IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 5, 2019, accepted November 18, 2019, date of publication November 21, 2019,
date of current version December 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2954892

Accurate Light Field Depth Estimation Using Multi-Orientation Partial Angular Coherence

ZHENGHUA GUO[1,2,3], JUNLONG WU[1,2,3], XIANFENG CHEN[1,2,3], SHUAI MA[1,2,3], LICHENG ZHU[1,2,3], PING YANG[1,2], AND BING XU[1,2]

[1]Key Laboratory on Adaptive Optics, Chinese Academy of Sciences, Chengdu 610209, China
[2]Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China
[3]University of Chinese Academy of Sciences, Beijing 100039, China

Corresponding author: Bing Xu (bing_xu_ioe@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61875203 and Grant 61805251, in part the Young Scientists Fund of the National Natural Science Foundation of China under Grant 11704382, and in part by the Funds for International Cooperation and Exchange of the National Natural Science Foundation of China under Grant 1171101412.

ABSTRACT Occlusion is a critical issue that affects the accuracy for light field depth estimation. In the presence of occlusions, the photo-consistency property is broken, making ambiguity near occlusion areas. In this paper, we proposed using multi-orientation partial angular coherence to achieve accurate depth estimation by which the occlusions are explicitly treated. Unlike previous approaches, the correspondence is partially measured along several lines with different directions. We show that depending on the occlusion edge orientation, the pixels along the occlusion orientation still preserve photo-consistency. By computing the partial angular coherence in four potential directions, we can obtain a sharp initial depth map. Since occlusions are precisely tackled, the outliers can be mostly removed by fast guided filtering on the cost volume. As a result, the depth accuracy at occlusion boundaries is greatly improved. The proposed method can obtain sharp transition edges at occlusion boundaries and has no requirement for extra edge information. Experimental results on a recent benchmark demonstrate that the proposed method outperforms the state-of-the-art algorithms in the accuracy metrics. Further experiments on real-world scenes also show the superior performance of the proposed method.

INDEX TERMS Depth estimation, light field, angular coherence, occlusion handling.

I. INTRODUCTION

Light field (LF) imaging combines optics with computation to provide more attractive scene capture and analysis than conventional imaging. For a light field camera, the sensor records not only the intensity of the rays but also the directions. Therefore, angular information is maintained in LF imaging [1]. To achieve this, the special optical imaging system is required to identify the light rays' direction, such as multi-view camera systems [2], micro-lens based plenoptic cameras [3], [4] and aperture-coded plenoptic cameras [5]. The precise combination of the novel imaging system and algorithms enables improved scene analysis, e.g., post refocusing [6], saliency detection [7] and depth estimation [8].

In light field imaging, accurate depth estimation is a hot and challenging topic, and it is important for further applications such as 3D reconstruction and ligh field editing. In Ref. [9], a taxonomy of dense light field depth estimation algorithms is presented. The state-of-art algorithms [10]–[15] and many more cited in the references can provide acceptable results. According to the LF representations they rely upon, these algorithms can be classified into several categories. Frequently-used representations include sub-aperture views, epipolar plane images (EPIs), focal stacks and angular patches. Depth from sub-aperture views is an ancient technique, and it was recognized as multi-view stereo [16] in the early days. Because it requires matching the same image elements in several views, the amount of computation is huge. EPI is a novel approach to represent depth information as line features [17], which re-arrange the multi-view images according to their view positions. Several methods transform the problem of estimating depth into detecting the slopes of the lines, thus computation can be greatly reduced compared with multi-view stereo matching. Depth from the focal stack

The associate editor coordinating the review of this manuscript and approving it for publication was Alexandros Iosifidis.

is also recognized as depth from defocus/focus (DfD/DfF), and depth can be inferred based on the in-focus measurement. The angular patch will have a constant value for a point when focused at the correct depth, and depth can be estimated based on minimizing angular patch variance. However, these traditional methods have problems around occlusion regions, resulting in outliers and fuzzy transitions at occlusion boundaries.

Removing the influence of occlusions on the light field depth estimation is a tough problem. Since the photo-consistency assumption no longer holds in the presence of occlusions, the rays from an object will be partially blocked from the light field camera when an occlusion occurs. As a result, the occluded points are only visible in several sub-aperture views. Consequently, the methods based on the photo-consistency assumption will fail at occluded pixels, causing smooth transitions and outliers around occlusion boundaries. There have been several methods of handling occlusions. Chen *et al.* [18] proposed a bilateral metric on angular patches to indicate the probability of occlusions. Wang *et al.* [13] found that although pixels at occlusion boundaries do not preserve photo-consistency in general, they are still consistent in a subset of regions. Moreover, the occlusion edge has the same orientation as that of the line separating the two regions. Strecke *et al.* [8] used the partial focal stack symmetry to handle occlusions, and proposed a refinement method using joint regularization of depth and normals.

In this paper, we proposed a new method to deal with occlusions. The proposed method is based on partial angular coherence and is free of being affected by occlusions. We compute the coherence with only a line subset of the angular patches in multiple orientations. Assumption is made that if the occlusion is present, it occurs only in one direction. This is true when the occlusion boundary is a straight line or the baseline is small enough. For light field cameras and real scenes, the assumption is reasonable. Under such the assumption, the partial angular patch will have the minimum cost when the direction is parallel to that of the occlusions. By comparing the partial coherence in several directions, we can tackle problems around occlusions. The proposed method is insensitive to occlusions and has no requirement for the occlusion information. After the initial depth estimation, further refinement is applied by a simple guided filtering process on the cost volume.

II. OCCLUSIONS AND PARTIAL ANGULAR COHERENCE

Angular coherence will play an important role in our method, so we discuss the traditional angular coherence and explain why occlusions can be handled by the proposed partial coherence.

