Robust Light Field Depth Estimation over Occluded and Specular Regions

Xuechun Wang, Wentao Chao, and Fuqing Duan ✉

School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China
fqduan@bnu.edu.cn

Abstract. Traditional methods for light field depth estimation establish the cost data to measure the photo consistency of pixels refocused into a specific depth range, with the highest level of consistency indicating the correct depth. These methods are based on the photo consistency of Lambertian surface. However, the photo consistency is broken when occlusion and specular reflection occur. In this paper, a new depth estimation algorithm is proposed to solve the problem that the photo consistency is broken. Firstly, the central view image is segmented into multiple superpixel regions. The cost ranges of the un-occluded points and occluded points in the refocusing process are analyzed, and a penalty term is added to the pixel whose color deviation exceeds an adaptive threshold to detect the occluded points. Because the un-occluded pixels in the angular sampling image still keeps the photo consistency, we propose a voting method to select the un-occluded pixels to obtain the initial depth of the occluded point. We use a method to determine the specular region based on similar features of color and texture in the superpixel region and then present an optimization energy function to obtain the depth of the specular region. Finally, a more accurate depth map is obtained by using a globally optimization. Experimental results show that the proposed method is superior to other comparison algorithms, especially in the cases of the specular regions and multi-occlusion.

Keywords: Light field · depth estimation · occlusion detection · specluar reflection · 3D reconstruction

1 Introduction

Different from traditional cameras, light field [12] can record spatial and angular information simultaneously by a single shot, and it has been used in microscopic scene [28] for life sciences, the temperature distribution reconstruction [32] and 3D reconstruction [7]. Since a single light field image has multi-view information, it can be applied to surveillance security [33] and fake face recognition [8]. Depth information is essential to accomplish these tasks. At present, many depth estimation methods have been proposed using different light field image formats, such as stereo matching in sub-aperture images [6, 4], slope estimation in epipolar plane image (EPI) [30, 29] and focusness estimation in focal stacks [2, 31]. These methods are based on the property of Lambertian surface that the image of a point obtained from different views should satisfy the photo consistency. However, occlusions and specular reflections are always present in real

scenes. The photo consistency assumption will be broken on the occlusions or specular reflection regions, which results in inaccurate depth estimation on these regions. Therefore, it is very important to accurately handle occlusion and specular reflection problems for depth estimation. Both the occluded and specular regions violate the photo consistency, so we denote the occluded and specular regions as non-photo consistency regions (NPCR).

The previous works [22, 21, 34] always used the Canny detector to detect edge points in the central view image and identified these edge points as occlusion points. However, some occlusion points in non-center view images will be missed. These methods are hard to detect occlusion correctly, which limits obtaining an accurate depth map. For the specular regions, the previous works [19, 27] estimated an initial depth map based on the assumption of the photo consistency, and each point was refocused to the estimated initial depth. The angular sampling image was clustered into two clusters to remove the specular reflection component for obtaining the final depth. However, the clustering result will be affected by the error of the initial depth, and it is not suitable for the saturated specular points. In addition, it is time-consuming to process each point in the central view image. At present, it is still challenging to well solve the NPCR problem.

In this paper, we propose a novel depth estimation method which can well deal with the NPCR problem. Our main contributions have four aspects, including

- The central view image is segmented into multiple superpixel regions. The NPCR can be determined based on the characteristic of superpixel which can save more time.
- An adaptive threshold is determined by combining the spatial domain and angular domain. We introduce a penalty term for the pixel whose color deviation exceeds the adaptive threshold to detect the occluded points.
- We propose a voting method to choose the un-occluded pixels from the angular sampling image of the occluded point to estimated the initial depth.
- We use the chrominance similarity among pixels in the same superpixel region to detect specular regions. Then, we develop an energy function to compute the depth of the specular region.

2   Related Work

The main reasons that cause the photo consistency of Lambertian surface to be broken are occlusion and specular reflection. A lot of algorithms have been proposed to solve the two problems separately.

