Supplementary Material: Geometric Calibration of Micro-Lens-Based Light-Field Cameras using Line Features

Yunsu Bok, Hae-Gon Jeon and In So Kweon

KAIST, Korea

As the supplementary material, we provide detailed derivation of the projection model in Sect. 3 and the linear form in Sect. 4. We also present the whole process of generating sub-aperture images using our calibration result.

1 Projection model of micro-lens-based light-field cameras

Micro-lens-based light-field cameras contain two layers of lenses; main lens and micro-lens array. We apply 'thin lens model' to main lens and 'pinhole model' to micro-lenses, similar to [1].

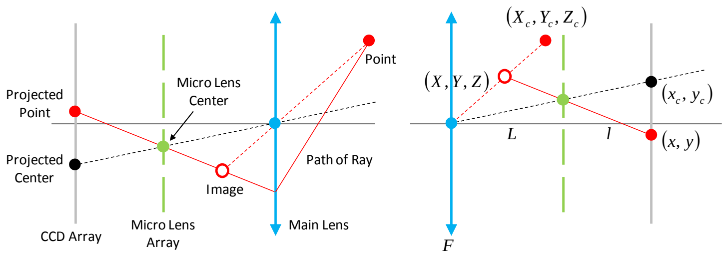


Fig. 1. Projection model of micro-lens-based light-field cameras.

Projected location (x, y) of the 'image' in Fig. 1 is computed by extending line connecting the image and a micro-lens center as follows:

$$
\begin{aligned}
(L+l)\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} &= \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \frac{L+l-Z}{L-Z}\left(L\begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} - \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) \\
&= \left(1-\frac{L+l-Z}{L-Z}\right)\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \frac{L+l-Z}{L-Z}\cdot L\begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \\
&= -\frac{l}{L-Z}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \frac{L+l-Z}{L-Z}\cdot L\begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix}.
\end{aligned}
\tag{1}
$$

The third term of (1) is simplified as

$$-\frac{l}{L-Z}Z+\frac{L+l-Z}{L-Z}L=\frac{-lZ+L^2+Ll-LZ}{L-Z}=\frac{(L-Z)(L+l)}{L-Z}=L+l. \quad (2)$$

The first and second terms of (1) divided by $(L+l)$ are

$$\begin{bmatrix} x \\ y \end{bmatrix} = -\frac{l}{(L-Z)(L+l)}\begin{bmatrix} X \\ Y \end{bmatrix} + \frac{L(L+l-Z)}{(L-Z)(L+l)}\begin{bmatrix} x_c \\ y_c \end{bmatrix}. \quad (3)$$

Equation (3) is simplified by subtracting projected micro-lens center (x_c, y_c) from it:

$$\begin{aligned}
\begin{bmatrix} x-x_c \\ y-y_c \end{bmatrix} &= -\frac{l}{(L-Z)(L+l)}\begin{bmatrix} X \\ Y \end{bmatrix} + \left(\frac{L(L+l-Z)}{(L-Z)(L+l)}-1\right)\begin{bmatrix} x_c \\ y_c \end{bmatrix} \\
&= -\frac{l}{(L-Z)(L+l)}\begin{bmatrix} X \\ Y \end{bmatrix} + \frac{L(L+l-Z)-(L-Z)(L+l)}{(L-Z)(L+l)}\begin{bmatrix} x_c \\ y_c \end{bmatrix} \\
&= -\frac{l}{(L-Z)(L+l)}\begin{bmatrix} X \\ Y \end{bmatrix} + \frac{(L^2+Ll-LZ)-(L^2+Ll-LZ-lZ)}{(L-Z)(L+l)}\begin{bmatrix} x_c \\ y_c \end{bmatrix} \quad (4) \\
&= -\frac{l}{(L-Z)(L+l)}\begin{bmatrix} X \\ Y \end{bmatrix} + \frac{lZ}{(L-Z)(L+l)}\begin{bmatrix} x_c \\ y_c \end{bmatrix} \\
&= \frac{l}{(L-Z)(L+l)}\left(-\begin{bmatrix} X \\ Y \end{bmatrix} + Z\begin{bmatrix} x_c \\ y_c \end{bmatrix}\right).
\end{aligned}$$