A. ANGULAR COHERENCE

Firstly, we introduce our notations. The light field is parameterized by two-plane parameterization (2PP), as shown in Fig.1(a). In 2PP, the positions and directions of a ray are recorded by its intersections with two parallel planes whose separation is D. Thus a light field $L_F(u, v, s, t)$ is parameterized by the (u,v) and (s,t) planes. Therefore, (s,t) are the spatial coordinates and (u,v) are the angular coordinates. For different light field cameras, the two parameterization planes are distinct. As shown in Fig.1(b), for the micro-lens based light field cameras, the u-v plane is on the exit pupil plane of the main lens or on the sensor, because the two planes are conjugate. As shown in Fig.1(c), for the camera arrays, the positions of each viewpoints indicate the (u,v) position. It can be seen that the two parameterization planes for the micro-lens based light field cameras are located in the image side, and the planes for camera-array based light field cameras are in the object side. The angular patch can be obtained from the 4D light field by fixing the (s,t) coordinates. The values of the pixels in the angular patch denote the radiance distribution of the spatial point in different directions.

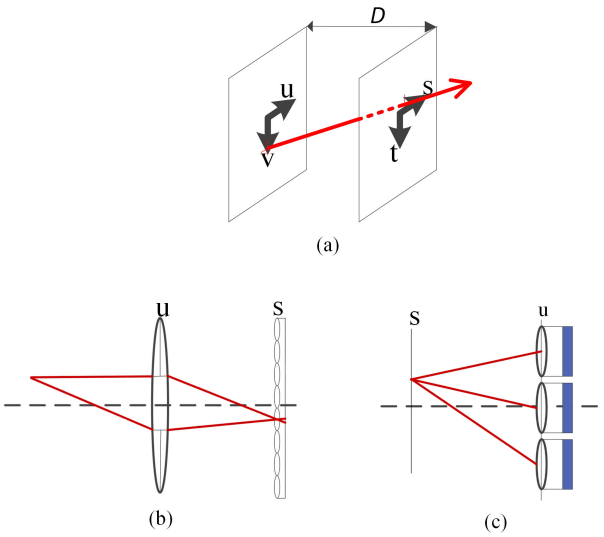

FIGURE 1. Illustration of parameterization of the light field camera. (a) 2PP parameterization. (b) The two parameterization planes for the micro-lens based light field camera are both in the image side. (c) The two parameterization planes for the camera-array based light field camera are both in the object side.

For an input light field L_0, we can shear the light field to propagate it to a new (s,t) plane which is αD from the u-v plane [3], where α is the ratio of the initial separation to the new separation. The shearing process is as follows:

$$L_\alpha(u, v, s, t) = L_0(u, v, s^f(\alpha), t^f(\alpha)), \tag{1}$$

where $s^f(\alpha) = s + u(1 - \frac{1}{\alpha})$ and $t^f(\alpha) = t + v(1 - \frac{1}{\alpha})$. Given the ground-truth depth $\alpha_{gt}(s, t)$ of every spatial point, we can refocus each spatial pixel to its corresponding depth as

$$L_{\alpha_{gt}}(u, v, s, t) = L_0(u, v, s^f(\alpha_{gt}(s, t)), t^f(\alpha_{gt}(s, t))). \tag{2}$$

Under Lambertian assumption, when a point is refocused to the correct depth, the pixels of the angular patch will exhibit angular coherence. In other words, for non-occluded spatial pixels, the values of these pixels in the angular patch are

almost the same since all the rays come from a single (Lambertian) spatial point in the scene, just as depicted in Fig.1(b) and (c). The phenomenon is called angular coherence. There are two forms to evaluate the angular coherence. The first one is done by calculating the variance among the angular pixels as

$$\sigma^2 = \frac{\sum_{u,v} (L(u, v, s, t) - \overline{L}(u, v, s, t))}{N_u N_v}, \tag{3}$$

where $\overline{L}$ is the mean value of all the angular pixels, and (N_u, N_v) are the widths of the u-v planes. This metric is widely used in multi-view stereo. The other one is based on a unique property of the central view where the angular coordinates are at $(u, v) = (0, 0)$. According to Equation (1), the shearing amount for the central view is independent of α. At every α,

$$L_\alpha(0, 0, s, t) = I^c(s, t), \tag{4}$$

where I^c is the central view of the light field. That is, regardless of the focus, the camera at the central angular coordinate always images the same spatial point. According to Equation (4), the angular coherence can also be calculated by

$$\sigma^2 = \frac{\sum_{u,v} (L(u, v, s, t) - I^c(s, t))}{N_u N_v}. \tag{5}$$

Generally, the two metrics are in alignment with each other. But confusion may occur in textureless regions for the first metric. Even at incorrect depth, the angular pixels may exhibit low variance. However, it is hard for the second metric to deal with a noisy light field, especially for the light fields in which the central view has low signal-to-noise-ratio (SNR).

B. LIGHT FIELD OCCLUSION MODELING

Occlusion is an unavoidable part of light field imaging, and it is one of the main causes of errors in light field depth estimation. We firstly analyze the impact of occlusions on the angular patches based on the physical image formation. We assume that the pixel is occluded by only one occluder in one direction. This is true when occluder boundaries are straight or when the equivalent baseline is small enough, ensuring that if the occlusion occurs, it separates the angular patch into two parts with a straight line. In addition, the occlusion edge can also be approximated by a line since the spatial patch the light field camera looks at is small. The assumption is quite reasonable in real-world scenes and for current light field cameras, and has been adopted by Wang *et al.* [19]. However, in our method, this assumption is not strict and will be described in the next section in detail.

