Pattern Recognition 000 (2017) 1–13

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Occlusion-aware depth estimation for light field using multi-orientation EPIs

Hao Sheng[a,b], Pan Zhao[a], Shuo Zhang[a,*], Jun Zhang[c], Da Yang[a]

[a] State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing, PR China
[b] Shenzhen Key Laboratory of Data Vitalization, Research Institute in Shenzhen, Beihang University, Shenzhen, PR China
[c] Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, 53201, USA

ARTICLE INFO

Article history:
Received 31 January 2017
Revised 19 June 2017
Accepted 5 September 2017
Available online xxx

Keywords:
Light field
Depth estimation
Multi-orientation EPIs
Occlusion analysis

ABSTRACT

Epipolar plane images (EPIs) contain special linear structures that reflect the disparity of a 3D point and are widely used in light field depth estimation. However, previous EPI-based approaches only utilize horizontal and vertical EPIs to estimate local disparities and ignore diagonal directions. In order to make full use of the regular grid light field images, we develop a strategy to extract epipolar plane images in all available directions. Based on the multi-orientation EPIs, a specific EPI in which the point is not occluded is found and used to calculate robust depth estimation. We also design a novel framework to estimate the depth information which combines the local depth with edge orientation. The multi-orientation EPIs and optimal orientation selection are proved to be effective in detecting and excluding occlusions. Experimental results show that the proposed method outperforms state-of-the-art depth estimation methods, especially near occlusion boundaries.

1. Introduction

Due to the large amount of information 4D light field cameras capture, they have become a popular technology for image acquisition. Unlike other cameras, a 4D light field camera provides information about the accumulated intensity of each image point, and also captures the light direction. Commercial light field cameras such as Lytro [1] and Raytrix [2] are successful in the market, because they capture scenes from different directions in one shot and provide information about the structure of a scene. The light field allows a wide range of applications, such as light field rendering [3,4], super-resolution [5,6], digital refocusing [7] and 3D reconstruction [8–10].

A lot of attention has been given to efficient and robust algorithms for light field depth estimation, and various algorithms are developed depending on different kinds of images, such as multi-view images, focal stack, and Epipolar Plane Images (EPIs). Occlusion is always a tough problem for depth estimation [9,11]. It occurs when points closer to the camera obstruct points that are farther from the camera. Consequently, the occluded points are visible only in some sub-aperture images. In multi-view stereo matching, an angular sampling image which contains the possible imaging intensities of one point in all sub-aperture images is available in light field images. It is supposed to be consistent if the scene is Lambertian and the corresponding point is not occluded. However, the state-of-the-art techniques fail when the occlusion breaks the photo consistency assumption, which is difficult to estimate the correct depth label. Some methods are proposed to estimate the slopes of the lines in the EPI due to the continuous sampling, but these methods [12–14] exhibit problems in occluded and noisy regions.

Existing EPI-based methods only utilize horizontal and vertical slices, since these two EPIs are easier to extract and manipulate. In this paper, we propose to extract more EPIs along other available directions, which clearly makes more use of the rich light field angular information. Multi-orientation EPIs are perceived as slices through the light field across lines with different orientation in the center view, and the occluded scene points are visible along one of the lines of sight. We prove that the direction of the optimal EPI is parallel to the boundary of the occlusion.

We develop a novel framework that computes disparities by utilizing multi-orientation EPIs. The proposed method can produce different directions of EPIs, as shown in Fig. 1. We use a spinning parallelogram operator (SPO) [15] as the basic approach to locate lines and compute their orientations in each EPI for local depth estimation. The local depth labels are consistent for unoccluded points. We propose to predict the occlusion boundaries using the depth differences in different EPIs. The robust depth label of the

* Corresponding author.
E-mail address: shuo.zhang@buaa.edu.cn (S. Zhang).

http://dx.doi.org/10.1016/j.patcog.2017.09.010
0031-3203/

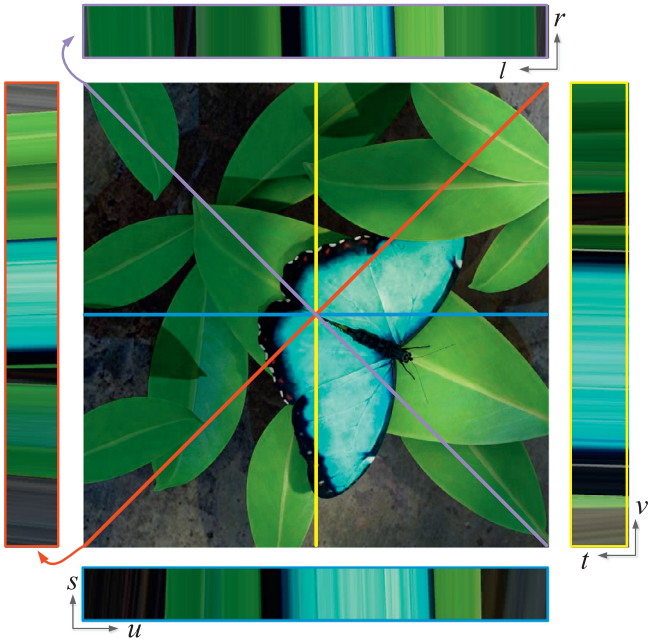

Fig. 1. Multi-orientation epipolar plane images. The picture shows the horizontal, vertical, 45°-orientation and 135°-orientation EPIs. The slope of the line in the EPIs is related to disparity and the slopes on different EPIs corresponding to the center point on the center view are the same, as shown on the bottom and right.

occluded point is then chosen from the appropriate EPI whose orientation is parallel to the occlusion boundary direction. We finally propose a method which combines the initial depth with the occlusion boundary and then perform a global optimization with a filter-based method to obtain a final high quality depth map. The integration of EPIs significantly improves robustness to occlusion and noise.

Experimental results show that our method achieves better performance near occlusion boundaries, which is effective for both real camera images and synthetic light field images. The main contributions of this paper are:

1. We introduced a method to extract various directions of EPIs and showed that the optimal EPI has the same direction as the occlusion boundary.
2. We proposed to detect the occlusion boundaries based on the variance and gradient of local depth values, and find the accurate depth label from the optimal EPI.
3. We developed a novel framework that combines the local disparity map with the occlusion boundary to deal with occlusions so as to obtain the robust depth maps.

2. Related work

Over the years, various methods have been introduced to estimate depth using light field images, including multi-view based methods [11,16,37], focus stack based method [17], and epipolar plane image (EPI) based methods [14,18].

2.1. Depth from light-field images

Multi-view stereo matching: Traditional stereo matching that used multi-view images was available because light field images had a dense collection of views. Yu et al. [19] encoded 3D line constraints into a light field image, and computed depth maps through line matching between sub-aperture images. Chen et al. [9] adopted a new bilateral consistency metric on the angular patch as a cost, to measure the probability of significant occlusions. Heber and Pock [20] proposed a low rank structure regularization technique and aligned the sub-aperture images to estimate depth maps. However, all of the forementioned methods relied on the dense sampling in angular resolution.

Depth from focus stack: Depth can be estimated by defocus cue because light field images are able to refocus at different depths in the scene [7]. Tao et al. [10] combined the correspondence and defocused measure to estimate accurate depth values. They used the patch-based variance measurements to measure defocus. This theory was extended by Tao et al. [17] who added a shading constraint as the regularization term. They used the average of the intended difference between the angular patch in the refocus and the corresponding point in the center pinhole images in order to modify the original defocus measure.

Depth from EPI: Several methods based on EPI have been developed due to the advantage of continuous space in angular direction. Wanner et al. [14] used the first order structure tensor for orientation estimation. However, the tensor structure relied on a high angular resolution and became too random to calculate when occlusion occurs. Tosic et al. [21] proposed a local method for ray detection and dense depth estimation by using the normalized second derivative of the Ray Gaussian kernel. Zhang et al. [15] proposed a spinning parallelogram operator (SPO) to calculate depth map, which measured the slopes of the lines on the EPI by maximizing distribution distances between the two parts of the parallelogram window. All of these methods calculated the local disparity only based on horizontal and vertical EPIs, which utilize less angular information.

2.2. Occlusion analysis for light-field images

Since occlusion is an important part of light field applications, there are different methods developed to solve the occlusion problem: stereo with occlusion [22,23], defocus and correspondence respond [24,25], and EPI analysis [14].

Stereo with occlusion: There were some researches done with multi-view stereo matching in order to solve occlusion problems. Kang et al. [26] separated the matching term into two parts based on a cost function, and the part with a small variance was regarded as the cost without occlusion. Kolmogorov et al. [27] utilized the visibility constraint to model occlusion and optimized it by the graph-cut framework. The method observed the visibility of a pixel in corresponding images to design the occlusion cost. But it remains difficult to address the method in a huge number of views, such as light field. To handle occlusion problems, Chen et al. [9] developed a bilateral metric that measured possible occlusion parts in the stereo matching calculation. More specifically, they calculated the similarity between the angular patch pixels and the central pixels. However, their method was biased when the input light field images became noisy.

