RECTIFYING PROJECTIVE DISTORTION IN 4D LIGHT FIELD

Chunping Zhang, Zhe Ji, Qing Wang

School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

ABSTRACT

The accuracy of calibration will significantly affect the post processing capability of light field imaging. The geometry of the reconstructed scene is related to the parameters of light field closely, involving the accuracy of decoded rays and ambiguities from ray correspondences. Through exploring the ray correspondence, we derive a transformation matrix to describe the projective distortion on reconstructed scene in 4D light field. Based on our derivation, we simplify the light field camera geometry as a 4-parameter model and calibrate its intrinsic parameters, including a linear initialization and a nonlinear refine process. The proposed light field calibration can simply be implemented with a parallel bi-planar board. Experiments on both simulation and real scene data validate the performance of the calibration.

Index Terms— Light field camera, projective distortion, intrinsic parameter, calibration

1. INTRODUCTION

The light field cameras, e.g. the plenoptic camera [1] and the focused plenoptic camera designed by Georgiev [2], capture both angular and spatial information of rays in 3D space. In contrast to the principle of the conventional camera, its 2D raw image can be decoded into 4D light field via its intrinsic parameters [3], which is able to recover the 3D structure of the real world scene from a single shot by point correspondence [4]. However, the reconstruction process is still an ill-posed problem, especially with unknown intrinsic parameters [5].

Due to projective distortion, imaging deviation and ambiguous parameterization, to reconstruct the real world scene accurately, one crucial step is decoding and calibrating light field with reasonable rectification model. Prior work has explored with the calibration of the intrinsic parameters. Dansereau et al. [3] presented a 15-parameter plenoptic camera model to relate pixels physical rays in the target scene. While its initialization was similar to conventional camera calibration on sub-aperture images, that may not guarantee the convergence of the solution. Johannsen et al. [6] estimated the intrinsic and extrinsic parameters for a focused plenoptic camera by direct depth estimating from the raw image. Due to the ideal model that the geometric center of micro image lies on its corresponding optical center of micro-lens, there was non-negligible measurement error [7]. Bok et al. [8] formulated a geometric projection model to calibrate the intrinsic and extrinsic parameters by utilizing raw images directly, including linear initialization and nonlinear optimization. In this paper, we present a simple and generic model to describe a light field camera base on two-parallel-plane (TPP) model [9, 10]. Our goal is to rectify the projective distortion of the reconstructed scene from 4D light field.

Furthermore, in traditional multi-view geometry, multiple cameras in different poses are defined as a set of unconstrained rays, which is also known as Generalized Camera Model (GCM) [11]. In GCM, correspondences are established by a set of rays intersecting at the same scene point to constrain the poses of cameras. Similarly, the rays decoded from a light field camera are equivalent to a camera array with camera centers aligned regularly on the aperture of its main lens [12, 13]. The ambiguous intrinsic parameters, which can be regarded as the pose of multiple conventional cameras, distort the scene structure recovered from the 4D light field. It is a special case of GMC. In this paper, we model the relationship between the recovered points and the parameters setting to calibrate a light field camera and eliminate the projective distortion of the recovered scene structure.

In summary, our main contributions are listed as follows:

(1) We deduce a transformation matrix to describe the relationship between the distorted points and the scene points based on TPP model.

(2) We propose a method to rectify and calibrate a light field camera, which includes a linear initialization to prove the uniqueness for the intrinsic parameters and a cost function to refine the result.

2. LIGHT FIELD PROJECTIVE DISTORTION

We use a TPP coordinate to describe a 4D light field, as shown in Fig.1, including the camera coordinate X-Y-Z, a 2D coordinate x-y with origin lying on $(0,0,0)^{\mathrm{T}}$, and u-v with origin lying on $(0,0,f)^{\mathrm{T}}$, where f is the distance of the two parallel planes in TPP model. It defines a unique ray $\vec{\boldsymbol{r}}$ as 4 parameters which intersects at points $(x,y,0)^{\mathrm{T}}$ and $(u,v,f)^{\mathrm{T}}$:

The work is supported by NSFC funds (61272287, 61531014) and research grant of State Key Laboratory of Virtual Reality Technology and Systems (BUAAVR-15KF-10).

$\vec{\boldsymbol{r}} = (x, y, u, v)^{\mathrm{T}}$.

