Laser & Optoelectronics Progress

Structured-Light-Field 3D Imaging System With Coaxial Projection

Zhang Zhijun[†], Wu Qingyang[*†], Jiang Yifan, Deng Yifeng

College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, Guangdong, China

Abstract　This paper presents a novel 3D measurement method for a light field camera (LFC) in which 3D information of object space is encoded by a microlens array (MLA). The light ray corresponding to each pixel of the LFC is calibrated. Once the matching points from at least two subviews exhibit sub-pixel accuracy, the 3D coordinates can be calculated optimally by intersecting light rays of these points matched through phase coding. Moreover, the proposed method obtains high-resolved results that exceed the subview resolution due to the virtual continuous phase search strategy. Finally, we combine the LFC and coaxial projection to solve the 3D data loss caused by shadowing and occlusion problems. Experimental results verify the feasibility of the proposed method, and the measurement error is about 30 μm in a depth range of 60 mm.

Key words　light field; 3D imaging; microlens array; structured light

1　Introduction

Equipped with a regular camera that can capture only the intensity of light rays, conventional surface reconstruction technologies must utilize either a structured light to build phase-height mapping or another camera to convert depth information into disparity. With the development of microlens array (MLA) manufacturing technology, the light field camera (LFC) system[1-2] composed of MLA and camera sensor has been widely used in academia and industries to robustly and implicitly capture 3D information with a single shot, thus solving the occlusion reflection problem. MLA acts similar to multiple small cameras and simultaneously captures reflectance properties from slightly different angles. As a result, 3D information on spatial points can be encoded by multiple views. The missing surface data of the measured object in one view due to occlusion reflection can be captured by other views. In conventional structured-light projection methods[3], shadowing and occlusion cause the loss of the phase map, leading to the loss of 3D data of the complex surfaces. The LFC provides an effective way to solve these problems.

Existing light field reconstruction approaches can be mainly categorized into depth-from-defocus[4-9] theory, multiview disparity[10-17], and phase mapping model[18-19]. In the depth-from-defocus theory, several images focusing at different depths are captured and scene depth can then be inferred by the analysis of the focal stack. Lin et al.[6] described a technique to retrieve depth information using two features of the light field focal stack. In the multiview disparity method, scene depth can be estimated from the disparity map calculated by at least two matching points from different viewpoints in the LFC system. However, passive techniques can only estimate depth and unfortunately are not robust enough in scenes that lack features, occlusion reflection, and repeating textures. Cai et al.[17] analyzed the angular variance by the sinusoidal distribution of the radiance in the structured-light field to achieve a single shot light field reconstruction. However, this method can only estimate depth. In addition to compensating the plenoptic imaging distortion, an auxiliary 3D measurement system must be calibrated in advance, which inevitably brings the cumulative error. For the phase mapping model, Cai et al.[19] proposed a method of ray calibration and phase mapping to achieve

structured-light-field 3D reconstruction that performed metric 3D reconstruction by ray calibration and derived the phase mapping in the structured-light-field so that 3D coordinates can be directly mapped from the phase. However, the method does not implement coaxial illumination and the relative position of the projector to the LFC system must remain fixed after calibration.

Light field 3D measurement obtains the mapping relationship between the depth value and the metric 3D coordinate and generally includes two steps: light field depth estimation and light field calibration. Most of the existing methods establish a complex mathematical model for the LFC system so that the depth value is converted into a metric 3D coordinate. However, the LFC system cannot be accurately described by light field calibration even with complex mathematical models due to the complex distortion and low resolution in each subview. In this paper, we describe a novel method to accurately achieve light field 3D measurements. We calibrate the ray equation for each pixel of the LFC instead of analyzing the LFC structure with a complicated mathematical model. The 3D coordinates are determined directly from the ray intersection of these matching points. The proposed method performs metric 3D reconstruction without using depth estimation algorithms and additional metric calibration steps[20-22].