As in Fig.2, the rays from point A is partially blocked by the yellow occluder. For a proof, we use more variables as shown in Fig.2, where γ is the slope of the occlusion edge and h is the height of the occluder. The normal of the blue plane in Fig.2 is

$$\begin{aligned}\mathbf{n} &= \mathbf{n_1} \times \mathbf{n_2} \\ &= (-\gamma(D - d_1), D - d_1, r(x_0 - x_1) - (y_0 - y_1)).\end{aligned} \tag{6}$$

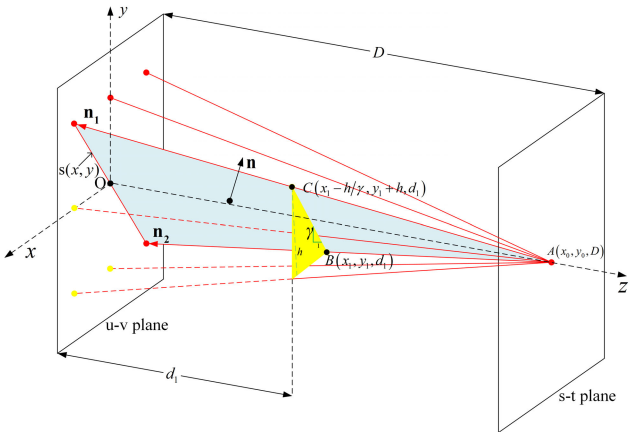

FIGURE 2. The occlusion model and its effects on the captured light field.

The plane equation is

$$\mathrm{P}(x, y, z) = \mathbf{n} \cdot (x - x_0, y - y_0, z - D) = 0. \tag{7}$$

Thus, the slope due to the occlusion in the angular patch is the intersection of the u-v plane and $\mathrm{P}(x, y, z)$:

$$\mathrm{s}(x, y) = y - \gamma x - \frac{y_0 d_1 - y_1 D}{D - d_1} - \frac{\gamma(x_1 D - x_0 d_1)}{D - d_1} = 0. \tag{8}$$

Therefore, the slope of the projection line on the angular patch is γ. This indicates that the occlusion edge in the angular patch has the same orientation as the occlusion edge in the spatial domain.

The intercept of the line is a function of the object position, occluder position and the occluder's slope. For a more intuitive understanding about the intercept, we use 1-D simplification as in Fig.3, and rewrite the intercept part as

$$\begin{aligned}\frac{y_0 d_1 - y_1 D}{D - d_1} &= \frac{y_0 d_1 - y_1 D + y_1 d_1 - y_1 d_1}{D - d_1} \\ &= \frac{d_1(y_0 - y_1)}{D - d_1} + y_1.\end{aligned} \tag{9}$$

With a geometric analysis in Fig.3, Equation (9) is easily derived, and here we do not conduct a detailed derivation. Similar 1-D analysis has been presented by Sheng *et al.* [20]. Another part of the intercept is $\gamma(x_1 D - x_0 d_1)/(D - d_1)$, which is similar with Equation (9), but with a ratio factor of γ. This is true because the intercept is the intersection with y-axis, and when the occlusion edge moves along x-axis, the corresponding change in the intercept will be scaled according to the slope.

We also give an intuitive explanation of the above proof. Consider a simple light field camera with 3×3 angular sampling where the orientation of the occluder is diagonal. The sub-aperture images and the angular patch of this point become what is shown in Fig.4(a) and Fig.4(b). Although the whole angular patch loses photo-consistency, the pixels which are not occluded still holds photo-consistency property. Moreover, the line separating the un-occluded and

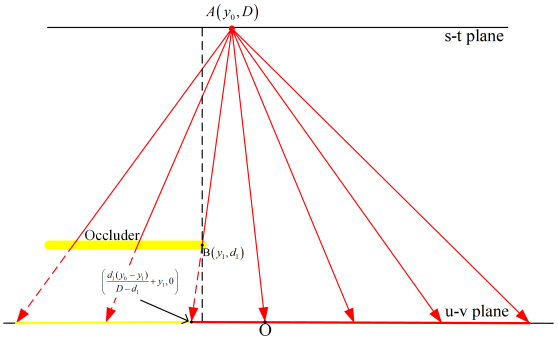

FIGURE 3. 1-D illustration of the occlusion model. Using the similarity relationship, the position of the occlusion edge is easily obtained.

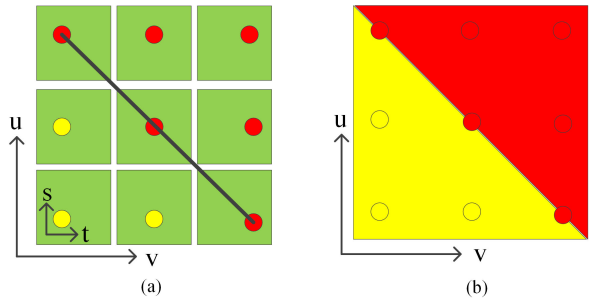

FIGURE 4. The effects on the captured light field due to occlusions. (a) The sub-aperture images of the occluded point. (b) The angular patch of the occluded point.

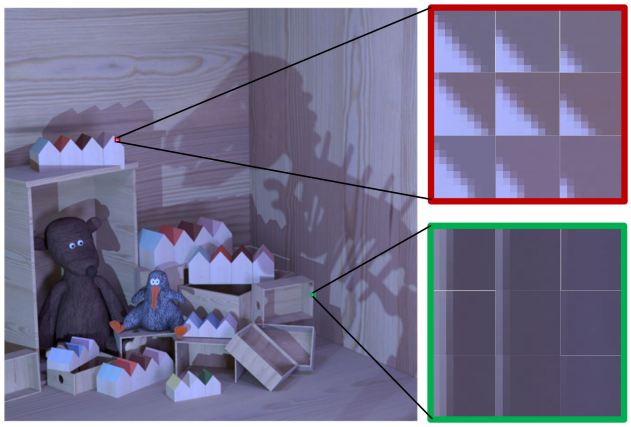

FIGURE 5. Some occlusion cases in real light field images. The two parts at the right side are the angular patches of the spatial points in the two rectangles in the left-side central view, which are colored red and green.

occluded pixels in the angular patch has the same orientation as that of the occlusion edge. We also present some cases in real light field images, as shown in Fig.5.