Defocus and correspondence respond: In order to handle occlusion problems, Wang et al. [28] separated the angular patch into two parts based on the boundary orientation in spatial patches, and discovered the minimum cost of the two regions to calculate the defocus and correspondence response. However, this method estimated occlusion possibilities based on the calculated responses of every depth label, which made the estimation uncertain for ambiguous regions and noises, and its performance was affected by how well the angular patch was divided. In a new matching technique proposed by Williem et al. [25], a corresponding method was designed to eliminate the influence of occlusion for consistency matching and focusing. With regard to the consistence matching, a new method was designed to calculate the matching value by using the intensity value of the main pixel. For the defocus response, a variable focus measure was used to divide the region

Please cite this article as: H. Sheng et al., Occlusion-aware depth estimation for light field using multi-orientation EPIs, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.09.010

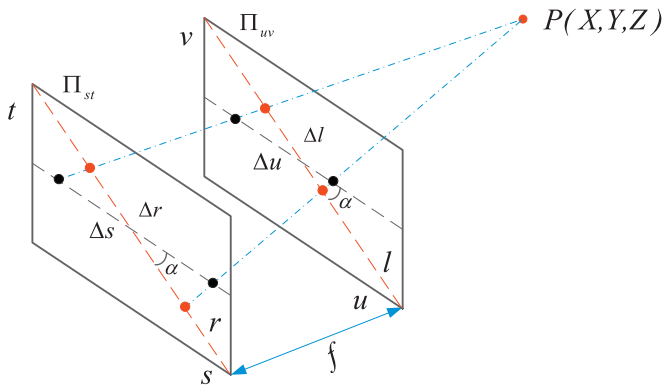

Fig. 2. Two-plane parametrization of a 4D light field by coordinates (u, v) in the image plane Π_{uv} and coordinates (s, t) in the camera plane Π_{st}, which describes the projection of every 3D Point P. Angle α is the orientation of the red slice, whose coordinate is (l, r). The relationship between l and r in α-orientation EPI is similar with the relationship between u and s.

into nine sub-regions. Then they calculated the degree of focus for each region, and selected the smallest matching corresponding to the depth of the measurement point. However, this method needs high angular resolution, since the angular entropy metric was less reliable when the occluder became more dominant than the occluded views in the angular patch.

EPI analysis: For epipolar plane images, Wanner et al. [14,18] introduced a global consistent visible constraint in the case of occluded points, *i.e.* only closer objects occluded relatively deeper objects. A corresponding penalty term was added in order that the point used calculating the structure tensor could be corrected in the process of minimizing the energy function. But since this method did not try to match pixels in different positions, it could not be used in light field images with sparsely sampled view points. Tosic et al. [21] proposed a method for constructing depth spaces of different sizes to represent the light field. This method also detected occlusion by analyzing overlapping rays and their ordering, *i.e.* the region with the smaller variance was considered to be the foreground. However, the visibility constraint was valid only when the initial depth estimate was correct and this condition was difficult to guarantee. Johannsen et al. [29] separately encoded upper and lower parts of the EPIs in order to explicitly handle occlusions. The orientations were divided into eight classes that corresponded to compass directions, but they can only get four directions based on horizontal and vertical EPI patches. In addition, none of the methods analyzed the disparity deviation caused by the different angles of EPIs so as to deal with the occlusion problem.

In this paper, we propose a novel depth estimation algorithm that is robust to occlusion by extracting multi-orientation EPIs and consider the orientation of occlusion boundaries, which is also useful for other light field applications, such as image segmentation.

3. Multi-orientation EPIs occlusion analysis

In this section, a common practice of using a 2*D* plane is adopted to parametrize the 4*D* light field, as shown in Fig. 2. The camera plane Π_{st} is at $z = 0$ and the image plane Π_{uv} is at $z = 1$. Light through Point P intersects the main lens plane at point(s, t) and intersects the imaging plane at point (u, v). The light field is then expressed as $L(s, t, u, v)$. Specifically, (s, t) can be expressed as the coordinate of the views, and (u, v) as the image coordinate captured in different views. When we fix the coordinates (s^*, t^*), the sub-aperture image $I_{s^*,t^*}(u, v)$ is obtained. We focus on representing EPIs in other orientations. The view directions of light field images include not only horizontal or vertical directions, but also other additional directions. To fully understand the occlusion relationship and structure information, more perspective information is better, which is the original intention of extracting multi-orientation EPIs. We use l and r to express the spatial and angular location in EPIs, respectively. The angle between the slice and the horizontal plane is α.

3.1. Multi-orientation epipolar plane images

We consider the method to extract more EPIs to make better use of the rich angle information of the light field images. For light field $L(u, v, s, t)$, the traditional epipolar plane image (EPI) $I_{v,t}(u, s)$ is obtained by fixing the coordinate (v^*, t^*) and changing coordinate (u, s) at the same time, as shown in Eq. (1), the same as $I_{u,s}(v, t)$. It can be regarded as a 2D slice in a constant angular image stack in the light field that reflects the changes of the same scene point in different sub-aperture images:

$$I_{v^*,t^*}(u, s) = L(u, v^*, s, t^*). \tag{1}$$

The EPI contains simple linear structures, which are projected by corresponding scene points. Fig. 2 shows that the coordinate u of a point P will change as the coordinate s varies. Specifically, the line in the EPI is able to reflect its depth and the actual depth Z is described according to Bolles et al. [30]:

$$Z = f\frac{\Delta s}{\Delta u}, \tag{2}$$

where f is the distance between the two parallel planes, Δs is the geometrical distance between the two cameras along the horizontal line, and Δu is the distance between the scene point in different local coordinate systems.

We note that the commonly used EPI $I_{v,t}(u, s)$ or $I_{u,s}(v, t)$ only utilize the horizontal or vertical direction angular coordinate in the camera plane Π_{st}, which means that the pervious depth from EPI-based methods do not make full use of the angular information of light field images. In this paper, we choose α to represent the direction of EPI. The projections of the EPI on the camera plane Π_{st} and the image plane Π_{uv} satisfy a same linear function $f(\cdot)$, which are expressed as $t = f(s)$ and $v = f(u)$, respectively. When the selected sub-aperture images superimposed, as seen in Fig. 2, we get a new EPI:

$$I_\alpha(l, r) = L(u, f(u), s, f(s)), \tag{3}$$

where α is the angle of the line, $f(\cdot)$ indicates the chosen sub-aperture images which compose the EPI in a straight line. The variable l represents the spatial information, and r represents the angle information. (l, r) can be transformed by (u, v) through:

$$\Delta l = \frac{\Delta u}{\cos\alpha}, \Delta r = \frac{\Delta s}{\cos\alpha}. \tag{4}$$

According to the geometry of Fig. 2, if we vary r, the coordinate l will change as:

$$\Delta l = \frac{f}{Z}\Delta r. \tag{5}$$

Then we can obtain

$$\frac{\Delta l}{\Delta r} = \frac{\Delta u}{\Delta s} = \frac{f}{Z}, \tag{6}$$

where $\Delta l/\Delta r$ is the slope of the lines in $I_\alpha(l, r)$, and the slope is independent of the α. The traditional horizontal EPI $I_{v,t}(u, s)$ and vertical EPI $I_{u,s}(v, t)$ are expressed as $I_{0^\circ}(l, r)$ and $I_{90^\circ}(l, r)$, respectively.

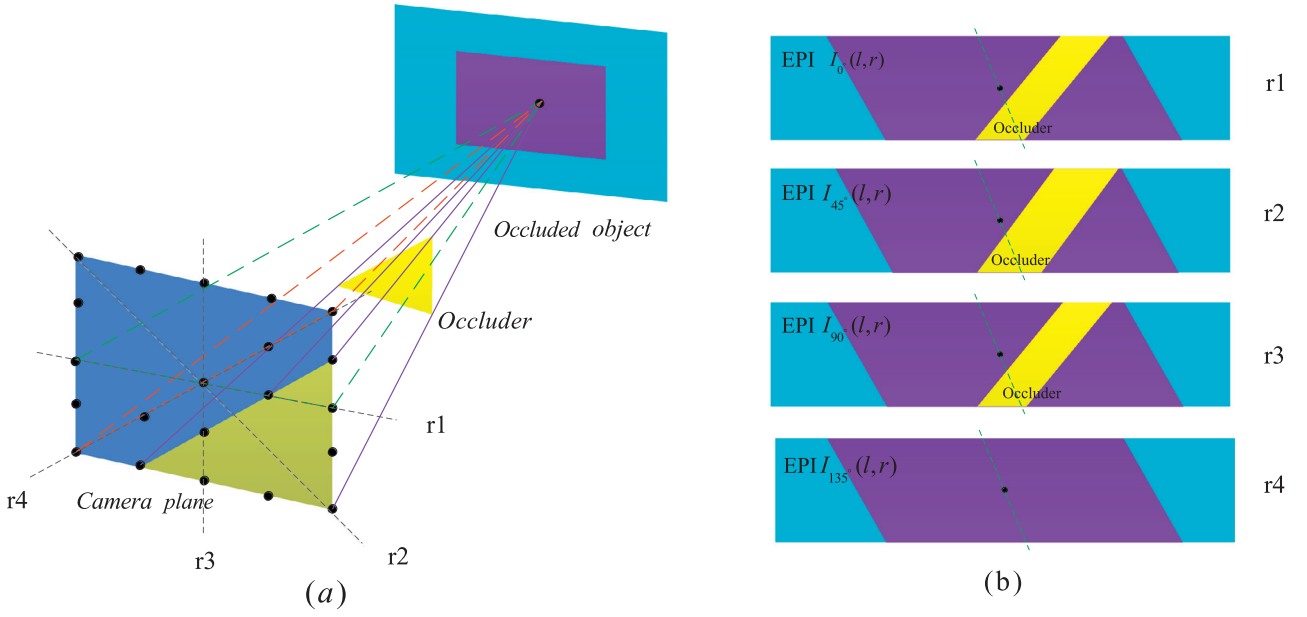

Fig. 3. Multi-orientation EPIs under the occlusion condition. (a) shows the occlusion model, when a point is occluded, the horizontal slice contains the views where the occluder is seen, while the red slice through the light field is produced by the unoccluded views. (b) shows the EPIs extracted along different directions. Since the occlusion boundary direction of occluded point is near 135°, EPI $I_{135^\circ}(l,r)$ is preferred.