Being substituted by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{F}{Z_c-F}\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}, \quad (5)$$

which is computed using the thin-lens model, (4) becomes

$$\begin{aligned}
\begin{bmatrix} x-x_c \\ y-y_c \end{bmatrix} &= \frac{l}{(L-\frac{F}{Z_c-F}Z_c)(L+l)}\left(-\frac{F}{Z_c-F}\begin{bmatrix} X_c \\ Y_c \end{bmatrix} + \frac{F}{Z_c-F}Z_c\begin{bmatrix} x_c \\ y_c \end{bmatrix}\right) \\
&= \frac{lF}{(L(Z_c-F)-FZ_c)(L+l)}\left(-\begin{bmatrix} X_c \\ Y_c \end{bmatrix} + Z_c\begin{bmatrix} x_c \\ y_c \end{bmatrix}\right) \\
&= \frac{LF}{(L-F)Z_c-LF}\cdot\frac{l}{L(L+l)}\left(-\begin{bmatrix} X_c \\ Y_c \end{bmatrix} + Z_c\begin{bmatrix} x_c \\ y_c \end{bmatrix}\right) \quad (6) \\
&= \frac{1}{(1/F-1/L)Z_c-1}\cdot\frac{1}{L(L/l+1)}\left(-\begin{bmatrix} X_c \\ Y_c \end{bmatrix} + Z_c\begin{bmatrix} x_c \\ y_c \end{bmatrix}\right) \\
&= \frac{1}{K_2(K_1Z_c-1)}\left(-\begin{bmatrix} X_c \\ Y_c \end{bmatrix} + Z_c\begin{bmatrix} x_c \\ y_c \end{bmatrix}\right),
\end{aligned}$$

where $K_1 \equiv 1/F - 1/L$ and $K_2 \equiv L(L/l+1)$.

2 Calibration of micro-lens-based light-field cameras

As we have mentioned in the paper, we use a simple intrinsic model (i.e., zero skew, single focal length and zero center) to compute an initial solution:

$$\begin{bmatrix} u \\ v \end{bmatrix} = f \begin{bmatrix} x \\ y \end{bmatrix}, \tag{7}$$

where f is the focal length of intrinsic parameters. Actually it indicates the number of pixels in one measurement unit (millimeter in this paper). Adopting (7) to the projection model (6), it becomes

$$\begin{bmatrix} u - u_c \\ v - v_c \end{bmatrix} = \frac{1}{K_2(K_1 Z_c - 1)} \left(-f \begin{bmatrix} X_c \\ Y_c \end{bmatrix} + Z_c \begin{bmatrix} u_c \\ v_c \end{bmatrix} \right). \tag{8}$$

We apply (8) to line features extracted from raw images of a checkerboard pattern. Projections of corners are close enough to corresponding line features to use an approximation that they lie on the features. Let $ax + by + c = 0$ ($a^2 + b^2 = 1$) be the equation of a line feature, setting micro-lens center (u_c, v_c) as origin. Substituting corner projections into the line equation,

$$a(u - u_c) + b(v - v_c) + c = 0 \tag{9}$$

$$a(-fX_c + Z_c u_c) + b(-fY_c + Z_c v_c) + cK_2(K_1 Z_c - 1) = 0. \tag{10}$$

Let (X_w, Y_w, Z_w) be one of two corners which define a line segment in world coordinate system (i.e., checkerboard coordinate system). It must be transformed into camera coordinate system by an unknown transformation matrix with a 3×3 rotation matrix $\mathbf{R}$ and a 3×1 translation vector $\mathbf{t}$:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \mathbf{t} = \begin{bmatrix} r_{11} X_w + r_{12} Y_w + t_1 \\ r_{21} X_w + r_{22} Y_w + t_2 \\ r_{31} X_w + r_{32} Y_w + t_3 \end{bmatrix}, \tag{11}$$