Although the physical occlusion cases are complex, the angular patch images are simple. As is shown in Fig.5, the occlusion amount varies for different spatial points while the slopes of the occlusion edges remain constant, which is parallel to the occluder edge. For different spatial points occluded by the same occluder, when the spatial coordinates vary, the amount of being occluded changes. The shifting amount of the separating edge can be calculated according to Equation (9). We take the angular patches in the green box for an example. Because occlusion edge is vertical, the occlusion amount changes only in the horizontal direction, while the occlusion amount in the red box varies in both dimension with different scale factors.

C. MULTI-ORIENTATION PARTIAL ANGULAR COHERENCE

Traditional angular coherence is invalid in the presence of occlusions. However, according to the above analysis, the angular pixels along the separation line will still remain photo-consistent. This is the main motivation of the multi-orientation partial angular coherence. Instead of evaluating the angular coherence using all the angular pixels, we only utilize some subsets of the angular pixels.

Like most light field depth estimation methods, the reference view is set as the central view in this paper. It is hard to exactly extract the separating line in the angular patches due to the complex slopes and occlusion amounts. In this paper, the multi-orientation lines used to calculate partial angular coherence are parallel to the occlusion edge but go through the origin ($u = 0, v = 0$). In general, the optimal direction for calculating the partial angular coherence near occlusion boundaries is parallel to the occlusion edge. As shown in Fig.6(a), the angular pixels on the red lines, which is parallel to the occlusion edge, hold photo-consistency property.

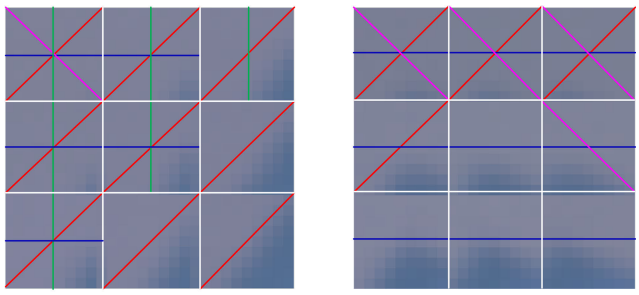

FIGURE 6. The optimal directions for evaluating partial angular coherence in some occlusion cases. (a) Straight line occlusion edge. (b) Ridge-shaped occlusion edge.

In order to obtain the optimal occlusion direction, we use four directions to calculate angular coherence: vertical, horizontal and two diagonal directions. After individually computing the partial angular coherence, we compare each with the other three. However, in the four directions, there may be no one that is parallel to the occlusion edge. More directions will provide better performance, but not essentially for validating our method. In addition, there may be more than one optimal direction. As shown in Fig.6(a), many cases of different occlusion amount are present. It can be seen that in one angular patch, there is often more than one direction in the four used directions that can ensure photo-consistency, such as the blue, green and purple lines. The other lines are not always parallel to the occlusion edge, but the pixels on these lines have similar pixel values. As shown in Fig.6(b), if the occluder is not a straight line and the angle is less than

180°, there will also be extra lines on which the pixels have photo-consistency. However, for the hill-shaped occluders, it is hard to find an optimal direction. In other words, if the angular pixels on the four candidate directions are from the un-occluded rays, our method will perform well. For complex occlusion situations, rotation will bring limitations to our method, but the degradation is not severe.

In fact, the partial coherence is valid by utilizing the rays that are not occluded. In the optics, this can be achieved by using a line mask on the main lens plane or a "line" lens. If the direction of the line is equal to that of the occlusion, the occluder will not be observed by the camera. With a traditional camera, this is not practical because of the uncertainty of the occlusion direction. A light field camera records the direction for each ray; thus, it is capable of recording and resolving the angular coherence in any direction in a single one snapshot. In other words, an adaptive line mask is formed for every spatial point by computation. By choosing the right direction, depth can be inferred near occlusion border. Generally, it is impossible to utilize such a physical adaptive line mask in photography. With the strong power brought by the combination of novel optics and computations, the light field camera enables complex analysis about the scene.

III. DEPTH ESTIMATION ALGORITHM

In this section, we present a compact algorithm for depth estimation using the proposed partial angular coherence. For the initial depth estimation, we consider four possible occlusion orientations and we use a guided filter [21] for cost volume refinement to generate the final depth map.

A. INITIAL DEPTH ESTIMATION

The reference view is chosen as the central view to define the direction of observation. To obtain the light field corresponding to different depth, we can shear the whole light field by Equation (1). In order to make the shearing process more concise, we re-write the right-side of Equation (1) to be

$$L_0(u, v, s^f(\alpha), t^f(\alpha)) = L_0^{(u,v)}(s^f(d), t^f(d)), \tag{10}$$

where $d = r \times u(1 - 1/\alpha)$ is the disparity between the current view and the central view. $r = \Delta u/\Delta s$ is a ratio to deal with the unequal sampling rates in the spatial and angular dimensions. In fact, the shearing process is a linear shift on each individual view with corresponding disparity. In practice, we translate each view within a disparity range.

Here we use line-shaped masks to pick up the desired angular pixels. The masks are applied to the sheared light field to calculate the variance according to Equation (5) by

$$\sigma_p^2(d) = \sum_{\{(u,v)|M^p(u,v)\neq 0\}} \frac{(L^{u,v}(s^f(d), t^f(d)) - I^c(s,t))^2}{N_u N_v}, \tag{11}$$

where $M^p(u, v)$ are the masks defined by the directions for evaluating the partial angular coherence. Considering the discussion in section II, we compute the partial angular coherence in four directions around the origin of the angular coordinates. Since we utilize four directions whose angle are 0°, 45°, 90° and 45° in our method, we will use the angle as the superscript. In addition, we still compute the full angular coherence and we denote its superscript as *ALL*. For example, assuming a light field with 3×3 sub-aperture images, in our setup, the masks are

$$M^{45} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad M^{-45} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$M^{0} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad M^{90} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \text{ and}$$
$$M^{ALL} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \tag{12}$$

An advantage to apply such a mask is that we can take vignetting into account with little effort if necessary. The weights in the mask could be modified according to the angular coordinates of each view.