3.2. Optimal EPI selection

Occlusions are the main cause of errors in depth estimation. When an object is occluded in one direction α, it means that the object is occluded in the corresponding EPI I_α. In this section, we show a way to deal with occlusions thus yielding more accurate results at occlusion boundaries. If a point is occluded by some objects, many rays from the point actually hit occluders so we notice that some views are occluded by the front occlusions. In general, the shape of the occluded cameras on the camera plane are similar to the occluder based on physical image formation, as shown in Fig. 3.

When the occlusion occurs, the slices through the center view across horizontal or vertical lines contain the occluded cameras in the camera plane, as shown in Fig. 3(a). The line in the horizontal or vertical EPIs, which is projected by the corresponding scene point P, is a broken line. A part of the line is the projection of the point P in different views while the other represents the occlusion point. However, a new EPI $I_\alpha(l, r)$ through the light field across the line is obtained when we select a line that passes through the center point and is parallel to the occlusion boundary in the camera plane. The line in $I_\alpha(l, r)$ shows the correct disparity for point P.

Fig. 3(b) shows the EPIs extracted along different directions. Obviously, EPI $I_{135^\circ}(l,r)$ is preferred, since the occlusion boundary direction of occluded point is near 135°. When we choose the sub-aperture images along other directions that are not parallel to the occlusion direction, the projection of the obtained slice on the camera plane also hits the yellow area. Thus, it is optimal to choose EPI oriented in the same direction as the occlusion.

In addition, we also extract multi-orientation EPIs in several discrete angular directions since the cameras are arranged in a regular grid on the camera plane. The higher the angular resolution, the more effective EPIs can be extracted.

4. Occlusion-aware depth estimation

In this section, we develop a novel depth estimation framework which combines the occlusion prediction with the multi-orientation EPIs. Algorithm 1 shows our occlusion-aware depth estimation. The complete framework is shown in Fig. 4.

Algorithm 1 Depth estimation via multi-orientation EPIs.

initialize EPI $I_\alpha(l,r)$
{for each θ and each α-orientation EPI, compute depth response}
for all $\alpha \in$ A (A represents the optional angular set of EPI) **do**
  for $(\theta = \theta_{min};\ \theta \le \theta_{max};\theta += \theta_{step})$ **do**
    $d_\alpha(l,r,\theta) = spo(I_\alpha(l,r))$ {depth cost}
  end for
  {For each pixel, compute response optimum}
  $\theta_\alpha = \arg\max(d_\alpha(l,r,\theta))$
end for
{Occlusion predictor}
Compute P_{occ}^{var} *based on* Eq. 9
Compute P_{occ}^{d} *based on* Eq. 10
$P_{occ} = P_{occ}^{var} \cdot P_{occ}^{d}$ {Global optimization}
$Depth = Guidedfilter(\theta_\alpha, occ)$
return *Depth*

4.1. Local depth maps

Multi-orientation EPIs can easily be applied in different EPI-based depth estimation frameworks, such as [14,15], because they are independent of the cost volumes computation. In this paper, the SPO in Zhang et al. [15] is used as our local slope estimation in each EPI, because it is more robust to estimate the slope of the line and easier to implement. SPO measures the slopes of the lines in an EPI by maximizing distribution distances between two parts of the parallelogram window. It uses χ^2 distance of the color histograms to measure the difference between the distributions of pixel colors:

$$\chi^2(g_\theta, h_\theta) = \sum_p \frac{(g_\theta(p) - h_\theta(p))^2}{g_\theta(p) + h_\theta(p)}. \tag{7}$$

where $g_\theta(p)$ and $h_\theta(p)$ are the histograms of the separated parts.

In this subsection, a set of local depth maps are aggregated from various orientation EPIs. We first construct a cost volume with axes (l, r, θ), which is known as disparity space image (DSI) in stereo [31]. We define $d_\alpha(l, r, \theta)$ as the histogram distance mea-

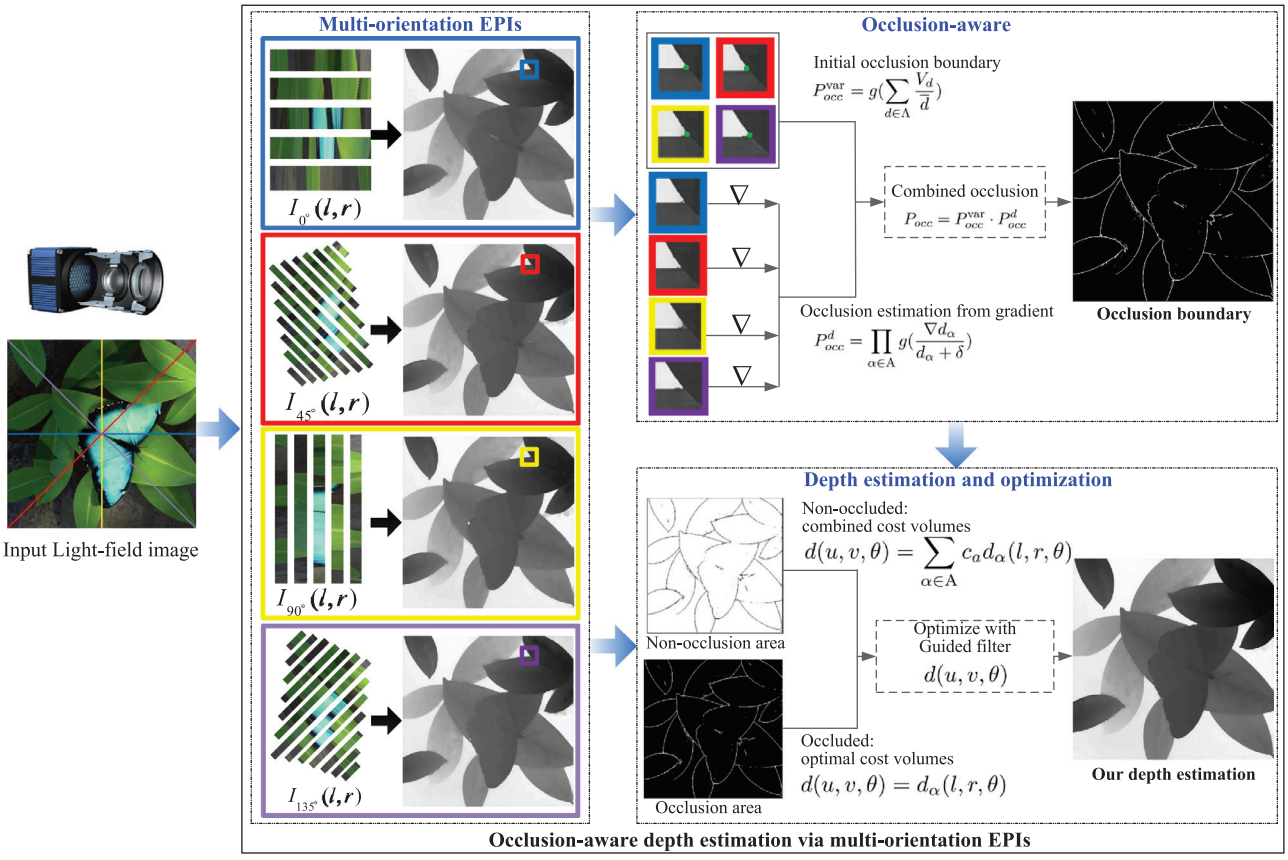

Fig. 4. Depth estimation using Multi-orientation EPIs. The occlusion boundaries are extracted based on different occlusion estimation. The proposed method refines the disparity estimation within the occlusion area by combining the local depth maps with occlusion boundaries.

sured by the SPO on an EPI $I_\alpha(l, r)$. The initial depth estimation is calculated by:

$$\Theta_\alpha(l, r) = \arg\max_\theta d_\alpha(l, r, \theta), \tag{8}$$

where $\alpha \in$ A, A represents the optional angular set of EPI. And θ is the angle corresponding to the maximum response $\Theta_\alpha(l, r)$, which reflects the depth of the point in the center view. It means that we can get the depth map for the reference view $\Theta_\alpha(l, r)$ after calculating $d_\alpha(l, r, \theta)$ in the α- orientation slice image.