where r_{ij} and t_i are elements of $\mathbf{R}$ and $\mathbf{t}$ at i-th row and j-th column. Without loss of generality, the z-coordinate of the checkerboard pattern is set to zero ($Z_w = 0$ for all corners). Substituting (X_c, Y_c, Z_c) by (11), (10) becomes

$$\begin{aligned} & a(-f(r_{11}X_w + r_{12}Y_w + t_1) + (r_{31}X_w + r_{32}Y_w + t_3)u_c) \\ & + b(-f(r_{21}X_w + r_{22}Y_w + t_2) + (r_{31}X_w + r_{32}Y_w + t_3)v_c) \\ & + cK_2(K_1(r_{31}X_w + r_{32}Y_w + t_3) - 1) \\ & = -afr_{11}X_w - afr_{12}Y_w - aft_1 + au_c r_{31}X_w + au_c r_{32}Y_w + au_c t_3 \\ & - bfr_{21}X_w - bfr_{22}Y_w - bft_2 + bv_c r_{31}X_w + bv_c r_{32}Y_w + bv_c t_3 \\ & + cK_1K_2r_{31}X_w + cK_1K_2r_{32}Y_w + cK_1K_2t_3 - cK_2 \\ & = -aX_w \cdot fr_{11} - af \cdot r_{12}Y_w - a \cdot ft_1 - bfr_{21}X_w - bfr_{22}Y_w - bft_2 \\ & + (au_c + bv_c) \cdot r_{31}X_w + (au_c + bv_c) \cdot r_{32}Y_w + (au_c + bv_c) \cdot t_3 \\ & + cX_w \cdot K_1K_2r_{31} + cY_w \cdot K_1K_2r_{32} + c \cdot K_2(K_1t_3 - 1) = 0, \end{aligned} \tag{12}$$

which can be expressed in $\mathbf{Ax} = \mathbf{0}$ form:

$$\begin{bmatrix} -aX_w \\ -aY_w \\ -a \\ -bX_w \\ -bY_w \\ -b \\ (au_c + bv_c)X_w \\ (au_c + bv_c)Y_w \\ (au_c + bv_c) \\ cX_w \\ cY_w \\ c \end{bmatrix}^\top \begin{bmatrix} fr_{11} \\ fr_{12} \\ ft_1 \\ fr_{21} \\ fr_{22} \\ ft_2 \\ r_{31} \\ r_{32} \\ t_3 \\ K_1K_2r_{31} \\ K_1K_2r_{32} \\ K_2(K_1t_3 - 1) \end{bmatrix} = 0. \tag{13}$$

3 Generating sub-aperture images

A sub-aperture image is a collection of camera rays which pass through a common point. Using the calibration result by the proposed method, we can compute a camera ray corresponding to each point in a raw image. The generalized version of the projection model is described in the paper:

$$\begin{bmatrix} u - u_c \\ v - v_c \end{bmatrix} = \frac{1}{K_2(K_1Z_c - 1)} \left(- \begin{bmatrix} f_xX_c \\ f_yY_c \end{bmatrix} + Z_c \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix} \right). \tag{14}$$

A camera ray corresponding to (u, v) in a micro-lens image centered at (u_c, v_c) passes through a 3D point (X_c, Y_c, Z_c). According to (14), the 3D point is independent of micro-lens center (u_c, v_c) if Z_c is equal to zero. It means that any ray corresponding to a pixel with same displacement $(d_u, d_v) = (u - u_c, v - v_c)$ from micro-lens center passes through a common point $(X_s, Y_s, 0)$:

$$\begin{bmatrix} u - u_c \\ v - v_c \end{bmatrix}_{Z_c=0} = \frac{1}{K_2} \begin{bmatrix} f_xX_s \\ f_yY_s \end{bmatrix} \tag{15}$$

$$\begin{bmatrix} X_s \\ Y_s \end{bmatrix}_{Z_c=0} = K_2 \begin{bmatrix} (u - u_c)/f_x \\ (v - v_c)/f_y \end{bmatrix} = K_2 \begin{bmatrix} d_u/f_x \\ d_v/f_y \end{bmatrix}, \tag{16}$$

so that $(X_s, Y_s, 0)$ becomes the center of a sub-aperture image at (d_u, d_v).