2　Methods

2.1　Ray calibration

The perspective projection model is suitable for describing the conventional camera imaging system but not the LFC system because of the complex distortion brought about by the combined lens system composed of a single microlens and the camera's main lens and the low resolution in each small view. In addition, the low resolution could lead to fuzzy features, which hinder the accurate extraction of calibration pattern features, such as the corner points of high-precision 3D or 2D targets. As a result, traditional camera calibration methods based on internal parameters and external parameters cannot accurately describe all the rays recorded by the camera and the relative positions of perspective projection coordinate systems from different views. However, the ray calibration for each pixel is slightly affected by complicated distortion. A fixed pixel on the image plane can record a series of points that make up a straight ray in space as shown in Fig. 1.

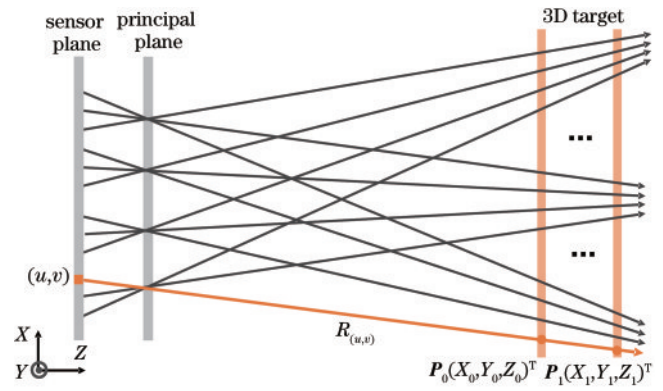We can use the following two equations (1)−(2)[19]

Fig. 1　The LFC system records the light ray

to describe the rays that pass through the image plane and recorded by pixel. In this way, the relation of 3D coordinates from these collinear spatial points is established.

$$X = a_{(u,v)}Z + b_{(u,v)}, \qquad (1)$$
$$Y = c_{(u,v)}Z + d_{(u,v)}, \qquad (2)$$

where a spatial point lying on the light ray, denoted as $\boldsymbol{P}(X, Y, Z)^{\mathrm{T}}$, is projected to (u, v) on the pixel plane, and parameters $(a_{(u,v)}, b_{(u,v)}, c_{(u,v)}, d_{(u,v)})$ involve ray position and direction. As long as the spatial positions of two points on the ray are known, the ray equation can be computed uniquely. The ray represented as $R_{(u,v)}$ is determined by $\boldsymbol{P}_0(X_0, Y_0, Z_0)^{\mathrm{T}}$ and $\boldsymbol{P}_1(X_1, Y_1, Z_1)^{\mathrm{T}}$, as the spatial coordinate of $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ can be obtained by moving the 3D target. Ray calibration is a simple way to accurately describe ray position and direction in the LFC system without establishing a complex mathematical model and could satisfy the global optimum. In addition, the effect of plenoptic imaging distortion can be reduced. Given that only two spatial points are enough to achieve ray calibration, this approach is susceptible to noise and outliers. The use of additional spatial points in ray calibration can guarantee strong robustness.

2.2　Ray intersection

In the LFC system, a spatial point can be recorded by the 4D light field in a 2D image sensor. When an in-focus object point is imaged in an LFC system, light rays emitted from the object point reach the image sensor plane through different sub-apertures (i. e., 4D light field light can collect light rays of different directions from the same point). The sub-aperture and sensor planes are represented as the angular (s, t) and spatial planes (u, v), respectively. Therefore, the 4D light field records the direction information of rays emitted from the target point. According to the reversibility of light rays, the 3D coordinate of the

target point can be obtained by intersecting the ray equation of these matching points once multiple matching points on the pixel plane are determined as shown in Fig. 2 (a). Considering that the ray is calibrated on pixel-level and the matching point can be sub-pixel coordinate, we need to obtain the ray equation at sub-pixel level. As illustrated in Fig. 2 (b), four nearest pixel-level rays obtained by ray calibration can be found around the known sub-pixel coordinate (u, v). These rays intersect a normalized plane that parallels the camera plane at four points, forming a quadrangle.