Occlusion problem. Williem et al. [25, 26] defined the constrained angular entropy cost and the constrained adaptive defocus cost to estimate the initial depth map. Zhang et al. [14] proved that if the depth of occlusion is constant in the angular sampling image, the occlusion boundary in the angular sampling image is similar to that of the reference image. Therefore, they estimated the depth map by developing an integral guided filter to estimate the occlusion probabilities in angular sampling images. Han et al. [2] estimated the depth map by counting the number of pixels whose deviation from the refocused pixel to the central view pixel was less than an adaptive threshold. Although they used

the adaptive threshold to reduce the influence of the occlusion, the threshold will be affected in the real scene with severe noise. Chen et al. [1] detected partially occluded boundary regions based on superpixel segmentation and applied shrinkage operation on the label confidence map and reinforcement operation on the occluded boundary to estimate the depth map. But the whole algorithm process takes a long time. Wang et al. [22, 21] demonstrated the occlusion edges in both spatial and angular domains are consistent when refocused to the proper depth. They divided the angular domains into two regions with the guidance of the spatial domain, and the depth is estimated using the region with less variance. Their method is limited to a single occluder and is highly dependent on edge detection. Zhu et al. [34] improved the occlusion model of Wang et al. [22, 21], and proposed a model suitable for both single-occluder and multi-occluder occlusions. They selected the un-occluded pixels using the K-means strategy to estimate the depth map. However, it is time-consuming and is not always effective when multiple occluders are distributed at different depths.

All of these methods use Canny detector to detect the occlusion points in the central view image, but the occlusion points in the other views may be neglected. Different from these methods, we analyze the range of data cost and add a penalty term to a pixel whose color difference exceeds the adaptive threshold to detect the occluded points. We choose the un-occluded pixels in the angular sampling image of the occluded point according to the voting method using an adaptive threshold.

Specular reflection. Wang et al.[19] used the depth estimation method by Tao et al.[17] to obtain initial depth. Each point is refocused to the initial depth to classify its angular sampling image into two clusters using k-means clustering. Light source was estimated using the difference between two clusters through several iterations to remove the specular reflection component. Wang et al. [27] used a threshold strategy to classify the points in the central view image into saturated specular points and unsaturated specular points, after estimating initial depth using the method by Tao et al. [17]. For unsaturated specular points, they used k-means clustering to divide the angular sampling image into the part without specular and the part with specular. The specular components will be removed using the difference between the two clusters. These methods all use initial depth obtained based on photo consistency to remove specularity. Moreover, all these methods can not handle the saturated specular points. Wang et al. [18] proposed point-consistency to estimate the initial depth and line-consistency to estimate light source color through analyzing the dichromatic reflection model. They estimated the specular intensity using light source color and removed the specular components. Unlike these methods, we utilize the characteristic of superpixel and energy function to estimate the depth of the specular reflection region.

Deep learning methods. There are also some learning based methods for light field depth estimation. Shin et al. [15] used a multi-stream convolutional neural network to estimate accurate depth maps with subpixel accuracy. However, their network fails to infer accurate disparities in reflection and textureless regions. Tsai et al. [20] fused multi-view information using an attention-based views selection module and used a spatial pyramid pooling to extract more context information of the image for depth estimation. Huang et al. [5] designed a lightweight disparity estimation model with physical-based multi-disparity-scale cost volume aggregation to estimate depth map. It
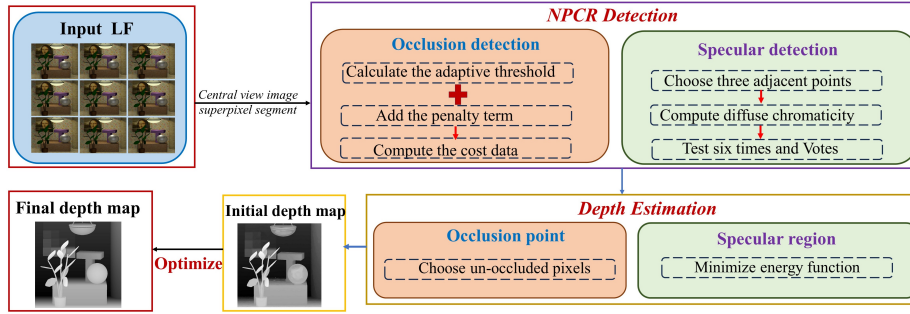
Fig. 1. The proposed algorithm overview. The proposed method mainly consists of four parts: Segment the central view image using SLIC algorithm; Detect the NPCR including occlusion and specular regions; For the occlusion, the un-occluded pixels in the angular sampling image are used to compute the initial depth; For the specular regions, the depth is estimated according to the depth of the adjacent non-specular regions; Depth optimization.