We compute the local cost volumes of some specific points in Fig. 5, where the maximum response value corresponds to the correct depth label, and the cost volumes of different EPIs are shown.

4.2. Occlusion boundary detection

As shown above, a scene point may be occluded in some EPIs and unoccluded in others, hence the local disparity values obtained from different EPIs are different. After the local depth estimation stage, several depth labels are acquired for a point and the occlusion points are calculated based on these labels. We define a predictor P_{occ} to measure whether a point in the center view is occluded by combining cues from various EPIs and depth.

Both the occlusion and the various EPIs have directivity. In occlusion regions, the depth labels obtained from different EPIs have large differences, as shown in Fig. 5(c) and (d). In contrast, if the point is not occluded, the depth values from all of the EPIs should be the same, as shown in Fig. 5(b). We use the variance of the depth values estimated based on different EPIs for a scene point to determine the possibility of the point being occluded. In order to reduce noise and obtain more robust results, we choose the eight-point neighborhood system to compute the degree of depth difference. In this way, we obtain an initial occlusion boundary,

$$P_{occ}^{\text{var}} = g\left(\sum_{d\in\Lambda} \frac{V_d}{\overline{d}}\right). \tag{9}$$

where Λ represents all of the depth labels for points in the eight-point neighborhood, V_d is the variance of the set of depth values, $\overline{d}$ is the average, and $g(\cdot)$ is a robust clipping function that saturates the response larger than a certain threshold. We divide the variance by $\overline{d}$ to increase robustness.

In addition, occlusion areas are related to the gradient of the local depth, or the boundaries in depth map. Hence we estimate the possibility of the current point being at an occlusion using P_{occ}^d as:

$$P_{occ}^d = \prod_{\alpha\in\text{A}} g\left(\frac{\nabla d_\alpha}{d_\alpha + \delta}\right), \tag{10}$$

where d_α is the initial depth obtained from EPI $I_\alpha(l, r)$. δ is the compensation value to d_α, since for regions with small depth values, the depth change across pixels becomes larger as the depth gets larger. In order to increase robustness, we determine a point as an occlusion point only when it is predicted as an occlusion point based on all initial depth maps.

Finally, we compute the combined occlusion response P_{occ} by

$$P_{occ} = P_{occ}^{\text{var}} \cdot P_{occ}^d. \tag{11}$$

In this section, we apply the compass operator on the occlusion boundary maps to obtain the edge orientations. The compass operator [32] splits a circle into two semicircles with an edge as the diameter, and the possible edge orientation is directly detected from the separated two sides among the different orientations with the

Please cite this article as: H. Sheng et al., Occlusion-aware depth estimation for light field using multi-orientation EPIs, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.09.010

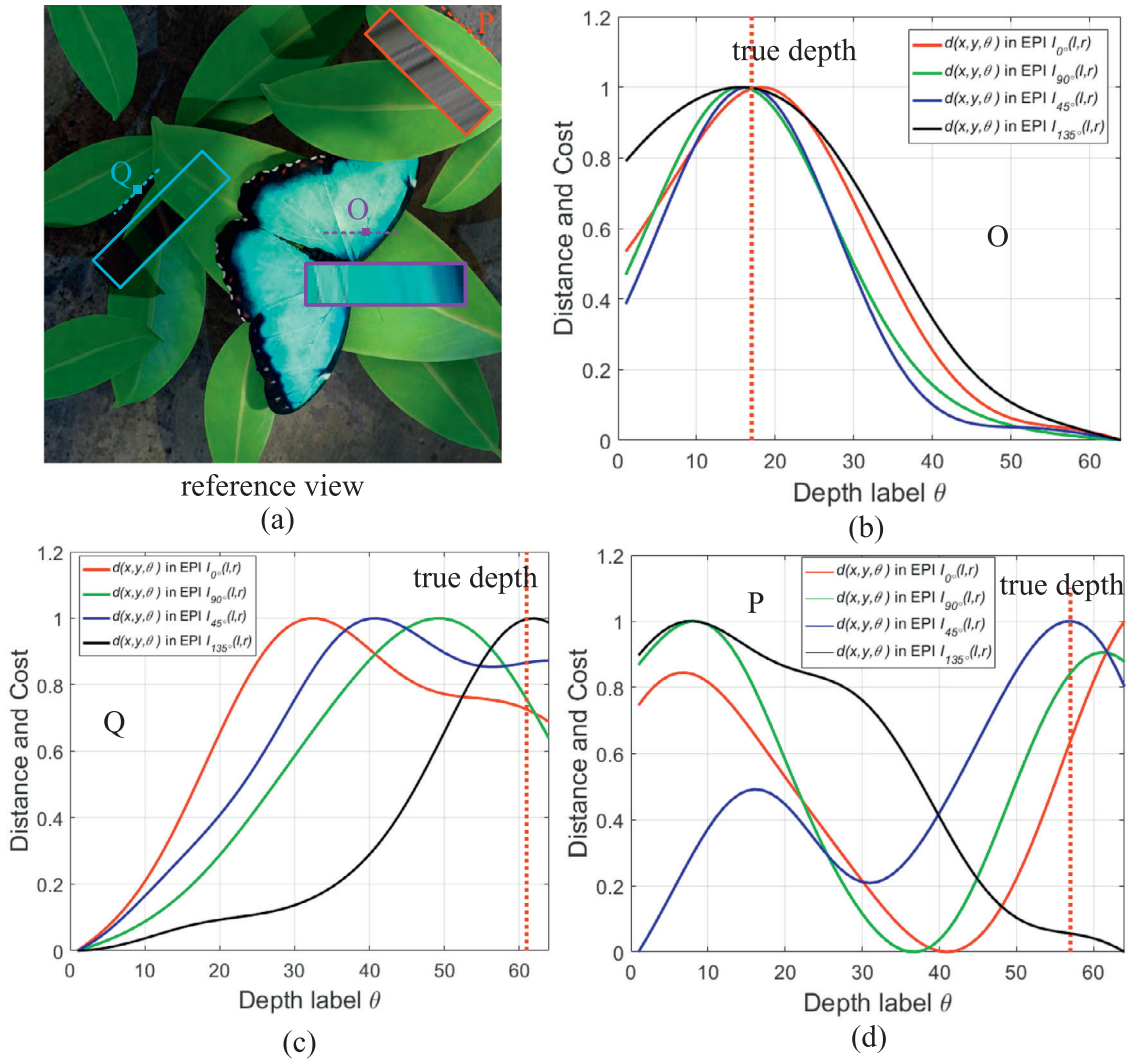

Fig. 5. Cost volumes of different EPIs. The EPI parallel to the occlusion direction tends to be considered as the optimal EPI for local depth calculation. The proposed method obtains the right depth labels for point P and Q using the 45°-orientation and 135°-orientation EPIs, respectively.

maximum response. The different occlusion boundary results are shown in Fig. 6.

4.3. Depth integration

We improve the depth estimation in occlusion regions by selecting the disparity value from the optimal EPI according to occlusion orientation. The key idea is that we choose the EPI that contains non-occluded views to calculate the cost volumes for each pixel in the reference view. The SPO maintains the correct distance information by picking up the maximum distance between the two regions. However, if heavy occlusions exist, the distance may not have the maximum value at the correct depth label. The proposed method extracts EPIs of various angles for a point in the center view, which provides more choices to obtain the optimal EPI for occlusion points in order to calculate the accurate depth value.

For example, points P and Q in Fig. 5 are occluded in some views, which causes errors when we use only horizontal or vertical EPIs. Since the occlusion edge of point P is close to 45°, the SPO produces the right depth label through the 45°- orientation EPI, the same as point Q through 135°- orientation EPI. For the occlusion points, the EPI parallel to the occlusion direction tends to be used as the optimal EPI for local depth calculation. In general, the proposed method uses the optimal EPI which is composed of non-occluded views to maintain the maximum value at the accurate depth label.

When occlusion occurs, the optimal cost volume is set as

$$d(u, v, \theta) = d_{\alpha}(l, r, \theta), \tag{12}$$

where $d(u, v, \theta)$ is the optimal cost volumes of point (u, v) on the center view. Angle $\alpha \in \mathrm{A}$ is closest to the orientation of the occlusion boundary.

For the points that are non-occluded, we adopt a confidence metric introduced by Zhang et al. [15], which reduces the influence of ambiguities. The confidence c is defined as:

$$c = \exp\left(-\frac{\overline{d}/d_{\max}}{2\sigma^2}\right), \tag{13}$$

where $d_{\max} = \max_{d} d(\theta)$, $\overline{d} = \sum_{\theta} d(\theta)$, and the term σ controls the sensitivity of the confidence. The computed cost volumes at each depth label are obtained by combining local cost volumes at different angles:

$$d(u, v, \theta) = \sum_{\alpha \in \mathrm{A}} c_{\alpha} d_{\alpha}(l, r, \theta), \tag{14}$$

where $d(u, v, \theta)$ is the combined cost volumes, and c_{α} is the confidence calculated by using EPI $I_{\alpha}(l, r)$.