The ray corresponding to the (u, v) also intersects at a point (x, y) inside the quadrangle. Therefore, the position of (x, y) can be computed by interpolation. The resolution of the camera sensor is high enough to improve the sampling rate of adjacent pixels. Hence, most interpolation technologies show similar performance. From the efficiency perspective, bilinear interpolation used in the sub-pixel ray equation calculated is more appropriate than other advanced interpolations. As a consequence, the ray equation of (u, v) is determined accurately.
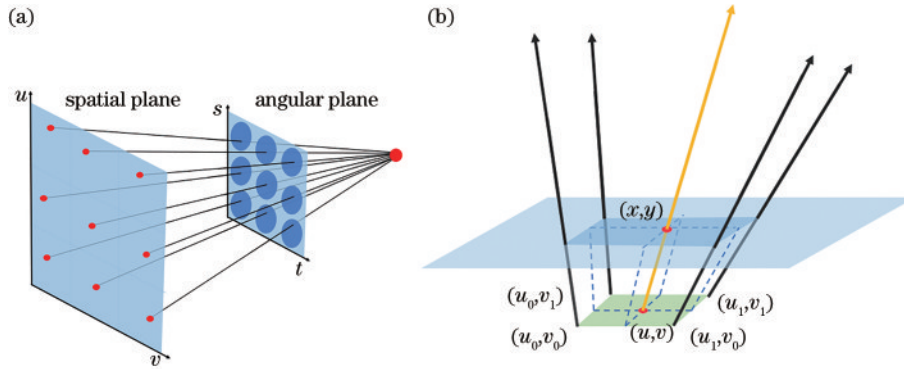

Fig. 2   Ray model of light field. (a) Spatial point determined by ray intersection; (b) schematic of sub-pixel ray equation

In theory, two rays are enough to work out the 3D coordinates of a target point. In practice, additional rays can be used to obtain accurate 3D coordinates for high-precision and robust reconstruction. Owing to the influence of measurement error, these rays will not intersect strictly. However, a spatial point is generally mutually closest to two or more rays in a least-squares sense. Therefore, the intersection point $\boldsymbol{P}_\mathrm{W}$ of a set of rays $\Re$ can be described as

$$\boldsymbol{P}_\mathrm{W} = \arg\min_{\boldsymbol{P}_\mathrm{W}} \sum_{R_{(u,v)} \in \Re} \left\| d_{R_{(u,v)}}(\boldsymbol{P}_\mathrm{W}, R_{(u,v)}) \right\|^2, \quad (3)$$

where $d_{R_{(u,v)}}$ is the distance of $\boldsymbol{P}_\mathrm{W}$ to $R_{(u,v)}$. Hence, the accurate 3D coordinates corresponding to the minimum computed value can be obtained. The proposed method allows the use of an appropriate algorithm strategy to select the matching points with high confidence to reduce the influence of noise and outliers. It also reduces the influence of noise and outliers by iteratively deleting the rays farthest from the intersection point $\boldsymbol{P}_\mathrm{W}$ during implementation. When the distance of the sets of rays from the intersection point $\boldsymbol{P}_\mathrm{W}$ is less than a certain threshold, the iteration is stopped to obtain the final intersection point $\boldsymbol{P}_\mathrm{W}$.