To further control the robustness against noise, we use the following cost function:

$$\varphi^p(d) = 1 - \exp(-\frac{\sigma_p^2(d)}{2\sigma_d^2}), \tag{13}$$

where σ_d controls the sensitiveness to noise. The value for every pixel will be globally minimal at the ground-truth depth in the partial angular patch of the same orientation with the occlusion boundary. Thus, the initial depth can be obtained by global optimization for each pixel within all the partial angular coherences:

$$d_{init} = \arg\min_{\alpha}(\min(\varphi^0(d), \varphi^{90}(d), \varphi^{45}(d), \varphi^{-45}(d), \varphi^{ALL}(d))). \tag{14}$$

We take the scene *dino* for an example to present the working process of the initial estimation, as shown in Fig.7. It can be seen that the initial depth map has sharp edges around occlusion boundaries, proving that our method can tackle occlusion problems in practice. The partial angular coherence cost of an occluded pixel marked in red in Fig.7(a) is computed in Fig.7(b). The cost function is depicted in Fig.7(c) and the initial depth map is Fig.7(d). The occlusion orientation is vertical, thus the partial angular pixels in this direction have significantly lower variance than that of other directions. Even if the point is defocused, the partial angular pixels that are parallel to the occlusion orientation will still have lower variance. For the point around the occlusion boundary marked red in Fig.7(a), its vertical partial angular coherence cost keeps at a low level in all disparity labels and reaches the globally minimum value at the ground-truth disparity.

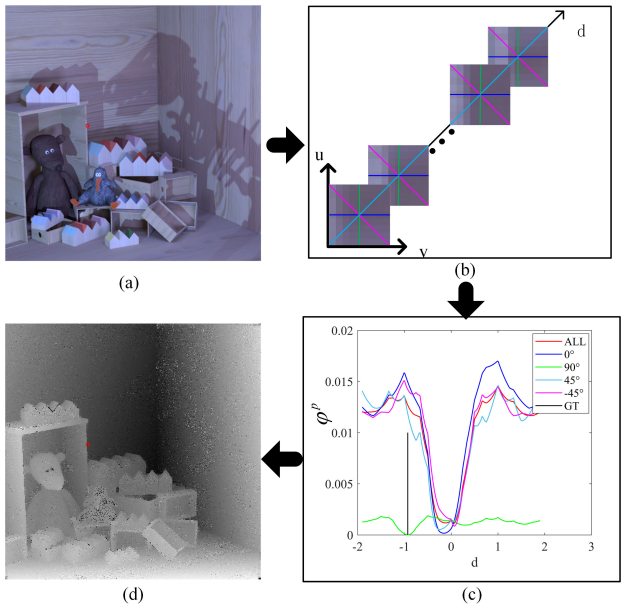

FIGURE 7. The working principle of the proposed method. (a) The input light field image. (b) The partial angular coherence is calculated in multiple directions. (c) The cost function of the spatial pixel marked red in (a). (d) The initial depth map.

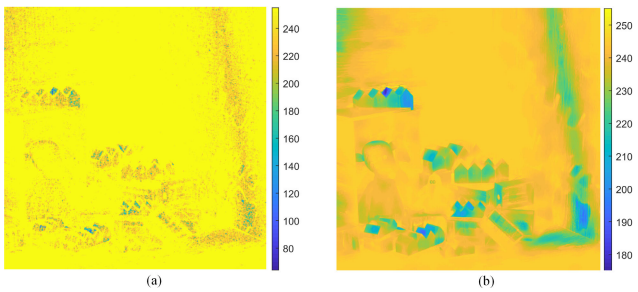

FIGURE 8. The comparison before and after guided filtering for a certain slice in the cost volume. (a) The initial cost. (b) After filtering guided by the central view. The values are normalized to [0,255] for display.

B. DEPTH OPTIMIZATION

Just as depicted in Fig.7, the initial depth map typically contains outliers due to noise and inherent matching uncertainty, especially in the noisy and textureless regions. Therefore, depth refinement methods are widely used in previous works. In order to refine the initial depth map, typical methods are either within a Markov random field (MRF) or a variational framework [22]. A light field image typically involves a large number of views and depth labels. Both frameworks are global approaches which are computationally expensive when being applied to light field depth estimation. For fast cost volume filtering, the guided filter has been widely used in labeling problems for computer vision tasks [23]. However, if occlusion regions are not dealt with strictly before filtering, the guided filter will propagate wrong information to similar regions.

The proposed initial estimation is able to handle occlusions during the process. Therefore, we can use guided filtering to fast optimize the cost volumes. The guided image is the

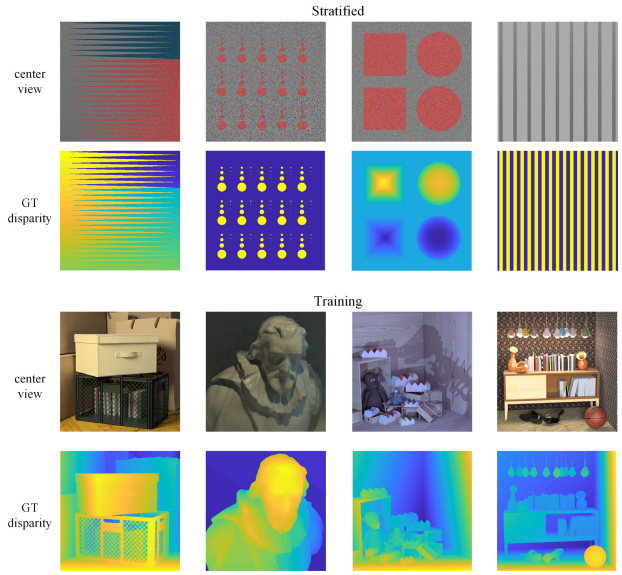

FIGURE 9. The center views and the ground-truth disparity of the dataset used in the experiment.