Please cite this article as: H. Sheng et al., Occlusion-aware depth estimation for light field using multi-orientation EPIs, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.09.010

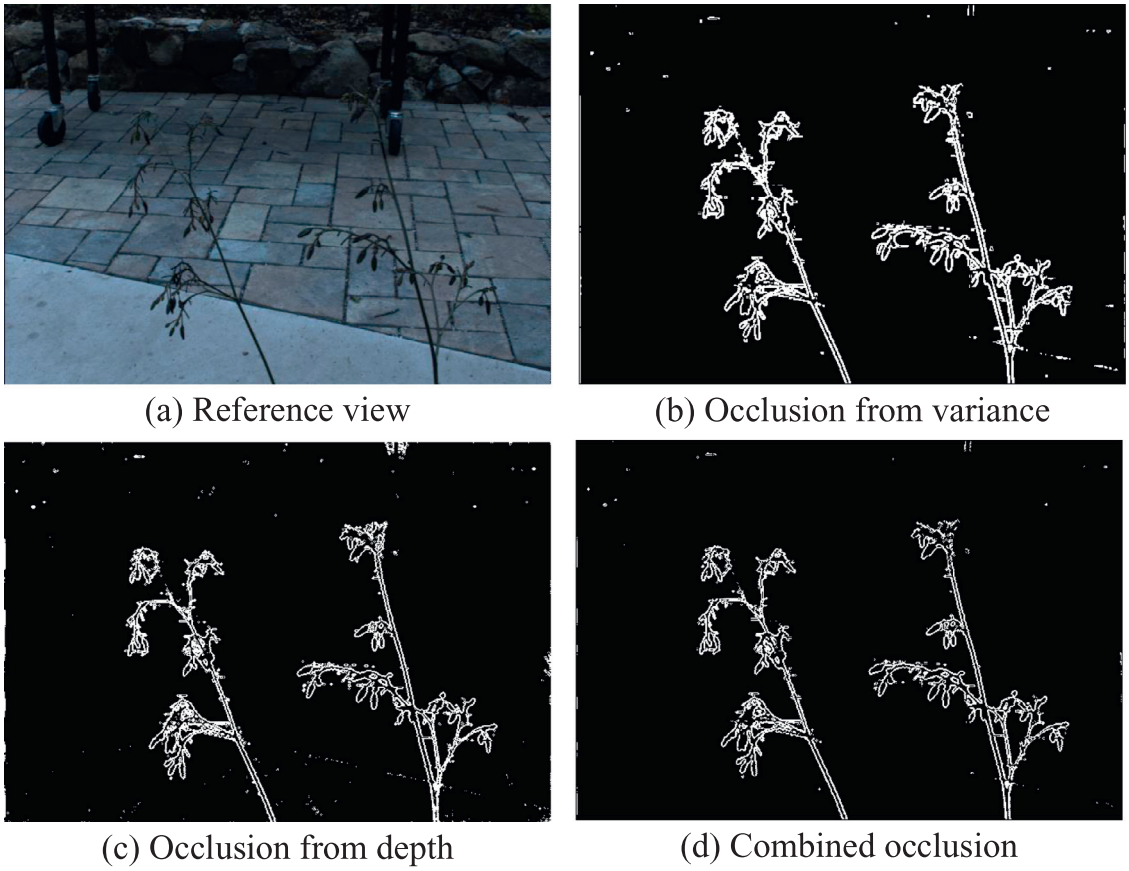

Fig. 6. Occlusion boundaries predicted on real scenes. (d) shows that we obtain a better estimation of occlusion edges by taking into account both the variance and gradient of the local depth maps.

Finally, cost volume filtering is used to improve the accuracy of the depth map. For Lytro images with lots of noises and aliasing, the graph cut is used for further optimization after guided filter.

Our framework replace SPO with the structure tensor [14]. Specifically, structure tensor is run on all the EPIs in multiple directions to estimate local depth maps and reliability, which can calculate final depth maps in a global optimization framework.

5. Experimental results

In this section, we show experimental results using the proposed depth estimation framework, and verify the effectiveness of the proposed occlusion prediction and depth integration method.

Datasets: The images we used in the experiments are divided into two parts. The synthetic light field images are collected from Wanner et al. [33], Wang et al. [28] and Honauer et al. [34] with available ground truth. The real light field images captured by Lytro Illum camera [28] are also evaluated.

Compared methods: In this paper, the performance of our occlusion-aware algorithm is mainly compared with those of Chen et al. [9] (BCM), Wanner et al. [14] (tensor structure based), Zhang et al. [15], Wang et al. [24] (occlusion-aware depth estimation), and Johannsen et al. [35] which have shown good performance on light field images recently. The methods designed for plenoptic light field camera images, Tao et al. [10], Jeon et al. [36], are also illustrated for a comparison.

Evaluation methods: As a quality measurement, we use the percentage of depth value below a relative error based on the ground truth (5% in our experiments as [14]). The accuracy for the overall (All) and occlusion regions (Occ) is calculated respectively. The *MSE* * *100* and *BadPix*(0.07) metrics are also used to evaluate the depth maps in the benchmark [34].

We choose $\alpha = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ in our experiments to calculate multi-orientation EPIs.

5.1. Occlusion boundary comparisons

In this section, we first show the estimated occlusion boundaries of different stages using our algorithm, as shown in Fig. 7. We obtain a more accurate occlusion edge by combining the gradients of four local depth maps. The variance of the depth labels for points in the eight-point neighborhood takes into account not only the differences in the depth values obtained with other EPIs, but also the differences between the depth values of the surrounding points. The result of Wang et al. [24] depends on the result of Canny edge detection on the central view. For some images, such as "Striped" in Fig. 8, Canny edge detection can lead to a smaller candidate set of occlusion points.

5.2. Optimal EPI selection

The image "Stripes" in the benchmark [34] has a special structure, whose occlusion direction is vertical, which is more conducive to verify the properties of the optimal EPI. As shown in Fig. 8, the depth map calculated by EPI $I_{90^\circ}(l, r)$ achieves lower *MSE* and *BadPix* than other EPIs, since the direction of the optimal EPI is the same with occlusion boundaries. It proves that the direction of the optimal EPI is parallel to the boundary of the occlusion. Our final depth map achieves the most accurate result.

Please cite this article as: H. Sheng et al., Occlusion-aware depth estimation for light field using multi-orientation EPIs, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.09.010

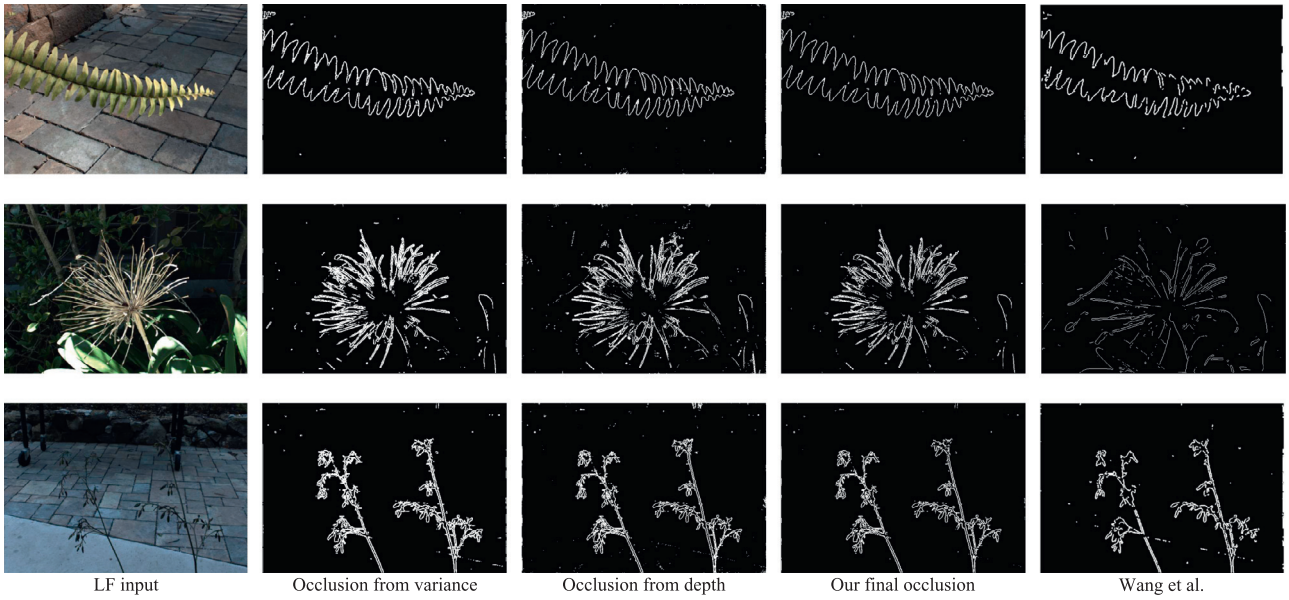

Fig. 7. Occlusion prediction results. The occlusion boundary is calculated using the variance and gradient of the estimated depth maps. Our final occlusion boundary is the combination of them. Compared with Wang et al. [24], our result is more accurate.