significantly improved the geometric details near edges by introducing a sub-network of edge guidance. Wang et al. [23] constructed an occlusion-aware cost by modulating pixels from different views and integrated pixels by using the convolutions with specifically designed dilation rates. They aggregated these costs constructed via convolutions to estimate the depth map.

The above methods only consider the occlusion or specular regions, so they cannot handle the two cases well at the same time. In this paper, we unify the two cases into a NPCR problem, and propose a new depth estimation algorithm to solve this problem better.

3   The Depth Estimation

In this section, we first segment the central view image using the SLIC superpixel algorithm. The initial depth map is estimated based on the characteristic of the superpixel region. The final depth map is obtained after optimization. The complete algorithm flow chart is shown in Fig. 1.

3.1   Consistency Data and Confidence

A 4D light field image can be represented as $L(x, y, u, v)$, where (x, y) means the spatial coordinate in the sub-aperture image, and (u, v) means the angular coordinate. According to the imaging characteristic of the light field, the light field image can be refocused to a new imaging plane [12]. Therefore, the refocused light field image $L_\alpha(x, y, u, v)$ is represented as

$$L_\alpha(x, y, u, v) = L(x + u(1 - \frac{1}{\alpha}), y + v(1 - \frac{1}{\alpha}), u, v) \tag{1}$$

where α is the candidate depth label.

According to the photo consistency of Lambertian surface, the pixels in the angular sampling image should have the same color when a point is refocused to the correct depth. We calculate the color deviation between the pixels in the central view and the non-central view in different candidate depths using Eq. 2.

$$d_\alpha(x, y, u, v) = \|L_\alpha(x, y, u, v) - L(x, y, 0, 0)\| \tag{2}$$

where $L(x, y, 0, 0)$ means the central view pixel of the angular sampling image, and $\|\cdot\|$ means L2-norm.

From Eq. 2, we calculate the mean and the variance of the color deviations to measure the consistency of the angular sampling image in different candidate depths,

$$D_\alpha(x, y) = \frac{1}{N} \sum_{u,v} \rho(d_\alpha(x, y, u, v)) \tag{3}$$

and

$$C_\alpha(x, y) = \frac{1}{N - 1} \sum_{u,v} \{\rho(d_\alpha(x, y, u, v)) - D_\alpha(x, y)\}^2 \tag{4}$$

where N is the number of pixels in the angular sampling image. $\rho(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}$ is the distance function, and σ controls the sensitivity of the function to large color difference.

The initial depth α^* of point (x, y) can be obtained using Eq. 5.

$$\alpha^*(x, y) = \arg\min_\alpha W_\alpha(x, y) \tag{5}$$

where $W_\alpha(x, y) = D_\alpha(x, y) + C_\alpha(x, y)$. The depth confidence is defined using the mean and minimum of data costs $W_\alpha(x, y)$ among all the candidates as Eq. 6.

$$w(x, y) = 1 - \exp(-\frac{mean\{W_\alpha(x, y)\}/\min\{W_\alpha(x, y)\}}{2\delta_w^2}) \tag{6}$$

where δ_w controls the sensitivity of ratio.