Let $\vec{\boldsymbol{r}}_i$ and $\vec{\boldsymbol{r}}_j$ intersect at point $(X, Y, Z)^{\mathrm{T}}$, we can get the relationship between the rays and the point by triangulation:

$$\begin{bmatrix} f & 0 & x_i - u_i & -fx_i \\ 0 & f & y_i - v_i & -fy_i \\ f & 0 & x_j - u_j & -fx_j \\ 0 & f & y_j - v_j & -fy_j \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0, \tag{1}$$

where $(X, Y, Z)^{\mathrm{T}}$ can be solved iff $\frac{u_i - u_j}{x_i - x_j} = \frac{v_i - v_j}{y_i - y_j}$.

Moreover, X, Y and Z can be represented by the elements of $\vec{\boldsymbol{r}}_i$ and $\vec{\boldsymbol{r}}_j$ in two equivalent forms:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{(x_i - x_j) - (u_i - u_j)} \begin{bmatrix} x_i u_j - x_j u_i \\ v_i(x_i - x_j) - y_i(u_i - u_j) \\ f(x_i - x_j) \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{(y_i - y_j) - (v_i - v_j)} \begin{bmatrix} u_i(y_i - y_j) - x_i(v_i - v_j) \\ y_i v_j - y_j v_i \\ f(y_i - y_j) \end{bmatrix} \tag{3}$$

If we change f into f', the point (u, v, f) intersected by $\vec{\boldsymbol{r}}$ becomes (u, v, f') and the rays' intersection becomes $\boldsymbol{X}'$. Substituting it into Eqs.2 and 3, we have:

$$\boldsymbol{X}' = \boldsymbol{P}_1(f) \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & f'/f & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{4}$$

where $\boldsymbol{X}'$ is in homogeneous coordinate. Let (u', v') be $(u + u_0, v + v_0)$, we get $\boldsymbol{X}'$ from Eqs.2 and 3:

$$\boldsymbol{X}' = \boldsymbol{P}_2(u_0, v_0) \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & u_0/f & 0 \\ 0 & 1 & v_0/f & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{5}$$

Let (x', y', u', v') be $(k_x x, k_y y, k_u u, k_v v)$, we get $\boldsymbol{X}'$ from Eqs.2 and 3:

$$s\boldsymbol{X}' = \boldsymbol{P}_3(\boldsymbol{k}) \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} fk_x k_u & 0 & 0 & 0 \\ 0 & fk_x k_v & 0 & 0 \\ 0 & 0 & fk_x & 0 \\ 0 & 0 & k_x - k_u & fk_u \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{6}$$

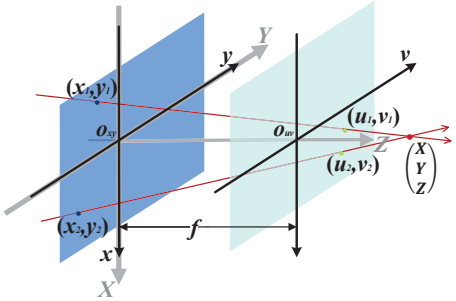

Fig. 1. TPP model with distance f. The origins of coordinate x-y and u-v lie on the axis Z. Two rays $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$ intersect at point $(X, Y, Z)^{\mathrm{T}}$.

$$t\boldsymbol{X}' = \boldsymbol{P}_4(\boldsymbol{k}) \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} fk_y k_u & 0 & 0 & 0 \\ 0 & fk_y k_v & 0 & 0 \\ 0 & 0 & fk_y & 0 \\ 0 & 0 & k_y - k_v & fk_v \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{7}$$

where $\boldsymbol{k} = (k_x, k_y, k_u, k_v)$. More importantly, Eqs.6 and 7 hold when $k_x/k_u = k_y/k_v$.

Indeterminate parameters of a TPP coordinate system result in projective distortion on the scene structure according to Eqs.4, 5, 6 and 7. As shown in the most left one of Fig.2, there is a scene with a Lambert cube recorded by a TPP light field. If the parameters f, (k_u, k_v), and (u_0, v_0) are changed respectively and the light intensity keeps constant, the intersections of the rays will be transformed by $\boldsymbol{P}_i (1 \leq i \leq 4)$. Therefore, the cube will be distorted by the change of the parameters setting of TPP (the right three of Fig.2).