2.3   Virtual continuous phase search

Most existing methods generate results with the same resolution as the single subview and only reconstruct the part of the surfaces from a subview. The visibility of the measured object in different subviews is inconsistent as shown in Fig. 3. Surface AB can be reconstructed in the left subviews but not in the right subviews. To reconstruct the measured object surface with high-resolution and without missing data, effective data must be used in all subviews for 3D reconstruction. We adopt the following strategy called virtual continuous phase search. First, the phase distribution range is obtained from the absolute phase map, and the orthogonal absolute phases are then uniformly sampled from a given sampling rate as required. A phase reference map is now established. The resolution of the phase reference map is related to the number of reconstructed point clouds. Second, rather than using a single subview as a template to search matching points in other subviews, these matching points based on the phase reference map are calculated by the interpolation technique in all subviews. This process is different from the existing active light field reconstruction method. Finally, 3D coordinates can be computed by Eq. (3) with these ray equations of matching points. The virtual phase continuous searching strategy performs sampling and sub-pixel interpolation in all subviews to utilize valid information in each subview for reconstructing the
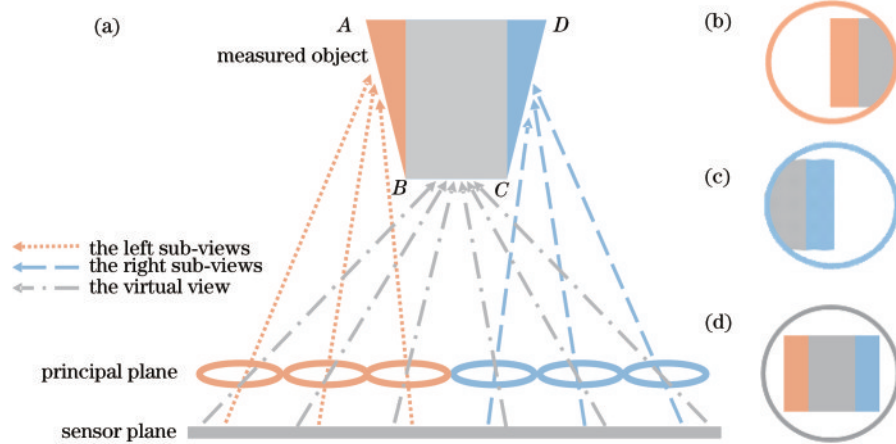
Fig. 3 Schematic of virtual continuous phase search. (a) Schematic of the sampling in different subview; (b) visibility of the left subview; (c) visibility of the right subview; (d) visibility using virtual continuous phase search

measured object. Therefore, surface $ABCD$ in Fig. 3(c) does not correspond to either the left subview or the right subview but is actually a virtual perspective that contains visible information for each subview. Therefore, the result contains a larger amount of effective 3D data than a subview.

3 Experiments

The experimental equipment of our LFC system shown in Fig. 4 consists of a camera (acA4112-30 μm, 3000×4096) from Basler, MLA (70 mm×70 mm, 19×17 apertures) and a 75 mm CHIOPT's lens FA7501C. An AOSIMAN screen (3840×2160) and a precise translation stage are necessary for calibrating the LFC system. A projector (DLP 4500) was used to add features to the object, and a semireflective membrane was applied to reflect the projecting beam. We also assembled a 3D target composed of a display screen and a precision translation stage. The display screen manufactured by large-scale integrated circuits and lithographic techniques has a high-resolution, and the size of pixel units is uniform and known.
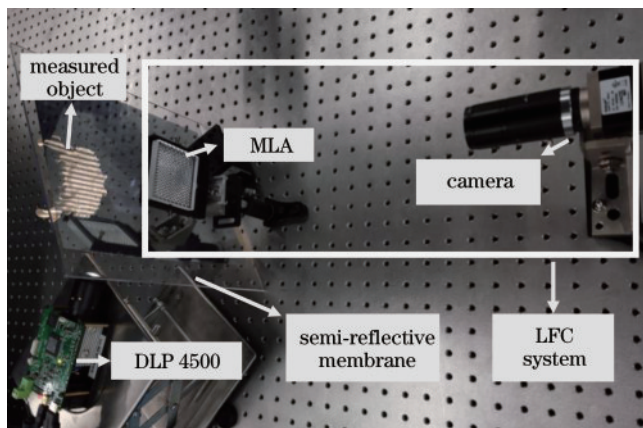