3.2   NPCR Depth Estimation

The occluded region is a small part of the image, and the un-occluded pixels in the angular sampling image still satisfy the photo consistency. Moreover, the specular region is a small part of most images, except when most of the objects in the image are metal or transparent. Therefore, we need to detect the occluded points and specular regions to reduce computational time.

A superpixel is a small region consisting of a series of adjacent points with similar features such as color, brightness, and texture. We use the SLIC superpixel segmentation algorithm [13] to segment the central view image. In the experiments, the size of superpixel segmentation is set as 15. We take the light field image *Mona* as an example shown in Fig. 2 (a). From Fig. 2 (a), we can find that the light field image is divided to many tiny regions, and each region is a superpixel.
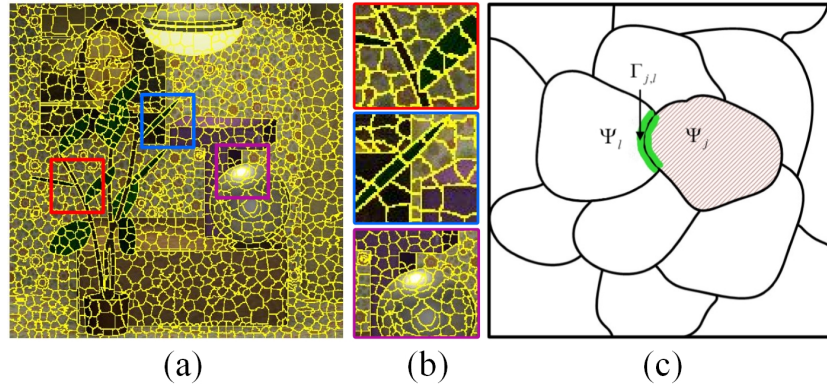
Fig. 2. (a) The superpixel segmentation result of the light field image *Mona*. (b) Zoom in on some small regions that contains occlusion and specular for a clearer observation. (c) The illustration of boundary points set $\Gamma_{j,l}$ between two adjacent superpixels Ψ_j and Ψ_l.

To more clearly show the relationship of adjacent superpixels and their boundary points, a concise illustration is shown in Fig. 2 (c). These superpixels are noted as $\Psi_j, j = 1, 2, ..., n$. It is not difficult to find that the occluded points only exist at the boundary of two adjacent superpixel regions, as shown in the green region $\Gamma_{j,l}$ in Fig. 2 (c). We define these pixels as the boundary region by dilating the edge of two adjacent superpixels. From Fig. 2, the occluded points are located at the boundary of the superpixel and in the region near the boundary of the superpixel, and the specular points are clustered in a superpixel. In conclusion, we identify occluded points and specular points from the superpixel boundary to the interior. It can reduce the time of detecting occluded points and specular regions.

Occluded Points If there exist occluded pixels in the angular sampling image, the photo consistency will be broken. It is key that the occluded points should be detected accurately. There is an important prior condition that the occluder and the occluded point have a large color difference.

We take the light field image *Mona* as an example, and the illustration is shown in Fig. 3. From Fig. 3 (a), the photo consistency of the angular sampling image is the highest when the un-occluded point is focused at the correct depth, and the corresponding data cost is also the minimum of all data costs. From Fig. 3 (c), the un-occluded pixels in the angular sampling image of the green point have the highest photo consistency when the occluded point is focused at the correct depth. However, the corresponding data cost is not the minimum among all data costs. In addition, from Fig. 3 (a) and (c), we can observe that the ranges of data cost between the occluded point and the un-occluded point are different.