reference view. The cost volume C to be filtered is the local minimum of the five partial angular cost function as

$$C(\alpha) = \min(\varphi^0(\alpha), \varphi^{90}(\alpha), \varphi^{45}(\alpha), \varphi^{-45}(\alpha), \varphi^{ALL}(\alpha)). \tag{15}$$

Thus, the cost volume C is a three-dimensional array storing the costs for deciding the depth label. In the filtering process, each slice of the cost volume is filtered guided by the reference view. To be more detailed, we take a pixel index i at depth label α as an example. The output is a weighted average of all pixels in the cost volume slice:

$$C_i'(\alpha) = \sum_j W_{i,j}(I^c) C_j(\alpha)$$

$$W_{i,j} = \frac{1}{\omega^2} \sum_{k:(i,j)\in\omega_k} (1 + (I_i^c - \mu_k)^T (\Sigma_k + \varepsilon \mathrm{U})^{-1} (I_j^c - \mu_k)), \tag{16}$$

where j is the pixel indexes defined by the filter window. For colorful central views, I_i^c, I_j^c and μ_k are 3×1 color vectors and Σ_k is the co-variance matrix. U is the identity matrix of size 3×3 and ε is a penalty coefficient to control the edge-preserving ability.

A major assumption here is that depth discontinuity occurs at occlusion boundaries. Therefore, each slice of the cost volume will have similar edge information to that in the central view, as shown in Fig.8(a). This ensures the cost volume slices are linear with the central view. The outlier in the initial depth map results from the outliers in the cost volumes. By guided filtering, we can propagate the correct information with higher confidence to the nearing regions with noise or less texture. In Fig.8(b), the outliers in the slice of the cost volume are mostly removed.

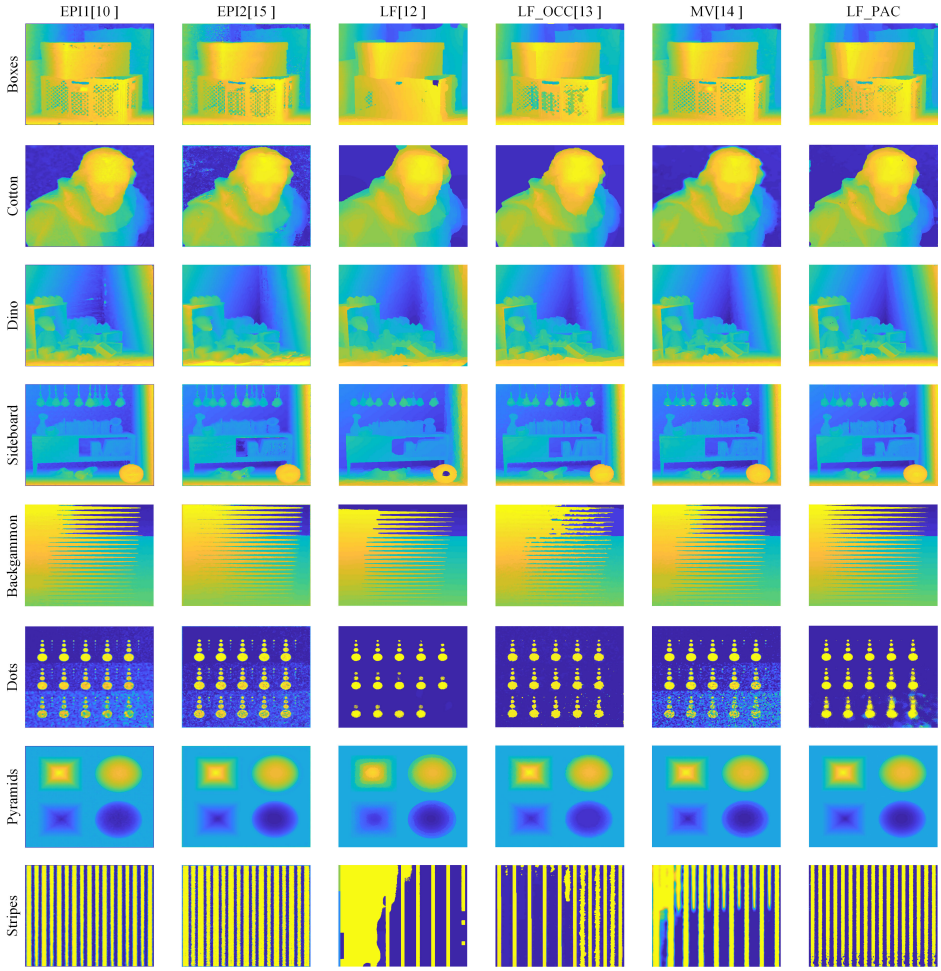

FIGURE 10. Our depth maps for the training and stratified datasets compared to that of the other methods. The pseudo-color depth maps are shown using the disparity value.

Therefore, the cost volume slices become smooth while still preserving sharp edges.

IV. EXPERIMENTS AND RESULTS

In this section, we validate our method both on a recent benchmark and real light field images taken by a Lytro camera. For the images in the benchmark, the depth maps obtained are converted to the corresponding disparity value to keep consistency with the dataset. For the real light field images, we calculate the depth label map and evaluate it with human perception. The σ_d in Equation (13) is 0.01, and the radius and the penalty coefficient in the Equation (16) are 5 and 10^{-4}. When we deal with several scenes in the stratified group, the σ_d is increased. The depth labels are 256 in all the experiments. The proposed algorithm is denoted as LF_PAC.

A. PERFORMANCE ON THE BENCHMARK DATASET

To test the accuracy of our algorithm, we use the light field benchmark dataset [14]. The central views and ground-truth disparity maps of the light fields used are shown in Fig.9.