Direction of the ray (x_d, y_d, z_d) is also computed using (14). Since the ray passes both 3D points (X_c, Y_c, Z_c) and $(X_c + x_d, Y_c + y_d, Z_c + z_d)$,

$$\begin{aligned} \begin{bmatrix} u - u_c \\ v - v_c \end{bmatrix} &= \frac{1}{K_2(K_1Z_c - 1)} \left(- \begin{bmatrix} f_xX_c \\ f_yY_c \end{bmatrix} + Z_c \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix} \right) \\ &= \frac{1}{K_2(K_1(Z_c + z_d) - 1)} \left(- \begin{bmatrix} f_x(X_c + x_d) \\ f_y(Y_c + y_d) \end{bmatrix} + (Z_c + z_d) \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix} \right) \end{aligned} \tag{17}$$

$$
\begin{aligned}
&K_2(K_1(Z_c + z_d) - 1)\left(-\begin{bmatrix} f_x X_c \\ f_y Y_c \end{bmatrix} + Z_c \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix}\right) \\
&= K_2(K_1 Z_c - 1)\left(-\begin{bmatrix} f_x(X_c + x_d) \\ f_y(Y_c + y_d) \end{bmatrix} + (Z_c + z_d)\begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix}\right)
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
&- K_1 K_2 z_d \begin{bmatrix} f_x X_c \\ f_y Y_c \end{bmatrix} + K_1 K_2 z_d Z_c \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix} \\
&= -K_2(K_1 Z_c - 1)\begin{bmatrix} f_x x_d \\ f_y y_d \end{bmatrix} + K_2(K_1 Z_c - 1) z_d \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix}.
\end{aligned}
\tag{19}
$$

Dividing (19) by $K_2 z_d$,

$$
(K_1 Z_c - 1)\begin{bmatrix} f_x x_d \\ f_y y_d \end{bmatrix} = K_1 z_d \begin{bmatrix} f_x X_c \\ f_y Y_c \end{bmatrix} - z_d \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix}. \tag{20}
$$

From (14),

$$
\begin{bmatrix} f_x X_c \\ f_y Y_c \end{bmatrix} = Z_c \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix} - K_2(K_1 Z_c - 1)\begin{bmatrix} u - u_c \\ v - v_c \end{bmatrix}. \tag{21}
$$

Substituting $(f_x X_c, f_y Y_c)$ by (21), (20) becomes

$$
\begin{aligned}
&(K_1 Z_c - 1)\begin{bmatrix} f_x x_d \\ f_y y_d \end{bmatrix} \\
&= K_1 z_d \left(Z_c \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix} - K_2(K_1 Z_c - 1)\begin{bmatrix} u - u_c \\ v - v_c \end{bmatrix}\right) - z_d \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix} \\
&= -K_1 K_2 z_d (K_1 Z_c - 1)\begin{bmatrix} u - u_c \\ v - v_c \end{bmatrix} + (K_1 Z_c - 1) z_d \begin{bmatrix} u_c - c_x \\ v_c - c_y \end{bmatrix}
\end{aligned}
\tag{22}
$$

$$
\begin{bmatrix} x_d / z_d \\ y_d / z_d \end{bmatrix} = -K_1 K_2 \begin{bmatrix} (u - u_c)/f_x \\ (v - v_c)/f_y \end{bmatrix} + \begin{bmatrix} (u_c - c_x)/f_x \\ (v_c - c_y)/f_y \end{bmatrix}. \tag{23}
$$

We set the size of sub-aperture images to 328×328 pixels because that of raw images is 3280×3280 pixels and the average distance between neighboring micro-lens centers is around 10 pixels. The parameters in their intrinsic matrix K_{sub} are set to 1/10 of the calibration result f_x, c_x, c_y:

$$
K_{sub} = \begin{bmatrix} f_x/10 & 0 & c_x/10 \\ 0 & f_x/10 & c_y/10 \\ 0 & 0 & 1 \end{bmatrix}. \tag{24}
$$

For each micro-lens image, we compute the correspondences between pixels in a raw image and those in a sub-aperture image. Let us assume that we want to generate a sub-aperture image at (d_x, d_y). We collect pixels with displacement of (d_x, d_y) from micro-lens centers, and compute their corresponding rays using (23). We remove radial distortion from the rays using k_1 and k_2 computed in the calibration process. Finally they are projected onto sub-aperture image by multiplying K_{sub}.