For a more intuitive comparison, we plot the cost curves of the occluded points and the un-occluded points in the same graph. The illustration is shown in Fig. 4. There are four points in Fig. 4 (a), which are un-occluded points **A, B** and the occluded points **C,**
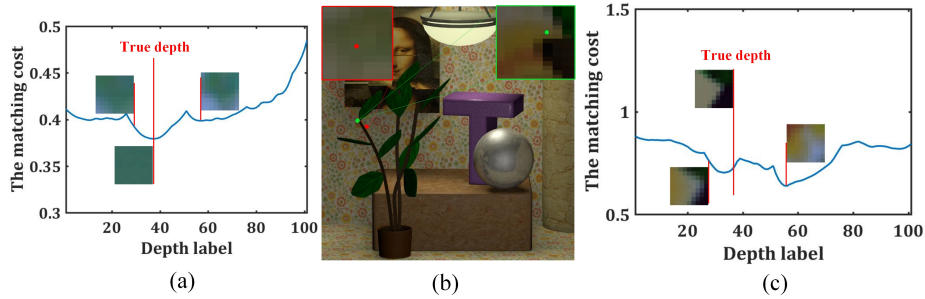
Fig. 3. The cost curves of the occluded point and the un-occluded point in the light field image *Mona* focused on different depth. (a) The cost curve of the red point. (b) The red and green rectangles are the spatial sampling images of the two points in the central view image. (c) The cost curve of the green point.
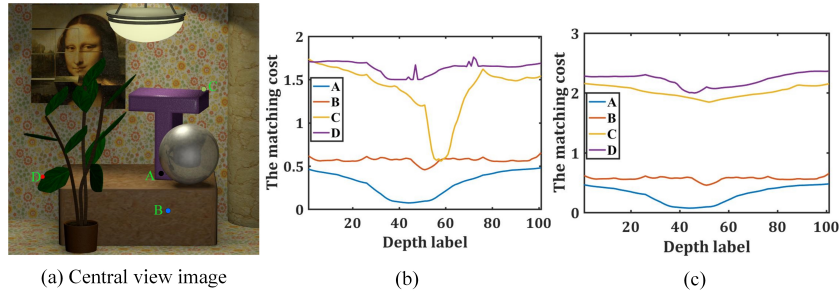
Fig. 4. The trend of the data cost of the four points in the light field image *Mona*. (b) The cost curves of four points A, B, C and D. (c) The cost curves of four points A, B, C and D after adding the penalty term.

D. From Fig. 4 (b), we can find that the maximum cost $W_{\max}$ and the minimum cost $W_{\min}$ of the points **C, D** are greater than those of the points **A, B**. We can identify the occluded point according to the minimum cost. However, the difference between the minimum cost of point **C** and the minimum cost of points **A, B** is small. The reason is that there are fewer occluded pixels in the angular sampling image of point **C**.

To avoid the impact of this case on determining occlusion, we add a penalty weight to pixels that the color deviation $d_\alpha(u, v)$ is too large. We propose an adaptive threshold $\Delta\varepsilon$ to control the color deviation $d_\alpha(u, v)$, which can increase the impact of the occluded pixel on the data cost. The new color deviation $\tilde{d}_\alpha$ can be obtained using Eq. 7, if the color deviation $d_\alpha(u, v)$ is greater than the adaptive threshold $\Delta\varepsilon$.

$$\tilde{d}_\alpha(u, v) = d_\alpha(u, v) + \varsigma(u, v) \tag{7}$$

where $\varsigma(u, v) = |d_\alpha(u, v) - \Delta\varepsilon|$ is the penalty term.

Because different occlusion situations have different color differences, we use an adaptive threshold to judge whether the pixel in the angular sampling image is from

the occluder. In Wang's method [22], they proved that the occlusion edge of the spatial domain is consistent with that of the angular sampling image when refocused to the correct depth. We build the local square window W_{UV} centered at $p = (x, y)$ in the central view image, and $U * V$ is same as the angular resolution of the light field image. We compute the color difference $E(p)$ between point p and each point $q = (s, t)$ in the local square window W_{UV},

$$E(p) = \|L(s, t, 0, 0) - L(x, y, 0, 0)\|, q \in W_{UV} \tag{8}$$

To avoid the effect of large color differences caused by occlusion or overly small color differences between point p and its adjacent points, we sort these color difference values and define them as a set A_{UV}. We take these values in the middle third of set A_{UV}. According to Eq. 9, the average of these color differences is calculated as the adaptive threshold.

$$\Delta\varepsilon = \frac{1}{k} \sum_{i=k+1}^{2k} A_i \tag{9}$$

where $k = \left\lceil UV/3 \right\rceil$.