As for a light field camera, every image pixel can be regarded as a ray passing through the coordinate on the image sensor and the optical center of its corresponding micro-lens [3, 7, 8] (in Fig.3). Due to the refraction of the main lens, directions of rays deviate by the focal lens. Consequently, to rectify the difference between the light field inside and outside the camera, we present a 4-intrinsic-parameter model to constrain the rays based on the projective transformation.

3. RECTIFICATION MODEL

We describe a light field camera by 4 intrinsic parameters equivalently, including the distance between the image sen-

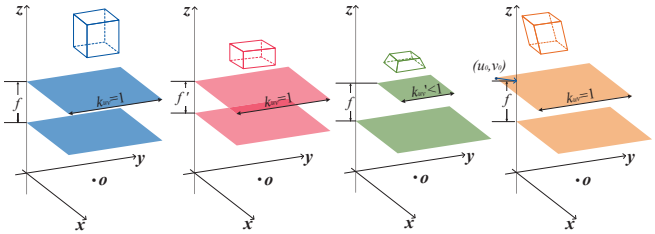

Fig. 2. TPP light field recording a Lambert cube. The left one is the original cube and the others are distorted cubes dependent on TPP parameters.

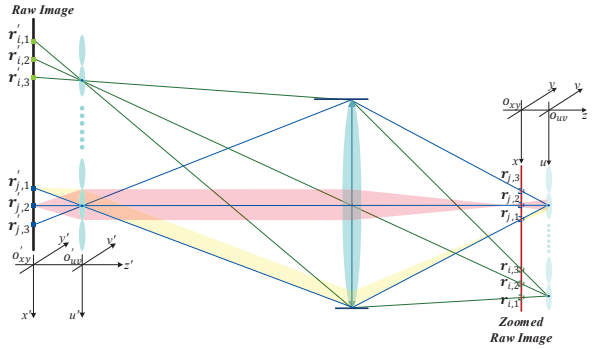

Fig. 3. The optical path of a plenoptic camera [1]. There are two TPP coordinates inside the camera and in the real world scene with different parameters.

sor and the micro-lens array f, the diameter of micro-lens d, and the offset of the optical center of the reference micro-lens (u_0, v_0) (in Fig.4). The coordinates of the optical centers of the micro-lenses are presented as $(id+u_0, jd+v_0, f)$, with label $i \in \mathbb{Z}$, $j \in \mathbb{Z}$, thus the label of reference micro-lens is $(0,0)$.

Consequently, a ray $\vec{\boldsymbol{r}}$ with a certain image coordinate (x, y) which intersects at the optical center of its corresponding micro-lens is constrained by the intrinsic parameters $\mathscr{X}=(u_0, v_0, d, f)$. Then the intersection of the corresponding rays $\boldsymbol{X}_c$ is obtained by Eqs.2 and 3. In addition, to simplify the discussion, we assume that the layout of the micro-lens array is square-like. For hexagon-like configuration, it is easy to partition the entire array into two square-like ones.

More importantly, (u_0, v_0, d, f) is not equal to the physical parameters. Otherwise there will be projective distortions on the recovered rays' intersections.

3.1. Linear Initialization

Given an arbitrary parameters setting $\mathscr{X}'=(u_0', v_0', d', f')$, the distorted intersection $\boldsymbol{X}_d$ is obtained. The relationship between $\boldsymbol{X}_d$, the undistorted intersections $\boldsymbol{X}_c$, $\mathscr{X}'$, and $\mathscr{X}$ is:

$$s\boldsymbol{X}_d = \boldsymbol{P}\boldsymbol{X}_c = \begin{bmatrix} fd'/d & 0 & u_0'-u_0 & 0 \\ 0 & fd'/d & v_0'-v_0 & 0 \\ 0 & 0 & f' & 0 \\ 0 & 0 & 1-d'/d & fd'/d \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}, \quad (8)$$

where $\boldsymbol{P} = \boldsymbol{P}_1(f')\boldsymbol{P}_2(u_0'-u_0, v_0'-v_0)\boldsymbol{P}_3(d'/d)$ are inferred from Eqs.4, 5, 6 and 7.