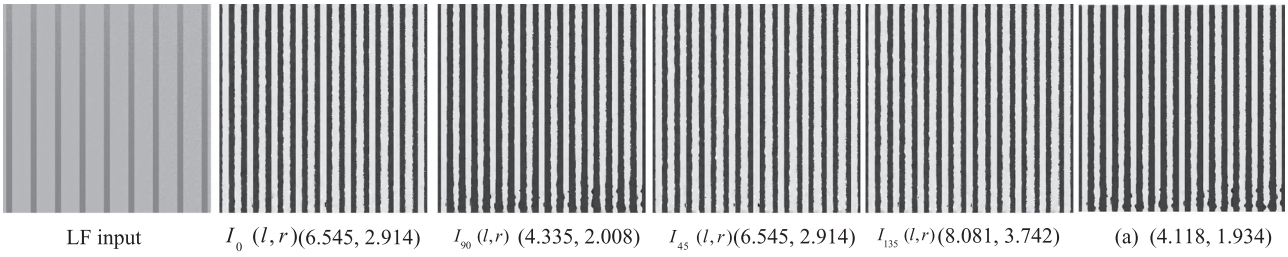

Fig. 8. The depth maps based on different EPIs on image "Stripes". The *MSE* and *BadPix* are illustrated in the bracket. The depth map calculated by the optimal EPI $I_{90^\circ}(l,r)$ achieves lower *MSE* and *BadPix* than other EPIs, and our final depth map achieves the most accurate result as shown in (a).

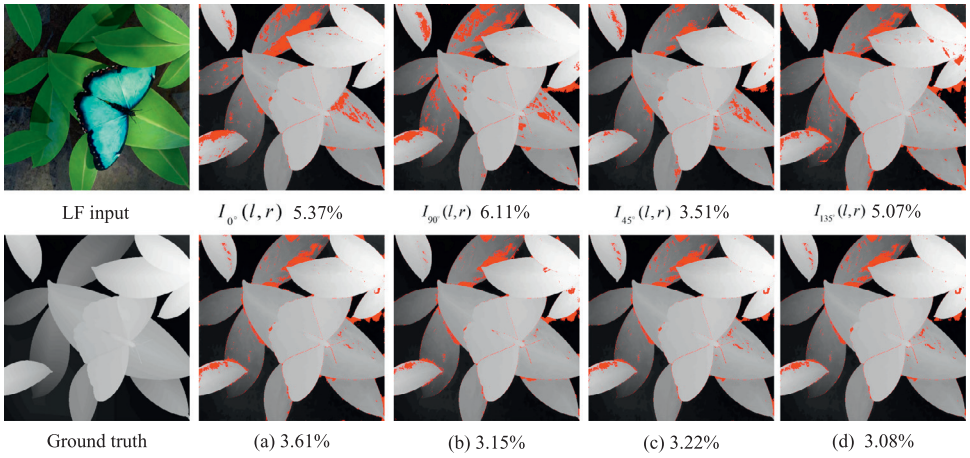

Fig. 9. The evaluation of different depth integration methods. The error pixels are marked red for distinct comparison in image "Papillon". The first row shows depth maps calculated by EPI $I_{0^\circ}(l,r)$, EPI $I_{90^\circ}(l,r)$, EPI $I_{45^\circ}(l,r)$ and EPI $I_{135^\circ}(l,r)$, respectively. (a) is the simple summation, (c) reflects the depth integration only based on confidence. (b) and (d) take the occlusion boundary into account on the basis of (a) and (c), respectively. The merged depth maps (d) based on our method achieves the lowest error rate.

5.3. Depth maps integration

To demonstrate the effectiveness of the proposed combination, we show detailed visual and quantitative results from the different depth map fusion methods, as shown in Fig. 9. The depth maps computed from four different EPIs are combined into a final depth image. To the points which are unoccluded, we choose the weight of each EPI based on their confidence. For occlusion regions, we utilize the winner-take-all strategy to select the EPI whose orientation is closest to the occlusion boundary. When utilizing the local depth information of all available EPIs for non-occluded areas

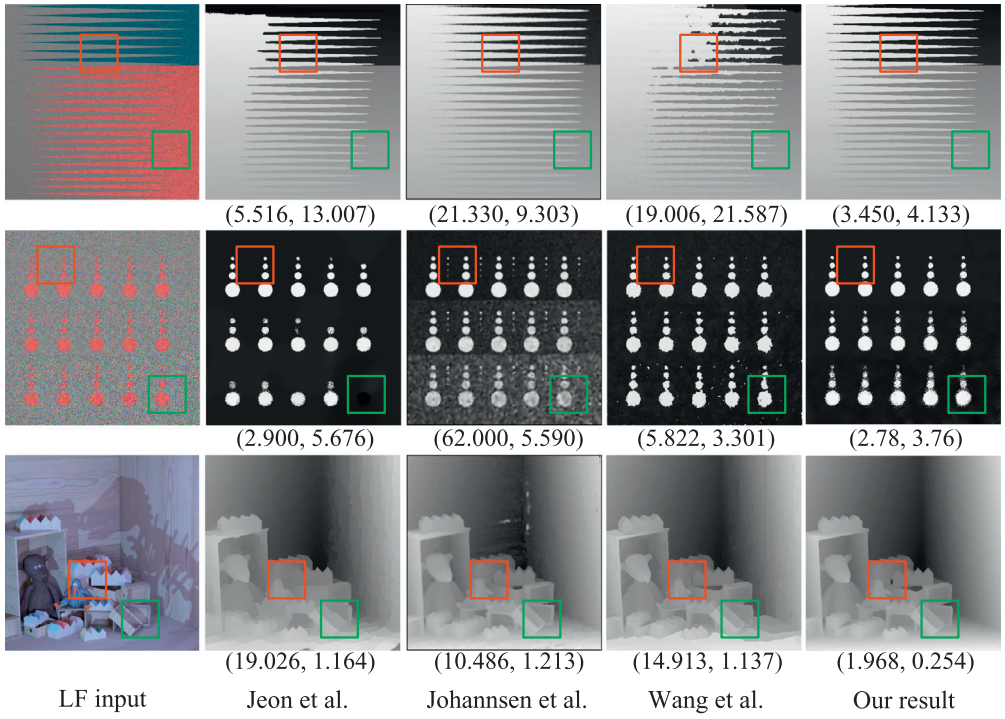

Fig. 10. The depth estimation results using synthetic light field images by Honauer et al. [34]. The three synthetic images are "Backgammon", "Dots" and "Dino" from top to bottom. Our method captures the details along the depth edges when the occlusion relationship is complex. The BadPix and MSE metrics are also illustrated in the brackets, respectively. Some occlusion regions are labeled and our depth maps show more accurate results.

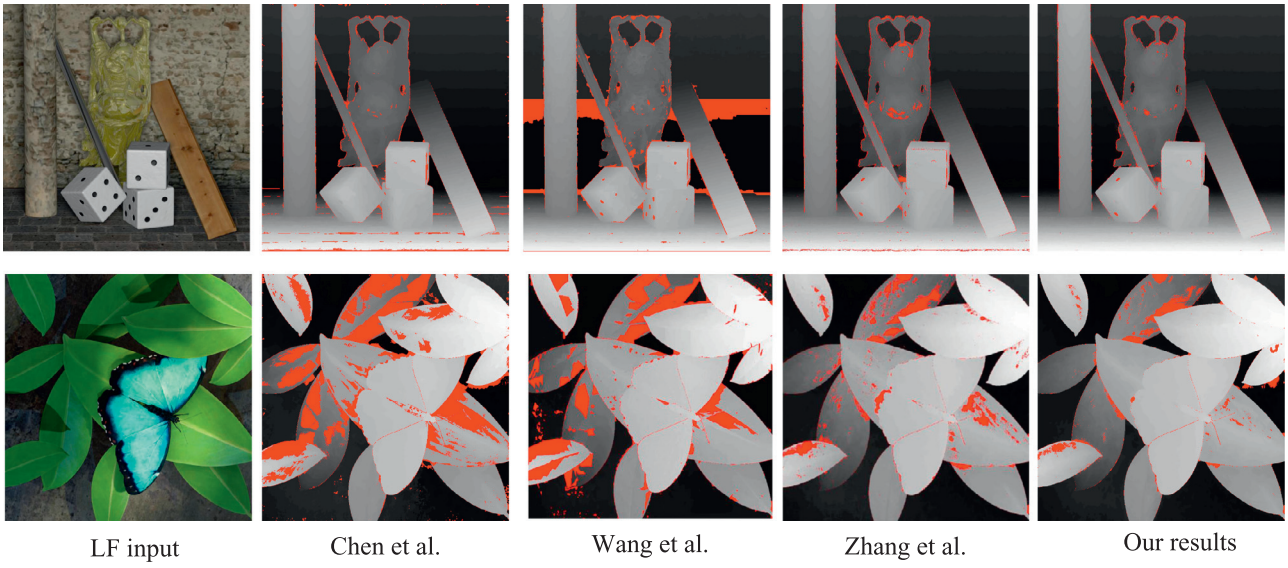

Fig. 11. Depth estimation from synthetic images "Buddha" and "Papillon".

to obtain more robust results and the optimal EPI to deal with occlusion, we can obtain a better final depth map.