We can obtain the new cost curves of the points **A, B, C, D** from Eq. 7 as shown in Fig. 4 (c). From this figure, we can find that the difference between the minimum cost of point **C** and the minimum cost of points **A, B** is greater than before. We can use Eq. 10 and the threshold δ_1 (set as 0.5 in the experiments) to determine whether the point $p = (x, y)$ is the occluded point,

$$t(p) = f \left\{ [W_{\min}(p) < \delta_1] \right\} \tag{10}$$

If $t(p) = 1$, the point p is an occluded point.

When a point is determined as the occluded point. It is obvious that the pixels from the occluder no longer satisfy the photo consistency with the central pixel in the angular sampling image. To obtain the accurate depth of the occluded point, we need to choose the un-occluded pixels from the angular sampling image. The color deviation $d_\alpha(u, v)$ of each pixel (u, v) in different label α is calculated using Eq. 2. If the color deviation $d_\alpha(u, v)$ is greater than the adaptive threshold $\Delta\varepsilon$, we set 0 as the label of this pixel (u, v). Otherwise, we set 1 as the label of this pixel (u, v).

$$label^\alpha_{(u,v)} = \begin{cases} 1, & d_\alpha(u, v) \leq \Delta\varepsilon \\ 0, & d_\alpha(u, v) > \Delta\varepsilon \end{cases} \tag{11}$$

Using Eq. 11, we can obtain all the labels of pixel (u, v) in the angular sampling image. The sum of these labels of pixel (u, v) is noted as $s(x, y, u, v)$. If $s(x, y, u, v)$ is more than half of the total number of candidate depth labels, the pixel (u, v) is un-occluded. Therefore, the un-occluded pixels can be accurately chosen using a threshold ξ (set as 60 in the experiments). That is

$$occ(u, v) = \begin{cases} 0, & s(x, y, u, v) \leq \xi \\ 1, & s(x, y, u, v) > \xi \end{cases} \tag{12}$$

From Eq. 12, the pixel can be determined as an un-occluded pixel when $occ(u, v) = 1$. Therefore, the initial depth of the occluded point can be determined using these un-occluded pixels in the angular sampling image by Eq. 5.

Specular Region As shown in the purple rectangle of Fig. 2 (b), the specular region is a small part of the image. The point p in the specular region can be described using the dichromatic reflection model [13],

$$I_c(p) = m_d(p)\Lambda_c(p) + m_s(p)\mathrm{T}_c(p) \tag{13}$$

where $c \in \{R, G, B\}$ denotes the color channels. $\Lambda_c(p)$ and $\mathrm{T}_c(p)$ represent diffuse and specular reflection components, respectively. $m_d(p)$ and $m_s(p)$ are the corresponding weight factors, respectively.

The diffuse components of adjacent diffuse points in the same superpixel region are very close. In addition, because of the small proportion of the specular component in the diffuse point, the weight factor $m_s(p)$ is very small. As shown in Fig. 2 (c), if there are three adjacent diffuse points p, p_1 and p_2 in the superpixel Ψ_j, $\Lambda_c(p)$, $\Lambda_c(p_1)$ and $\Lambda_c(p_2)$ should be approximately equal. For the two points p and p_1, we have

$$I_c(p_1) - I_c(p) \cong [m_d(p_1) - m_d(p)]\Lambda_c(p_1, p) \tag{14}$$

where $\Lambda_c(p_1, p)$ means either of $\Lambda_c(p)$ and $\Lambda_c(p_1)$, because $\Lambda_c(p)$ and $\Lambda_c(p_1)$ are approximately equal.