The results in Fig.10 demonstrate the superior performance of the partial angular coherence in depth estimation. It can be seen that our method can preserve sharper transitions around occlusion boundaries in all scenes than the five state-of-the-art algorithms, which serve as a baseline to stimulate further progress. Among these algorithms, LF_OCC is the most similar one to our method. It proposed to use the edges in the central view to perform occlusion-aware depth estimation and fails at the non-texture areas with depth discontinuity. The result can be seen in the scene *stripes*, which is designed to assess the influence of texture and contrast at occlusion boundaries. The amount of texture is gradually increasing from left to right. Our method produces fine depth map from left to right while LF_OCC fails in the left because it heavily relies on the edge information. Because the occlusion handling is independent, our method is occlusion-free.

For further quantitative assessment, an overview of the results and a comparison for *BadPixel(0.07)* and *MSE* are shown in Table.1 and Table.2. The *BadPixel(0.07)* is calculated by the percentage of the pixels that differ from the ground-truth disparity by more than 0.07 pixels. The number in bold shows the best metric among the results of the

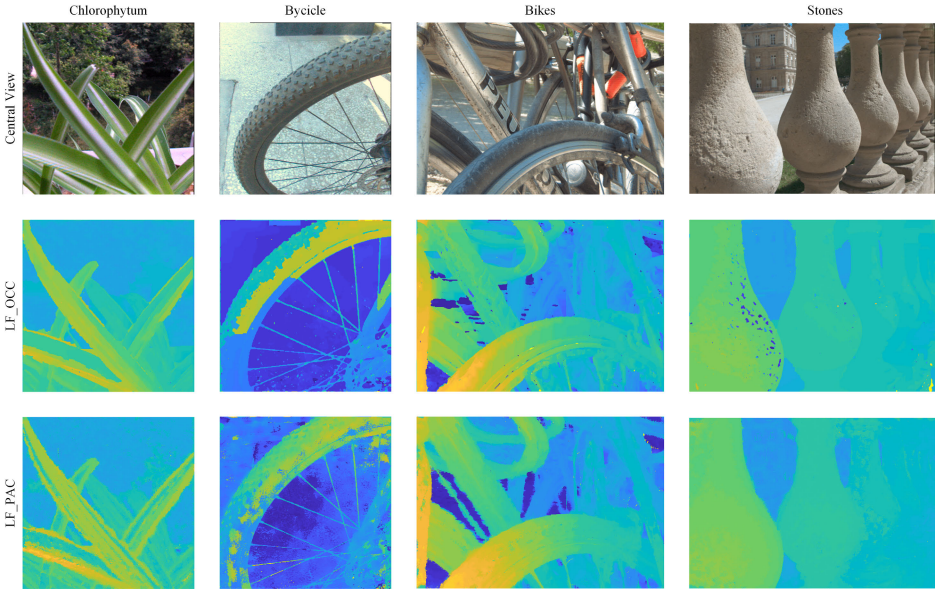

FIGURE 11. Our depth maps for real-world scenes. The first two columns are the results for images captured by a Lytro F01 camera, and the latter two columns for Lytro Illum.

TABLE 1. Badpixel(0.07) comparison with the baseline algorithms[a].

Scene	EPI1	EPI2	LF	LF_OCC	MV	LF_PAC
Boxes	24.45	29.80	23.02	26.52	21.64	**20.32**
Cotton	13.93	16.69	7.83	6.22	9.25	**4.20**
Dino	10.35	15.67	19.03	14.91	6.29	**4.08**
Sideboard	18.38	18.95	21.99	18.50	17.30	**8.65**
Backgammon	21.33	22.08	5.52	19.01	8.92	**2.29**
Dots	62.00	46.53	**2.90**	5.82	45.59	7.23
Pyramids	0.86	1.08	12.35	3.17	0.78	**0.21**
Stripes	25.81	23.81	35.74	18.41	46.19	**5.12**

[a] BadPix(0.07): The percentage of pixels with errors more than 0.07 pixels. Lower scores are better and the best score among the six algorithms are in bold.

TABLE 2. MSE comparison with the baseline algorithms[a].

Scene	EPI1	EPI2	LF	LF_OCC	MV	LF_PAC
Boxes	8.72	10.93	17.43	9.85	**8.59**	9.39
Cotton	2.25	4.32	9.17	1.07	3.44	**1.05**
Dino	1.23	2.07	1.16	1.14	0.75	**0.43**
Sideboard	2.85	4.65	5.07	2.30	1.89	**1.10**
Backgammon	9.56	20.78	13.01	21.59	13.23	**3.65**
Dots	5.73	6.66	5.68	**3.30**	7.26	4.92
Pyramids	0.03	0.02	0.27	0.10	0.05	**0.01**
Stripes	2.67	6.10	17.45	8.13	12.17	**1.68**

[a] MSE: The mean squared error over all pixels, multiplied with 100. Lower scores are better and the best scores among the six algorithms are in bold.

compared algorithms. We get better results than the previous methods in both metrics except for the scene *dots* and *boxes*. The scene named *boxes* contains complex occlusion orientations and the occlusion changes quickly. This breaks the basic assumption of our method, resulting in a large number of outliers around the holes. However, although the MSE is not the best, the *Badpixel(0.07)* is still best due to the accurate depth extraction near occlusion boundaries. A way to remedy this problem is to use more orientations and less angular pixels corresponding to the un-occluded rays. In addition, the improvement is huge on the quantitative metrics. It is worth noting that the *Badpixel(0.07)* of the scene *Backgammon* ranks first in the current benchmark is 2.937, and our score is 2.293. Another outlier is the performance on the stratified scene *dots*, with the unsatisfactory results in the lower right regions. Since our method uses partial angular pixels to compute coherence, the robustness against noise is inevitably reduced. A way to remedy this is to combine several cues like Tao *et al.* [24].