Then we assume that the coordinate of the world coordinate system $\boldsymbol{X}_w$ is related to the TPP coordinate system $\boldsymbol{X}_c$ by a rigid motion and a zoom, $\boldsymbol{X}_c = K(\boldsymbol{R}\boldsymbol{X}_w + \boldsymbol{t})$, with rotation $\boldsymbol{R} \in SO(3)$, $\boldsymbol{t} \in \mathbb{R}^3$, and $K \in \mathbb{R}$. Let's denote the i^{th} row of the rotation matrix $\boldsymbol{R}$ by $\boldsymbol{r}_i$. Then we substitute $\boldsymbol{X}_c$ in Eq.8 by the rigid motion to get an $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ form:

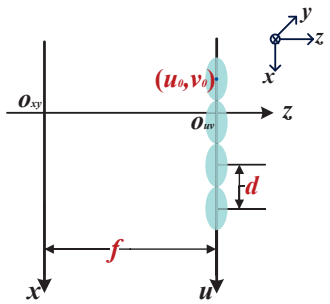

Fig. 4. A TPP describes a light field camera. x-o_{xy}-y is the image coordinate and u-o_{uv}-v is the micro-lens array plane whose origin lies on the z-axis.

$$\begin{bmatrix} d'X_d & d'Y_d & d'Z_d \\ (X_d-u_0')\boldsymbol{X}_w & (Y_d-v_0')\boldsymbol{X}_w & (Z_d-f')\boldsymbol{X}_w \\ -d'X_d\tilde{\boldsymbol{X}}_{\boldsymbol{w}} & -d'Y_d\tilde{\boldsymbol{X}}_{\boldsymbol{w}} & -d'Z_d\tilde{\boldsymbol{X}}_{\boldsymbol{w}} \\ -d'\boldsymbol{X}_w & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{X}_w & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -d'\boldsymbol{X}_{\boldsymbol{w}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_{\boldsymbol{w}} & \boldsymbol{0} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} f-Kt_3 \\ dK\,[\boldsymbol{r}_3\ t_3]^{\mathrm{T}} \\ K\boldsymbol{r}_3^{\mathrm{T}} \\ fK\,[\boldsymbol{r}_1\ t_1]^{\mathrm{T}} \\ u_0dK[\boldsymbol{r}_3\ t_3]^{\mathrm{T}} \\ fK\,[\boldsymbol{r}_2\ t_2]^{\mathrm{T}} \\ v_0dK[\boldsymbol{r}_3\ t_3]^{\mathrm{T}} \end{bmatrix} = \boldsymbol{0}, \quad (9)$$

where $\tilde{\boldsymbol{X}}_{\boldsymbol{w}} = (X_w, Y_w, Z_w)^{\mathrm{T}}$ denotes non-homogeneous coordinate.

Stacking all recovered intersections in different settings $\mathscr{X}'$ to the matrix $\boldsymbol{A}$, its right singular vector $\boldsymbol{v}$ corresponding to its smallest singular value is selected as an initial solution multiplied by an unknown scale factor λ ($\boldsymbol{v} = \lambda\boldsymbol{x}$). Let v_i be the i^{th} element of $\boldsymbol{v}$ ($1 \le i \le 24$). Utilizing the identity and orthogonality of $\boldsymbol{R}$, the parameters are recovered as follows:

$$d = \frac{v_2^2 + v_3^2 + v_4^2}{v_6^2 + v_7^2 + v_8^2}, \quad (10)$$

$$u_0 = \frac{v_{13}^2 + v_{14}^2 + v_{15}^2}{v_2^2 + v_3^2 + v_4^2}, v_0 = \frac{v_{21}^2 + v_{22}^2 + v_{23}^2}{v_2^2 + v_3^2 + v_4^2}, \quad (11)$$

$$f = \frac{v_9^2 + v_{10}^2 + v_{11}^2}{v_6^2 + v_7^2 + v_8^2} = \frac{v_{17}^2 + v_{18}^2 + v_{19}^2}{v_6^2 + v_7^2 + v_8^2}. \quad (12)$$

Moreover, Eq.9 indicates that the 4 parameters are sufficient to constrain a rectified 4D light field where the rays intersect at the undistorted scene points. Once there are enough prior scene points $\boldsymbol{X}_w$ and parameters settings $\boldsymbol{X}'$, the solution for $\mathscr{X}$ is unique.