Combining the three color channels, we have

$$\sum_c [I_c(p_1) - I_c(p)] = [m_d(p_1) - m_d(p)] \sum_c \Lambda_c \tag{15}$$

Let $\Lambda_R + \Lambda_G + \Lambda_B = 1$, Eq. 15 can be rewritten as

$$\sum_c [I_c(p_1) - I_c(p)] = m_d(p_1) - m_d(p) \tag{16}$$

Combining Eq.14 and Eq. 16, we can compute the diffuse chromaticity of p and p_1 in the superpixel Ψ_j,

$$\Lambda_c(p_1, p) = \frac{I_c(p_1) - I_c(p)}{\sum_c [I_c(p_1) - I_c(p)]} \tag{17}$$

Similarly, $\Lambda_c(p_2, p)$ can be also obtained. If the superpixel Ψ_j is not a specular region, the difference in chromaticity values should be very small. Therefore, we have $\Delta\kappa = |\Lambda_c(p_1, p) - \Lambda_c(p_2, p)|$. If $\Delta\kappa$ is larger than a predefined threshold (set as 0.1 in the experiments), the three points p, p_1, and p_2 are voted as the specular points. Because the point p has four adjacent points, each point in the superpixel Ψ_j is tested six times. If the number of votes for the point is greater than 4, then the point is a specular point. If the number of points with votes greater than 4 is more than half of the total points in the boundary region of Ψ_j, the superpixel Ψ_j is a specular region.

Let M_j be the set of all adjacent superpixels of the specular reflection region Ψ_j, and $\Gamma_{j,l}$ be the set of boundary points between Ψ_j and $\Psi_l(\Psi_l \in \mathrm{M}_j)$. Therefore, the depth

of the specular reflection region Ψ_j is computed by minimizing the following energy function,

$$\sum_{p \in \Psi_j} w(p) \| \psi(j) - \alpha^*(p) \|^2 + \lambda \sum_{\Psi_l \in M_j} \sum_{q \in \Gamma_{j,l}} \frac{\| \psi(j) - \psi(\Psi_l) \|^2}{|\nabla I(q)|} \tag{18}$$

where $\nabla I(q)$ is the gradient of the boundary point $\Gamma_{j,l}$, $\alpha^*(p)$ is the initial depth of point p, and λ (set as 0.05 in the experiments) is a weight of the depth of the adjacent superpixel. $\psi(\Psi_l)$ is the weighted mean depth of the superpixel Ψ_l. It is calculated using the following equation,

$$\psi(\Psi_l) = \frac{\sum\limits_{t \in \Psi_l} \alpha^*(t) \big/ \min\{W_\alpha(t)\}}{\sum\limits_{t \in \Psi_l} 1 \big/ \min\{W_\alpha(t)\}} \tag{19}$$

In order to avoid the depth difference of adjacent superpixel regions in the specular region being too large, which will affect the depth accuracy of the specular region, a depth difference threshold is added when optimizing the energy function to discard the unreliable adjacent region. The depth of $\psi(j)$ can be solved efficiently using Eq. 18 within a weighted least squares framework [16].

3.3   Depth Refinement

We optimize the initial depth map using the global optimization. The energy function is

$$E = E_1(p, \alpha^*(p)) + \lambda_s E_2(p, q, \alpha^*(p), \alpha^*(q)) \tag{20}$$

where q is the adjacent point of point p, λ_s is a weight factor (set as 5 in the experiments). The unary term E_1 is defined as Eq. 21

$$E_1 = \sum_p w(p) \| \alpha(p) - \alpha^*(p) \|^2 \tag{21}$$

The smooth regularization term E_2 is defined as Eq. 22

$$E_2 = \sum_{p,q} \frac{\exp(-\frac{(\alpha(p) - \alpha(q))^2}{2\delta^2})}{|\nabla I(p) - \nabla I(q)| + \mu |t(p) - t(q)|} \tag{22}$$

where μ is a weighting factor, $t(p)$ and $t(q)$ mean whether the two points are occlusion points or not, which can be obtained from Eq. 10. $\nabla I(q)$ is the gradient of the point q. The final depth can be derived by solving Eq. 20 using the standard graph cut algorithm [9].