5.4. Depth maps for synthetic images

The synthetic images have fewer noises and cross-talk artifacts and are more like multi-view images. The numerical results of comparing with advanced methods on synthetic light field images with 9 × 9 views [33] are illustrated in Table 1, which summarizes the accuracies of the proposed method and compares with the results of the state-of-the-art techniques for the whole images (ALL) and occlusion regions (Occ). The best results are displayed in bold. The commonly used $MSE * 100$ and $BadPix(0.07)$ metrics used in the benchmark [34] are also illustrated in Fig. 10.

The detailed comparisons on discontinuous depth regions with state-of-art methods are shown in Figs. 10–12. The "Dots" scene in the new benchmark [34] are designed to show the effects of camera noises. Our method can better handle noise by the combination of more local depth values, as shown in Fig. 10. These synthetic images in Fig. 10 also contain many occlusions. Compared with Jeon et al. [36], Johannsen et al. [35] and Wang et al. [24], our results have less noise in image "Dots" and clearer edges in images "Backgammon" and "Dino". The complete results have been submitted to Honauer et al. [34], where our results are ranked in the top few.

Please cite this article as: H. Sheng et al., Occlusion-aware depth estimation for light field using multi-orientation EPIs, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.09.010

Fig. 12. Depth estimation result on synthetic dataset by Wanner et al. [14] and Wang et al. [24]. In the first row, we obtain the clear boundaries of the leaves, thin shape of the chandelier in the second row, and accurate shapes of the droplight and headboard in the third row. Our method also successfully captures the small occluder in the bottom row.

Fig. 11 compares results for synthetic light field image "Papillon" and "Buddha" (9 × 9 views) with ground truth. Both images contain a lot of occlusions. Chen et al. [9] produced many errors in ambiguous regions, because only half pixels of the angular sampling images are taken into account for the matching process. But the excluded half pixels may be very useful for ambiguous regions. Wang et al. [24] used Canny edge detection and dilated the boundaries detected in the center view, such that pixels that have a distance from the boundary may be considered as occlusions candidates. Zhang et al. [15] only used two EPIs to merge into a single depth map, while the depth map obtained from EPI $I_{45^\circ}(l, r)$ sometimes achieves lower error rate compared with EPI $I_{0^\circ}(l, r)$ and EPI

Please cite this article as: H. Sheng et al., Occlusion-aware depth estimation for light field using multi-orientation EPIs, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.09.010

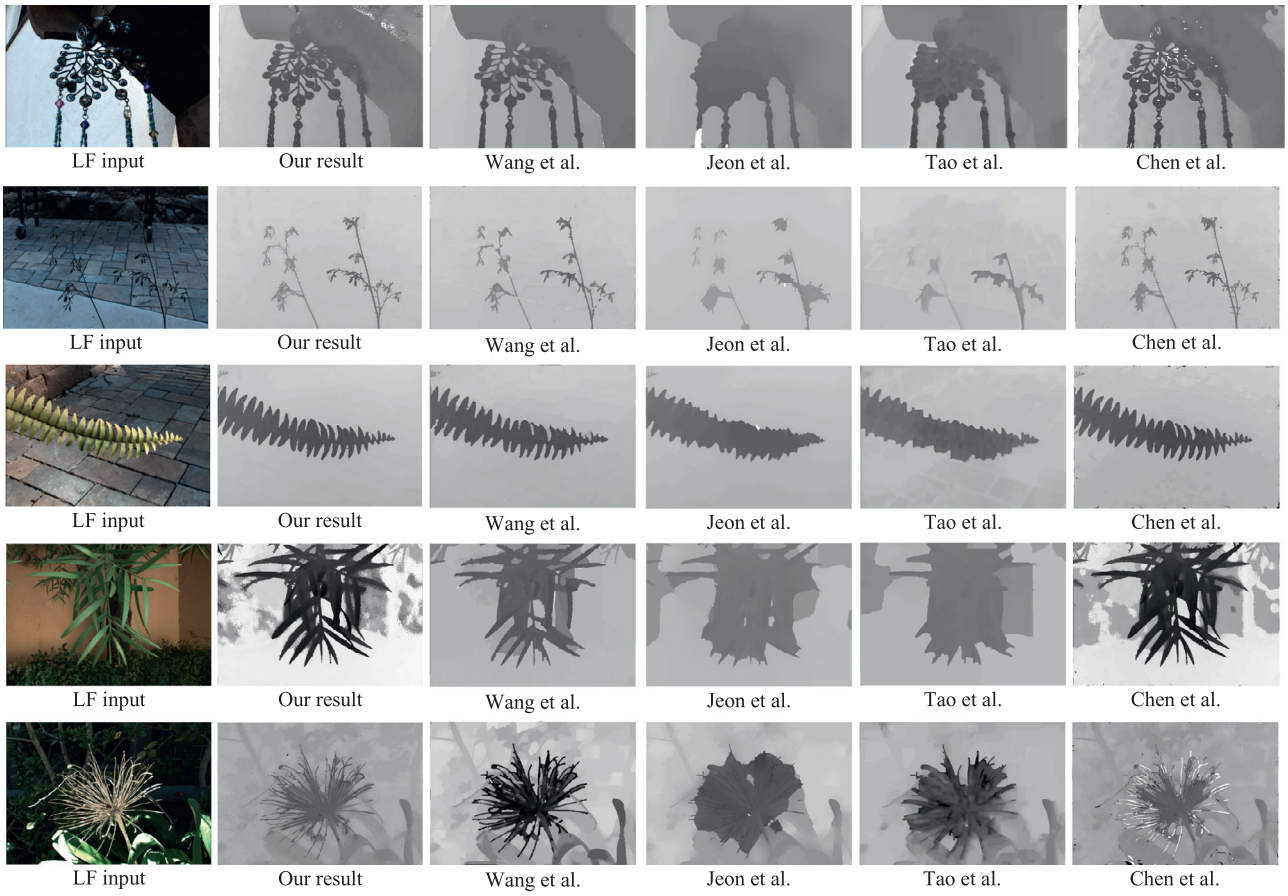

Fig. 13. Depth estimation of Lytro Illum images provided by Wang et al. [24]. Our method successfully obtains the accurate shapes of each object in the images, while other methods fail, or generate weaken results.

Table 1
The error rate of the estimated depth compared with ground truth (%).

Image	Wanner et al. [14]		Chen et al. [9]		Zhang et al. [15]		Wang et al. [28]		Ours	
	Overall	Occlusion	Overall	Occlusion	Overall	Occlusion	Overall	Occlusion	Overall	Occlusion
Cube(9)	0.92	13.29	1.11	9.72	0.98	9.96	1.44	11.46	**0.86**	**8.99**
Buddha(9)	2.41	15.01	1.72	8.34	1.50	7.99	6.203	13.524	**1.29**	**7.29**
Stilllife(9)	4.30	8.35	1.51	6.53	1.61	5.81	3.99	12.79	**1.29**	**5.12**
Papillon(9)	19.24	24.44	12.86	12.83	5.02	7.50	10.9	17.9	**3.08**	**7.18**

$I_{90^\circ}(l, r)$ as shown in Fig. 8. Hence the overall and occlusion performance of Zhang et al. [15] is worse than the proposed method.

Fig. 12 shows some close-up views for a more clear comparison on the synthetic dataset by Wanner et al. [14] and Wang et al. [24]. We can observe that Chen et al. [9] produced some errors in ambiguous regions, and Wang et al. [24] failed to handle occlusions problem when the occlusion relationship becomes complex. On the contrary, the proposed method used more accurate occlusion boundaries to handle occlusion problems, and utilized the surrounding regions to approach the accurate depth to effectively deal with ambiguity problems. It can be seen that our results show fewer errors in occluded areas. We obtained more clear boundaries of the leaves in the top row, thin shape of the chandelier in the second row, and accurate shapes of the droplight and headboard in the third row. Our method also successfully captured the small occluder in the bottom row.

5.5. Depth maps for real images

In this section, the performance of our method on Lytro Illum camera images is evaluated. The real images obtained from the Lytro camera are more noisy than synthetic images. Combined with the occlusion boundaries and Multi-orientation EPIs, the proposed method is able to maintain more details. Moreover, the integration of local depth estimations makes it more robust to noise. Fig. 13 compares the depth maps on real scenes with occlusions. Compared with other methods, our depth maps have more clear boundary features and fully reflect the structure of the object in the scene. Our method produced the complicated structure of the strap in the first row, especially captured the small basket holes, and properly obtained the thin edges of the stem, plant, leaves and flower.