Fig. 4 Experimental system architecture

When the sinusoidal fringe patterns are displayed on the AOSIMAN screen, X and Y information on the display screen plane could be modulated into intersecting phase information using fringe analysis. Furthermore, the precision translation stage provides Z information on each plane. Compared with traditional 3D targets, this kind is not affected by the blurred calibration pattern and is convenient to realize the camera's pixel-based calibration. The screen is placed on the precise translating stage, adjusted to be perpendicular to its moving track, and precisely moved at different positions. The scale value, which is Z coordinates for every calibration plane, is simultaneously recorded. Moreover, the screen displays horizontal and vertical sinusoidal fringe patterns, which convert the spatial x and y coordinates to intersecting phase information. The metric 3D coordinates of these space points are then obtained. However, the measurement process may be influenced by phase errors and ray calibration errors. For phase errors, we used root-mean-square-error (RMSE) to evaluate the deviation between the actual observed absolute phase value and the ideal absolute phase value. As shown in Fig. 5(a), a set of data at pixel coordinates $u = 1400$ were selected for comparison. A RMSE value of 0.0471 was obtained, demonstrating minimal deviation between the two sets of data. For the ray calibration error, one set of spatial points recorded at pixel coordinates $(u, v) = (1400, 2000)$ was taken as an example. A maximal value of 0.0177 mm and a root-mean-square value of 0.0109 mm of the fitting error were derived illustrated in Fig. 5(b).

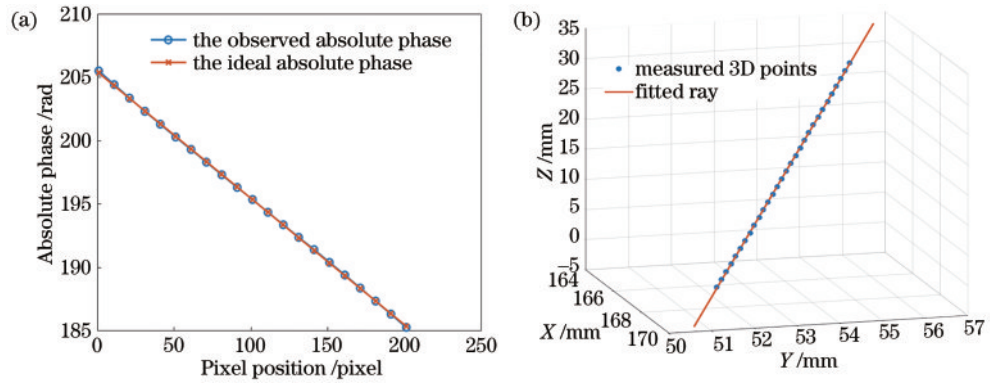In the experiment, the screen was moved to 30 positions with a 2 mm interval in a measurement

Fig. 5　Error analysis of phase and ray calibration.（a）Actual observed absolute phase and corresponding ideal absolute phase；
（b）one set of the recorded 3D points along with the corresponding fitted ray

volume of 90 mm×110 mm ×60 mm. Eventually，all the light rays recorded by the light field camera were calibrated and described by ray equations. The calibrated system was then applied to measure the metric scene for 3D imaging.

According to the reconstruction principle of the presented method，the orthogonal absolute phase is used only for feature marking. Therefore，the position of the projector relative to the camera is flexible. To prevent missing data caused by shadows，we adjusted the position of the semipermeable membrane to make the projected ray illuminate and camera coaxial. We chose a wooden carving model for the experiment to evaluate the system's performance on a complex surface. As shown in Figs. 6（a）–（c），the results from different angles exhibit some tiny features，such as sunken nostrils and eyebrows，that could be entirely recovered with details. Even the sharp edges can be reconstructed. As shown in Fig. 6（d），only a part of the measured object is captured with low resolution in the original image because of the limited imaging range of a single sub-aperture. However，the proposed method can reconstruct the complete high-resolution object at one time without stitching. This finding illustrated the advantages of the proposed method.
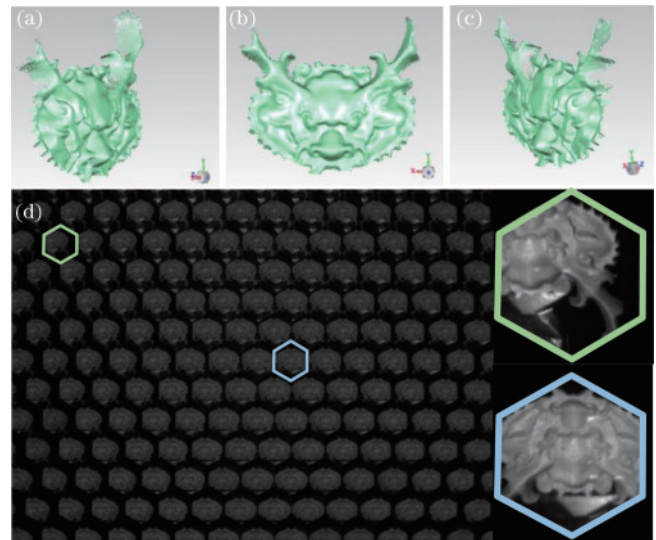
To further illuminate the performance of the

