}_{uv}$. Note that the matrix $\boldsymbol{H}_s$ occurring in Eq. (17) may be changed by multiplication by an arbitrary non-zero scale factor without altering the projective transformation. According to the last formula of Eq. (19) and unit norm of $\boldsymbol{r}_i$, we calculate a scale factor

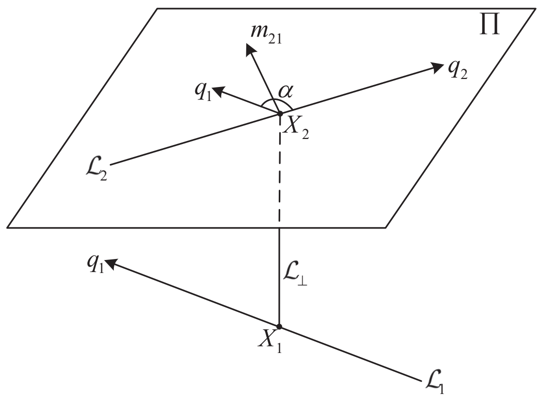

Figure 2. The distance between $\mathcal{L}_1$ and $\mathcal{L}_2$. The common perpendicular $\mathcal{L}_\perp$ intersects two lines at $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ respectively.

$\tau = 1/||\hat{\boldsymbol{K}}_{uv}^{-\top}\boldsymbol{h}_5^\top||$. Adopting Eq. (2) to Eq. (19), it can be rewritten as,

$$\begin{cases} -[\boldsymbol{r}_1]_\times t = \tau\hat{\boldsymbol{K}}_{uv}^{-\top}\boldsymbol{h}_3^\top \\ -[\boldsymbol{r}_2]_\times t = \tau\hat{\boldsymbol{K}}_{uv}^{-\top}\boldsymbol{h}_4^\top \end{cases}. \tag{20}$$

These equations can be solved by linear least-squares techniques to obtain the translation $\boldsymbol{t}$, *i.e.*

$$\begin{aligned} \boldsymbol{t} &= (\boldsymbol{G}^\top\boldsymbol{G})^{-1}(\boldsymbol{G}^\top\boldsymbol{g}), \\ \boldsymbol{G} &= (-[\boldsymbol{r}_1]_\times, -[\boldsymbol{r}_2]_\times)^\top, \\ \boldsymbol{g} &= (\tau\hat{\boldsymbol{K}}_{ij}^{-\top}\boldsymbol{h}_3^\top, \tau\hat{\boldsymbol{K}}_{ij}^{-\top}\boldsymbol{h}_4^\top)^\top. \end{aligned} \tag{21}$$

1.5. Ray-to-Ray Cost Function

In the submission, a ray-to-ray cost function is established for non-linear optimization. The ray-to-ray cost is described as the distance between rays, as shown in Fig. 2. It illustrates the distance between $\mathcal{L}_1 = (\boldsymbol{m}_1^\top, \boldsymbol{q}_1^\top)^\top$ and $\mathcal{L}_2 = (\boldsymbol{m}_2^\top, \boldsymbol{q}_2^\top)^\top$. The lines $\mathcal{L}_1$ and $\mathcal{L}_2$ are not parallel to each other. Refereing to Fig. 2, the plane Π containing $\boldsymbol{X}_2$ and $\boldsymbol{m}_{21}$ is orthogonal to $\mathcal{L}_\perp$ and parallel to $\mathcal{L}_1$. $\boldsymbol{m}_{21}$ is the moment vector of $\mathcal{L}_2$ about a point $\boldsymbol{X}_1$ on $\mathcal{L}_1$. This moment is defined as,

$$\begin{aligned} \boldsymbol{m}_{21} &= (\boldsymbol{X}_2 - \boldsymbol{X}_1) \times \boldsymbol{q}_2 \\ &= \boldsymbol{m}_2 - \boldsymbol{X}_1 \times \boldsymbol{q}_2 \end{aligned}. \tag{22}$$

In Fig. 2, α is the angle of rotation from $\boldsymbol{q}_1$ to $\boldsymbol{q}_2$. We can obtain $|\sin\alpha| = ||\boldsymbol{q}_1 \times \boldsymbol{q}_2||/(||\boldsymbol{q}_1|| \cdot ||\boldsymbol{q}_2||)$. Since $\boldsymbol{m}_{21} \perp \boldsymbol{q}_2$, we drive,

$$\begin{aligned} \boldsymbol{q}_1^\top\boldsymbol{m}_{21} &= ||\boldsymbol{m}_{21}|| \cdot ||\boldsymbol{q}_1|| \cos(\alpha - \frac{\pi}{2}) \\ &= ||\boldsymbol{m}_{21}|| \cdot ||\boldsymbol{q}_1||| \sin\alpha| \end{aligned}, \tag{23}$$

The above yields the distance between $\mathcal{L}_1$ and $\mathcal{L}_2$,

$$\begin{aligned} d = ||\boldsymbol{X}_2 - \boldsymbol{X}_1|| = \frac{||\boldsymbol{m}_{21}||}{||\boldsymbol{q}_2||} &= \frac{|\boldsymbol{q}_1^\top\boldsymbol{m}_{21}|}{||\boldsymbol{q}_1|| \cdot ||\boldsymbol{q}_2||| \sin\alpha|} \\ = \frac{|\boldsymbol{q}_1^\top\boldsymbol{m}_{21}|}{||\boldsymbol{q}_1 \times \boldsymbol{q}_2||} &= \frac{|\boldsymbol{q}_1^\top(\boldsymbol{m}_2 - \boldsymbol{X}_1 \times \boldsymbol{q}_2)|}{||\boldsymbol{q}_1 \times \boldsymbol{q}_2||} \\ &= \frac{|\boldsymbol{q}_1^\top\boldsymbol{m}_2 - (\boldsymbol{q}_1 \times \boldsymbol{X}_1)^\top\boldsymbol{q}_2|}{||\boldsymbol{q}_1 \times \boldsymbol{q}_2||} \\ &= \frac{|\boldsymbol{q}_1^\top\boldsymbol{m}_2 + \boldsymbol{m}_1^\top\boldsymbol{q}_2|}{||\boldsymbol{q}_1 \times \boldsymbol{q}_2||} \end{aligned} \tag{24}$$

References

[1] Richard Hartley. Self-calibration of stationary cameras. *IJCV*, 22(1):5–23, 1997.

[2] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[3] William Vallance Douglas Hodge, WVD Hodge, and Daniel Pedoe. *Methods of algebraic geometry*, volume 2. Cambridge University Press, 1994.

[4] Robert Pless. Using many cameras as one. In *IEEE CVPR*, volume 2, pages 587–593, 2003.