3.2. Non-linear Optimization

Moreover, with the initialization solved in sec 3.1, we minimize the cost function $E(\cdot)$ on the recovered intersections $\hat{\boldsymbol{X}}_c$ dependent on $\mathscr{X}$ to refine the solution in Eqs.10, 11 and12:

$$E(\mathscr{X}) = \sum_{j=1}^{4}\sum_{i=1}^{n_j} |\sin(\vec{\boldsymbol{q}}_j, \widehat{\vec{\boldsymbol{q}}_{j,i}})| + \sum_{1\le k,l\le 4} |\sin(\vec{\boldsymbol{q}}_k, \widehat{\vec{\boldsymbol{q}}_l}) - \theta_{k,l}|, \quad (13)$$

where $\vec{\boldsymbol{q}}_{j,i} (1 \le i \le n_j)$ are the directions of a set of parallel lines formed by the recovered intersections (by Eqs.2 and 3) with their mean directions $\vec{\boldsymbol{q}}_j = \frac{1}{n_j}\sum_i \vec{\boldsymbol{q}}_{j,i}$. $\theta_{k,l}$ is the prior included angle between $\vec{\boldsymbol{q}}_k$ and $\vec{\boldsymbol{q}}_l$.

With an arbitrary parameters setting, there is a projective transformation between the recovered scene points and ground truth. Specifically, the first term in Eq.13 constrains the affine property of the recovered scene points, and the second term constrains the similarity property [5].

A specific case of the cost function is illustrated in Fig.5. Moreover, the four non-parallel directions are in different planes. To minimize the cost function, the structure of a cube without ambiguity can be constrained which is the same as Fig.2 except for the euclidean property.

Table 1. Parameters estimated with a parallel bi-planar board with different poses on simulated data and the ground truth.

Parameters	Data 1	Data 2	Data 3	GT
$\boldsymbol{R}/deg$	(0, 0, 0)	(0, 0, 0)	(5, 15, 38)	–
$\boldsymbol{t}/mm$	(10,20,1000)	(20,40,900)	(10,20,1000)	–
$u_0/pixel$	-17.0622	-18.0623	-18.0620	-20
$v_0/pixel$	-58.9686	-57.9689	-58.3191	-60
d/mm	39.7498	39.7860	39.7506	39.70
f/mm	1047.3021	1053.1209	1054.1388	1050

4. EXPERIMENTAL RESULTS

To verify the projective distortion model, we apply our rectification method on simulated data and a self-assembly focused plenoptic camera. In addition, we design a parallel bi-planar board with grids (in Fig.5) to provide prior scene points in the world coordinate system. As shown in Fig.7 and 8, the focused images on the left are rendered with the calibrated result, and the images on the right are rendered with the assumption that the center of micro image lies on the optical axis of its corresponding micro-lens.

First, we render the simulated raw images in MATLAB where the micro-lens and the main lens are modelled as thin lens. The results and ground truth with different priors and pose are shown in Tab.1. The value of $E(\cdot)$ is shown in Fig.6, which indicates that the optimization is convergent. Then we shot a resolution test chart and rendered the images focused on the same depth with different $\mathscr{X}$ using the algorithm in [14] (in Fig.7), which indicates that the rays' directions effects the sharpness of rendered images.

Second, we calibrate a physical camera based on focused plenoptic camera model. The rendered images with different $\mathscr{X}$ are shown in Fig.8. In addition, $\mathscr{X}$ influences the geometric structure of the focused images from a 4D light field. It is more visible in the bottom one, where the chess board becomes more uniform after calibrated.

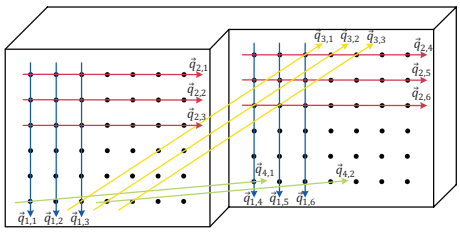

Fig. 5. The parallel bi-planar board we designed to provide prior scene points for calibration. The grids on the two planes are in square layout and $\vec{\boldsymbol{q}}_{1,i}$ is perpendicular to $\vec{\boldsymbol{q}}_{2,j}$. The distance between the two parallel planes and the distance between adjacent grids are known.

5. CONCLUSION

In this paper, we present a transformation matrix to describe the projective distortion on the recovered scene points caused by the uncertainty of parameters of a TPP model. Then we apply the transformation to the light field camera and simplify the camera model as a 4-intrinsic-parameter system, where the refraction of the main lens is substituted by the scale of the micro-lens array plane. Finally, by rectifying the projective distortion of the geometric structure of the scene, the intrinsic parameters of a light field camera can be calibrated. Experimental results show that our model is sufficient to rectify the light field data, proving the uniqueness of the solution for the intrinsic parameters.