Fig. 6　Experimental scene 1. （a）Left-view point cloud；
（b）front-view point cloud；（c）right-view point cloud；
（d）original image

proposed method，we selected a dental impression with evident occlusion and shadows as the measured object as shown in Fig. 7（a）. The traditional phase-height mapping method[23] was used for comparison. As shown in Fig. 7（b），a large number of point cloud loss appeared in the measurement results，and the 3D shape of the teeth was hardly recovered. The measurement results of the proposed method are shown in Fig. 7（c），the 3D shape of the tooth was generally recovered.

Fig. 7　Experimental scene 2. （a）Measured object of dental impression；（b）result of phase-height mapping method；
（c）result of the proposed method

Regardless of the tip or the edge of the tooth, the point cloud data can be obtained. This finding verified that the proposed method can solve the occlusion problem.

We also measured a pair of standard spheres made from ceramic for the quantitative evaluation of 3D reconstruction accuracy. The shape of the standard spheres was calibrated using a coordinate measurement machine. The two tested spheres whose diameters are 38.0946 and 38.0950 mm are shown in Fig. 8(a). We fitted the diameter of the tested spheres and analyzed the error range. The diameters of the reconstructed spheres are 38.1261 and 38.1257 mm, with deviations of 0.0315 and 0.0307 mm, respectively. As shown in Fig. 8(b), the standard deviation of the spheres are 0.0295 and 0.0215 mm. The experimental results demonstrated that the proposed method could achieve high-accuracy 3D measurements in an LFC system.
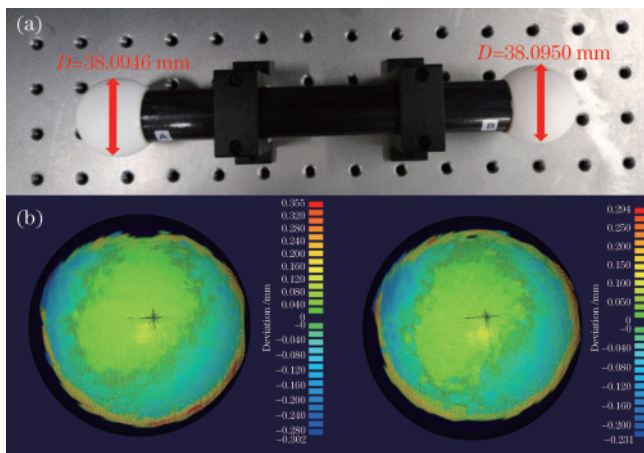

Fig. 8   Experimental scene 3. (a) Measured spheres; (b) point cloud deviation map

4   Conclusion

The proposed method realized light field 3D high-resolved reconstruction with high-accuracy, a problem that has been unsolved due to the complexity of the LFC system. Once the matching points on the image plane from at least views are determined, the 3D coordinates of the spatial points can be calculated directly. Several superiorities of our calibration are listed as follows. First, the light field reconstruction process is independent of the projector, which is not used in this method to project patterns to modulate scene depth but to add features. Hence, the technique is flexible and suitable for various applications. Second, given that all views from the MLA are calibrated, the actual position of a spatial point can be determined from all the high confidence matching points with sub-pixel

accuracy from multiple views by taking advantage of projected intersecting sinusoidal stripes, thus accomplishing high-accuracy and high-resolution reconstruction. Third, the combination of LFC and coaxial projection solves the problems of occlusion reflection and shadows that lead to the lack of 3D information in traditional reconstruction systems. Hence, complete reconstruction is achieved because the matching points can always be found in another microlens. Finally, multiple views from the MLA are calibrated based on the same coordinate system. Therefore, the sub-aperture point clouds can be fused without utilizing any stitching algorithm and detailed information of the object is restored in the virtual continuous phase search strategy.