Please cite this article as: H. Sheng et al., Occlusion-aware depth estimation for light field using multi-orientation EPIs, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.09.010

6. Conclusion

In this paper, we proposed a novel occlusion-aware depth estimation method based on multi-orientation EPIs. A strategy to extract Epipolar plane images on all available directions is developed by taking into account the special structure of light field images. Occlusion boundaries are better predicted by combining the variance and gradient of local depth maps. We also combined the local depth with occlusion orientation in order to treat occlusion effectively. The proposed algorithm improved the accuracy of the final depth map by using cost volume filtering. Compared with the state-of-the-art light field depth estimation methods, the proposed method achieves the higher accuracy in depth maps. Experimental results show that our method performs very well in both synthetic and real scenes, especially near occlusion boundaries.

Acknowledgments

This study is partially supported by the National Key R&D Program of China (No. 2017YFC0806502), the National Natural Science Foundation of China (No. 61472019), the Macao Science and Technology Development Fund (No. 138/2016/A3), the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under grant # SKLSDE-2017ZX-09. The authors would like to thank the support from HAWKEYE Group.

References

[1] R. Ng, Lytro redefines photography with light field cameras, (http://www.lytro.com).
[2] C. Perwass, L. Wietzke, Single lens 3d-camera with extended depth-of-field, in: IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2012, p. 829108.
[3] S.J. Gortler, R. Grzeszczuk, R. Szeliski, M.F. Cohen, The lumigraph, in: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, ACM, 1996, pp. 43–54.
[4] M. Levoy, P. Hanrahan, Light field rendering, in: ACM Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, 1996, pp. 31–42.
[5] T.E. Bishop, S. Zanetti, P. Favaro, Light field superresolution, in: Proceedings of the IEEE International Conference on Computational Photography (ICCP), IEEE, 2009, pp. 1–9.
[6] S. Wanner, B. Goldluecke, Spatial and angular variational super-resolution of 4D light fields, in: European Conference on Computer Vision (ECCV), Springer, 2012, pp. 608–621.
[7] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, P. Hanrahan, Light field photography with a hand-held plenoptic camera, Comput. Sci. Tech. Rep. 2 (11) (2005) 1–11.
[8] E.H. Adelson, J.Y. Wang, Single lens stereo with a plenoptic camera, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 99–106.
[9] C. Chen, H. Lin, Z. Yu, S.B. Kang, J. Yu, Light field stereo matching using bilateral statistics of surface cameras, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1518–1525.
[10] M.W. Tao, S. Hadap, J. Malik, R. Ramamoorthi, Depth from combining defocus and correspondence using light-field cameras, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2013, pp. 673–680.
[11] J. Sun, N.-N. Zheng, H.-Y. Shum, Stereo matching using belief propagation, IEEE Trans. Pattern Anal. Mach. Intell. 25 (7) (2003) 787–800.
[12] R.C. Bolles, H.H. Baker, D.H. Marimont, Epipolar-plane image analysis: an approach to determining structure from motion, Int. J. Comput. Vision 1 (1) (1987) 7–55.
[13] A. Criminisi, S.B. Kang, R. Swaminathan, R. Szeliski, P. Anandan, Extracting layers and analyzing their specular properties using epipolar-plane-image analysis, Comput. Vision Image Understanding 97 (1) (2005) 51–85.
[14] S. Wanner, B. Goldluecke, Globally consistent depth labeling of 4D light fields, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 41–48.
[15] S. Zhang, H. Sheng, C. Li, J. Zhang, Z. Xiong, Robust depth estimation for light field via spinning parallelogram operator, Comput. Vision Image Understanding 145 (2016) 148–159.
[16] T.E. Bishop, P. Favaro, Plenoptic depth estimation from multiple aliased views, in: Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), IEEE, 2009, pp. 1622–1629.
[17] M.W. Tao, P.P. Srinivasan, J. Malik, S. Rusinkiewicz, R. Ramamoorthi, Depth from shading, defocus, and correspondence using light-field angular coherence, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1940–1948.
[18] S. Wanner, B. Goldluecke, Variational light field analysis for disparity estimation and super-resolution, IEEE Trans. Pattern Anal. Mach. Intell. 36 (3) (2014) 606–619.
[19] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, J. Yu, Line assisted light field triangulation and stereo matching, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2013, pp. 2792–2799.
[20] S. Heber, T. Pock, Shape from light field meets robust PCA, in: European Conference on Computer Vision (ECCV), Springer, 2014, pp. 751–767.
[21] I. Tosic, K. Berkner, Light field scale-depth space transform for dense depth estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2014, pp. 441–448.
[22] M. Bleyer, C. Rother, P. Kohli, Surface stereo with soft segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 1570–1577.
[23] V. Kolmogorov, R. Zabih, Multi-camera scene reconstruction via graph cuts, in: European Conference on Computer Vision (ECCV), Springer, 2002, pp. 82–96.
[24] T.-C. Wang, A.A. Efros, R. Ramamoorthi, Occlusion-aware depth estimation using light-field cameras, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3487–3495.
[25] W. Williem, I.K. Park, Robust light field depth estimation for noisy scene with occlusion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4396–4404.
[26] S.B. Kang, R. Szeliski, J. Chai, Handling occlusions in dense multi-view stereo, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, IEEE, 2001, pp. I–103.
[27] V. Kolmogorov, R. Zabih, Multi-camera scene reconstruction via graph cuts, in: European Conference on Computer Vision (ECCV), Springer, 2002, pp. 82–96.
[28] T.-C. Wang, A. Efros, R. Ramamoorthi, Depth estimation with occlusion modeling using light-field cameras, IEEE Trans. Pattern Anal. Mach. Intell. 38 (11) (2016) 2170–2181.
[29] O. Johannsen, A. Sulc, B. Goldluecke, Occlusion-aware depth estimation using sparse light field coding, in: German Conference on Pattern Recognition, Springer, 2016, pp. 207–218.
[30] R.C. Bolles, H.H. Baker, D.H. Marimont, Epipolar-plane image analysis: an approach to determining structure from motion, Int. J. Comput. Vision 1 (1) (1987) 7–55.
[31] D. Scharstein, R. Szeliski, R. Zabih, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, in: Stereo and Multi-Baseline Vision (SMBV), IEEE, 2001, pp. 131–140.
[32] M.A. Ruzon, C. Tomasi, Color edge detection with the compass operator, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, IEEE, 1999, pp. 160–166.
[33] S. Wanner, S. Meister, B. Goldluecke, Datasets and benchmarks for densely sampled 4D light fields, in: Vision, Modeling & Visualization, The Eurographics Association, 2013, pp. 225–226.
[34] K. Honauer, O. Johannsen, D. Kondermann, B. Goldluecke, A dataset and evaluation methodology for depth estimation on 4D light fields, in: Asian Conference on Computer Vision (ACCV), Springer, 2016, pp. 19–34.
[35] O. Johannsen, A. Sulc, B. Goldluecke, What sparse light field coding reveals about scene structure, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3262–3270.
[36] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, I.S. Kweon, Accurate depth map estimation from a lenslet light field camera, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1547–1555.
[37] H. Sheng, S. Zhang, X. Cao, Y. Fang, Z. Xiong, Geometric Occlusion Analysis in Depth Estimation using Integral Guided Filter for Light-Field Image, IEEE Transaction on Image Processing, 2017.

Hao Sheng received his B.S. and Ph.D. degrees from School of Computer Science and Engineering, Beihang University, Beijing, China, in 2003 and 2009, respectively. He is currently an associate professor in School of Computer Science and Engineering, Beihang University (e-mail: shenghao@buaa.edu.cn). His research interests include computer vision, pattern recognition and machine learning.

Pan Zhao received his B.S. degree from School of Transportation Science and Engineering, Beihang University, Beijing, China, in 2014. He is currently pursuing the M.S. degree in School of Computer Science and Engineering, Beihang University (e-mail: zhaopan@buaa.edu.cn). His research interests include light field image processing, computer vision and machine learning.

Shuo Zhang received the B.S. degree from Schoolof Information Engineering, Zhengzhou University, Henan, China, in 2012. She is currently pursuing the Ph.D. degree in School of Computer Science and Engineering, Beihang University, Beijing, China (e-mail: shuo.zhang@buaa.edu.cn). From 2015 to 2016, she was a visiting student incomputer science and artificial intelligence laboratory, Massachusetts Institute of Technology (MIT), Cambridge. Her research interests include machine learning, computer vision and computational photography.

Jun Zhang recieved his M.S. and Ph.D. both in electrical engineering from Rensselaer Polytechnic Institute in 1985 and 1988, respectively. Now he is a professor at the Department of Electrical Engineering and Computer Science in University of Wisconsin-Milwaukee (e-mail: junzhang@uwm.edu). His research interests include image processing and computer vision, signal processing and digital communications.

Da Yang received his B.S. degree from School of Computer Science and Engineering, Beihang University, Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree (e-mail: da.yang@buaa.edu.cn). His research interests include machine learning and computer vision.

Please cite this article as: H. Sheng et al., Occlusion-aware depth estimation for light field using multi-orientation EPIs, Pattern Recognition (2017), http://dx.doi.org/10.1016/j.patcog.2017.09.010