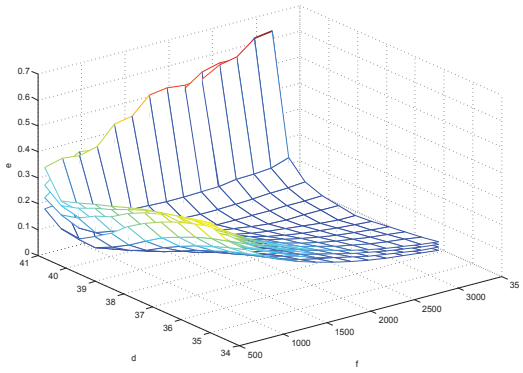

Fig. 6. The error maps from Eq.13. We keep arbitrary (u_0, v_0) constant to plot four error maps about f and d.

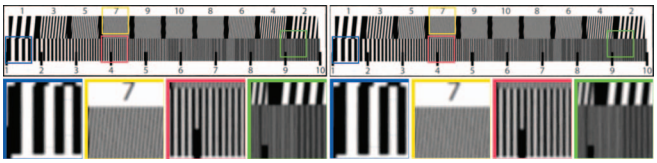

Fig. 7. The rendered images from our simulated data by ray tracing.

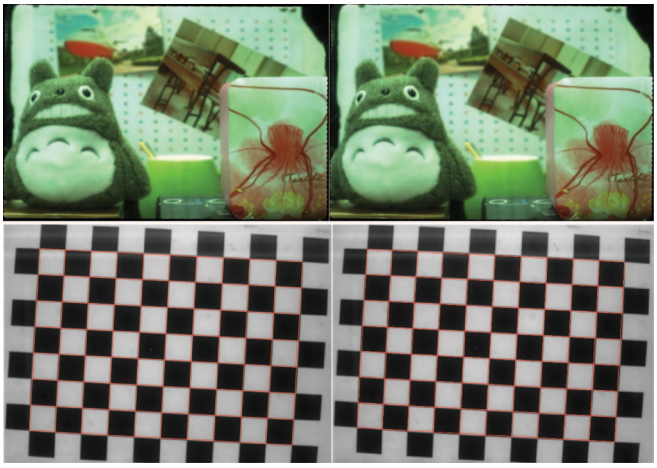

Fig. 8. The rendered images from the physical focused plenoptic camera by ray tracing.

6. REFERENCES

[1] Ren Ng, *Digital light field photography*, Ph.D. thesis, stanford university, 2006.

[2] Andrew Lumsdaine and Todor Georgiev, "The focused plenoptic camera," in *2009 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2009, pp. 1–8.

[3] Donald G Dansereau, Oscar Pizarro, and Stefan B Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013, pp. 1027–1034.

[4] Hae Gon Jeon, Jaesik Park, Gyeongmin Choe, and Jinsun Park, "Accurate depth map estimation from a lenslet light field camera," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, 2015.

[5] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.

[6] Ole Johannsen, Christian Heinze, Bastian Goldluecke, and Christian Perwaß, "On the calibration of focused plenoptic cameras," in *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pp. 302–317. Springer, 2013.

[7] Christopher Hahne, Amar Aggoun, and Vladan Velisavljevic, "The refocusing distance of a standard plenoptic photograph," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2015*. IEEE, 2015, pp. 1–4.

[8] Yunsu Bok, Hae-Gon Jeon, and In So Kweon, "Geometric calibration of micro-lens-based light-field cameras using line features," in *Computer Vision–ECCV 2014*, pp. 47–61. Springer, 2014.

[9] Marc Levoy and Pat Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.

[10] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen, "The lumigraph," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 43–54.

[11] Robert Pless, "Using many cameras as one," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. IEEE, 2003, vol. 2, pp. II–587.

[12] Ole Johannsen, Antonin Sulc, and Bastian Goldluecke, "On linear structure from motion for light field cameras," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 720–728.

[13] Christopher Hahne, Amar Aggoun, Shyqyri Haxha, Vladan Velisavljevic, and Juan Carlos Jácome Fernández, "Light field geometry of a standard plenoptic camera," *Optics express*, vol. 22, no. 22, pp. 26659–26673, 2014.

[14] Christian Perwass and Lennart Wietzke, "Single lens 3d-camera with extended depth-of-field," in *Proc. SPIE*, 2012, vol. 8291, p. 829108.