References

[1] Adelson E H, Bergen J R. The plenoptic function and the elements of early vision[M]//Computational models of visual processing. Cambridge: The MIT Press, 1991: 3-20.

[2] Nie Y F, Xiangli B, Zhou Z L. Advances in light field photography technique[J]. Journal of the Graduate School of the Chinese Academy of Sciences, 2011, 28(5): 563-572.

[3] Gorthi S S, Rastogi P. Fringe projection techniques: whither we are? [J]. Optics and Lasers in Engineering, 2010, 48(2): 133-140.

[4] Bishop T E, Favaro P. The light field camera: extended depth of field, aliasing, and superresolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(5): 972-986.

[5] Hahne C, Aggoun A, Velisavljevic V, et al. Refocusing distance of a standard plenoptic camera[J]. Optics Express, 2016, 24(19): 21521-21540.

[6] Lin H T, Chen C, Kang S B, et al. Depth recovery from light field using focal stack symmetry[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2016: 3451-3459.

[7] Nava F P, Luke J P. Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera[C]//2009 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, May 4-6, 2009, Potsdam, Germany. New York: IEEE Press, 2009.

[8] Tao M W, Hadap S, Malik J, et al. Depth from combining defocus and correspondence using light-field cameras[C]//2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2014: 673-680.

[9] Tao M W, Srinivasan P P, Hadap S, et al. Shape estimation from shading, defocus, and correspondence using light-field angular coherence[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(3): 546-560.

[10] Wanner S, Goldluecke B. Globally consistent depth labeling of 4D light fields[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 41-48.

[11] Kim C, Zimmer H, Pritch Y, et al. Scene reconstruction from high spatio-angular resolution light fields[J]. ACM Transactions on Graphics, 2013, 32(4): 1-12.

[12] Wanner S, Goldluecke B. Variational light field analysis for disparity estimation and super-resolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3): 606-619.

[13] Zhang S, Sheng H, Li C, et al. Robust depth estimation for light field via spinning parallelogram operator[J]. Computer Vision and Image Understanding, 2016, 145: 148-159.

[14] Zhang Y B, Lv H J, Liu Y B, et al. Light-field depth estimation via epipolar plane image analysis and locally linear embedding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(4): 739-747.

[15] Chen Y Q, Jin X, Dai Q H. Distance measurement based on light field geometry and ray tracing[J]. Optics Express, 2017, 25(1): 59-76.

[16] Jeon H G, Park J, Choe G, et al. Accurate depth map estimation from a lenslet light field camera[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1547-1555.

[17] Cai Z W, Pedrini G, Osten W, et al. Single-shot structured-light-field three-dimensional imaging[J]. Optics Letters, 2020, 45(12): 3256-3259.

[18] Cai Z W, Liu X L, Peng X, et al. Structured light field 3D imaging[J]. Optics Express, 2016, 24(18): 20324-20334.

[19] Cai Z W, Liu X L, Peng X, et al. Ray calibration and phase mapping for structured-light-field 3D reconstruction [J]. Optics Express, 2018, 26(6): 7598-7613.

[20] Su L J, Yan Q Q, Cao J, et al. Calibrating the orientation between a microlens array and a sensor based on projective geometry[J]. Optics and Lasers in Engineering, 2016, 82: 22-27.

[21] Cai Z W, Liu X L, Tang Q J, et al. Light field 3D measurement using unfocused plenoptic cameras[J]. Optics Letters, 2018, 43(15): 3746-3749.

[22] Zhou P, Cai W J, Yu Y L, et al. A two-step calibration method of lenslet-based light field cameras[J]. Optics and Lasers in Engineering, 2019, 115: 190-196.

[23] Li Y, Su X Y, Wu Q Y. Accurate phase-height mapping algorithm for PMP[J]. Journal of Modern Optics, 2006, 53(14): 1955-1964.