B. REAL LIGHT FIELD IMAGES

To test the performance of the proposed method on real-world images, we evaluate on light field images captured by Lytro F01 and Illum cameras. The light field images of the Illum camera is from a freely available EPFL [25] datasets. The images of the F01 camera is taken by ourselves and the decoding method for the F01 image is from Dansereau *et al.* [26]. The results are shown in Fig.11. Clearly, our depth maps have sharp edges but with remaining outliers in the low-SNR regions. This is due to the noise in the capturing and decoding process. The depth results from the LF_OCC are placed in the second row. It can be seen our depth maps are similar to that of the LF_OCC, while ours are a little better at occlusion edges. In some occlusion boundaries where occlusion textures are weak, LF_OCC fails to maintain sharp edges. The reason accounting for this is that they heavily rely on the edge information in the central view. If the edge detection fails, the depth estimation will also fail to keep sharp transitions. Our results can also be further optimized by combining more cues to provide better robustness.

V. CONCLUSION

In this paper, we proposed an accurate depth estimation method which can deal with occlusions explicitly for light field images. We show that when occlusions present, although the whole angular patch will lose photo-consistency property, a line subset of the pixels still exhibit photo-consistency. Utilizing this unique phenomenon, we can tackle occlusions without knowing any information about the occlusion edges and orientations. The initial depth estimation is accomplished by using four directions and can obtain sharp edges with several outliers. For depth map optimization, because the occlusions are well dealt with during initial estimation, we used guided filtering on each cost volume slice to remove the outliers in the initial depth map. We demonstrated the benefits and superior performance on a recent benchmark designed for light field depth estimation, which quantitatively emphasized the high-accuracy and the strong power for handling occlusions. In addition, we also validated our algorithm on light fields captured by real Lytro cameras, showing that our method performs excellently in real light field images, especially near occlusion boundaries. In future work, it would be interesting to exploit a rotation-invariant method to deal with complex occlusion areas and orientations.

REFERENCES

[1] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Refocusing distance of a standard plenoptic camera," *Opt. Express*, vol. 24, no. 19, pp. 21521–21540, 2016.

[2] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005.

[3] R. Ng, "Digital light field photography," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, 2006.

[4] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2009, pp. 1–8.

[5] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.

[6] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 972–986, May 2012.

[7] J. Zhang, M. Wang, J. Gao, Y. Wang, X. Zhang, and X. Wu, "Saliency detection with a deeper investigation of light field," presented at the 24th Int. Conf. Artif. Intell., Buenos Aires, Argentina, 2015.

[8] M. Strecke, A. Alperovich, and B. Goldluecke, "Accurate depth and normal maps from occlusion-aware focal stack symmetry," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2529–2537.

[9] O. Johannsen *et al.*, "A taxonomy and evaluation of dense light field depth estimation algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1795–1812.

[10] O. Johannsen, A. Sulc, and B. Goldluecke, "What sparse light field coding reveals about scene structure," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3262–3270.

[11] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar. 2014.

[12] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1547–1555.

[13] T. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3487–3495.

[14] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Proc. Asian Conf. Comput. Vis.*, 2016, pp. 19–34.

[15] S. Wanner, C. Straehle, and B. Goldluecke, "Globally consistent multi-label assignment on the ray space of 4D light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1011–1018.

[16] C. H. Esteban, G. Vogiatzis, and R. Cipolla, "Multiview photometric stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 548–554, Mar. 2008.

[17] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.*, vol. 1, no. 1, pp. 7–55, 1987.

[18] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu, "Light field stereo matching using bilateral statistics of surface cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1518–1525.

[19] T.-C. Wang, A. A. Efros, and R. Ramamoorthi, "Depth estimation with occlusion modeling using light-field cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2170–2181, Nov. 2016.

[20] H. Sheng, S. Zhang, X. Cao, Y. Fang, and Z. Xiong, "Geometric occlusion analysis in depth estimation using integral guided filter for light-field image," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5758–5771, Dec. 2017.

[21] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[22] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.

[23] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 504–511, Feb. 2013.

[24] M. W. Tao, P. P. Srinivasan, S. Hadap, S. Rusinkiewicz, J. Malik, and R. Ramamoorthi, "Shape estimation from shading, defocus, and correspondence using light-field angular coherence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 546–560, Mar. 2017.

[25] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *Proc. 8th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Lisbon, Portugal, 2016. [Online]. Available: https://mmspg.epfl.ch/downloads/epfl-light-field-image-dataset/

[26] 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.*, Jun. 2013, pp. 1027–1034.

ZHENGHUA GUO was born in Zigong, Sichuan, China, in 1992. He received the B.S. degree from Nanjing University. He is currently pursuing the Ph.D. degree with the Institute of Optics and Electronics, Chinese Academy of Science. His current research interests include light field technologies and deep learning in light field depth estimation.

JUNLONG WU received the B.S. degree from the University of Electronic Science and Technology of China. He is currently pursuing the Ph.D. degree with the Institute of Optics and Electronics, Chinese Academy of Science. His current research interests include light field camera decoding and calibration.

XIANFENG CHEN received the B.S. degree from the University of Electronic Science and Technology of China. He is currently pursuing the master's degree with the University of Chinese Academy of Sciences. His current research interests include stereo matching and computer vision.

SHUAI MA received the B.S. degree from the China University of Mining and Technology. He is currently pursuing the Ph.D. degree with the Institute of Optics and Electronics, Chinese Academy of Science. His current research interest includes light field technologies.

LICHENG ZHU received the B.S. degree from the Beijing Institute of Technology. He is currently pursuing the Ph.D. degree with the Institute of Optics and Electronics, Chinese Academy of Science. His current research interest includes adaptive optics for imaging.

PING YANG graduated from the Institute of Optics and Electronics, Chinese Academy of Science. He received the Ph.D. degree, in 2008. He has worked a Senior Research Scientist with 20-year experience in adaptive optics. He is currently leading research activities in adaptive optics (wavefront sensing, and control) and light field imaging.

BING XU is currently a Senior Research Scientist with the Institute of Optics and Electronics, Chinese Academy of Science. His current research interests include application of adaptive optics in improving laser beam quality, wavefront detector development, and application of light field cameras.