Using Lytro camera, micro-lens centers are placed in a triangular shape as shown in Fig. 2(a). We connect pixels with displacement of (d_x, d_y) from adjacent micro-lens centers to generate triangular meshes. They are transformed into a sub-aperture image (see Fig. 2(b)). For each pixel of the sub-aperture image, we find a mesh which contains the pixel and compute its intensity value by a triangle interpolation (also known as barycentric interpolation). As shown in Fig. 2(c), let (u_1, v_1), (u_2, v_2) and (u_3, v_3) be the points in a sub-aperture image which define a mesh, and y_1, y_2 and y_3 be their intensity value computed by bilinear interpolation using raw image. The intensity value y corresponding to (u, v) inside the mesh is computed as follows:

$$\begin{bmatrix} u_0 \\ v_0 \\ 1 \end{bmatrix} \sim \left(\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \times \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \right) \times \left(\begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} \times \begin{bmatrix} u_3 \\ v_3 \\ 1 \end{bmatrix} \right) \tag{25}$$

$$l_0 = \sqrt{(u - u_0)^2 + (v - v_0)^2} \tag{26}$$

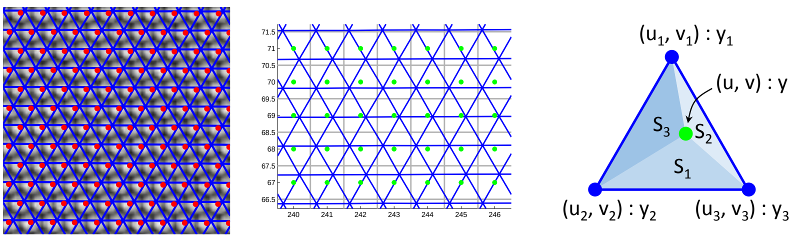
$$l_1 = \sqrt{(u - u_1)^2 + (v - v_1)^2} \tag{27}$$

$$l_2 = \sqrt{(u_0 - u_2)^2 + (v_0 - v_2)^2} \tag{28}$$

$$l_3 = \sqrt{(u_0 - u_3)^2 + (v_0 - v_3)^2} \tag{29}$$

$$\begin{aligned} y &= \frac{S_1 y_1 + S_2 y_2 + S_3 y_3}{S_1 + S_2 + S_3} \\ &= \frac{l_0}{l_0 + l_1} y_1 + \frac{l_1}{l_0 + l_1} \left(\frac{l_3}{l_2 + l_3} y_2 + \frac{l_2}{l_2 + l_3} y_3 \right), \end{aligned} \tag{30}$$

where (u_0, v_0) is the intersection of two lines: one connecting (u_1, v_1) and (u, v), and the other connecting (u_2, v_2) and (u_3, v_3).



(a) (b) (c)

Fig. 2. (a) An example of triangular meshes (blue line) generated by connecting pixels with displacement of $(2, -1)$ pixels from adjacent micro-lens centers (red dots). (b) The meshes are transformed into a sub-aperture image to compute intensity value of each pixel (green dots) of the sub-aperture image. Grey lines indicate pixel boundaries of the sub-aperture image. (c) Notation for triangle interpolation. The intensity value y at pixel (u, v) is computed by a weighted sum of three vertices of a mesh which contains the pixel. Weights of the vertices are proportional to the area of triangles at their opposite side.

References

1. Dansereau, D.G., Pizarro, O., Williams, S.B.: Decoding, calibration and rectification for lenselet-based plenoptic cameras. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1027–1034 (2013)