4   Experiment

In this section, the proposed algorithm is evaluated on the synthetic and real light field images. The synthetic light field images are from Wang et al. [22], Wanner et al. [24]

(a)   (b)
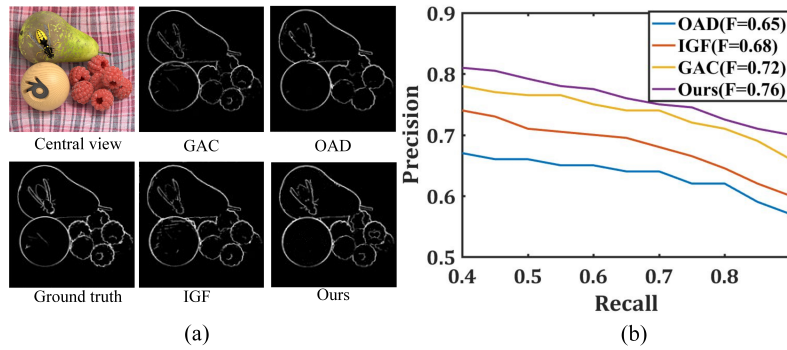
Fig. 5. (a) The visual comparisons of occlusion detection results. (b) The PR-curves of occlusion detection results using OAD [22], GAC [34], IGF [14], and Ours on the dataset from Wanner et al. [24].

and Honauer et al. [3], and the real light field images are collected using Lytro Illum camera [11] and Lytro camera [10]. We compare the proposed algorithm with traditional estimation methods including Wang et al. [22] (OAD), Zhu et al. [34] (GAC), Shen et al. [14] (IGF), Han et al. [2] (OAVC) and Wang et al. [18] (PLC), and the learning based methods including Shin et al. [15] (Epinet), Wang et al. [23] (OACC), Tsai et al. [20] (LFnet) and Huang et al. [5] (Fastnet). Because there are ground truth in synthetic light field datasets, we use the mean squared error (MSE) and badpixel(0.07) error [3] to evaluate the estimated depth results. F-measure [15] is computed to evaluate the performance of the occlusion detection methods. The values of the hyperparameters have been stated in Section 3. The values of the hyperparameters in the experiments are the ones that have been tested many times or generally recognized, and we will not describe them in detail due to the lack of space.

4.1 Occlusion Processing Comparisons

We detect the occluded points using different occlusion-aware methods including OAD [22], IGF [14], and GAC [34]. The comparison results are shown in Fig. 5 (a), which takes the light field image *StillLife* from Wanner et al. [24] as an example. From Fig. 6 (a), we can see our method can get more completed occluded points than the other three comparison methods. In addition, we quantitatively evaluate the occluded point detection method compared with OAD [22], IGF [14], and GAC [34] by computing the recall and precision to obtain the PR-curve, and the result is shown in Fig. 5 (b). The PR-curve also demonstrates the proposed method performs best. Experimental results shown in Fig. 5 demonstrate that the performance of the proposed method is better than these comparison algorithms. The reason is that the proposed method detects the occluded points according to the minimum cost of each point, while OAD [22] and GAC [34] use the Canny detector to determine the occluded points in the central view image, and the occlusions in other views may be ignored. For IGF [14], the accuracy of occluded point detection is highly dependent on the central view image. If there are too

Table 1. The badpix error of the occlusion points on the synthetic datasets [24, 22].

	Buddha	Medieval	Papillon	Buddha2	Horses	Mona	StillLife	Bedroom	Livingroom	Outdoor	Plant
Epinet[15]	18.22	27.67	43.97	58.71	23.91	22.51	25.81	16.51	13.24	32.56	21.74
OACC[23]	12.71	25.53	31.55	43.67	20.14	18.53	22.17	12.31	10.56	26.73	17.60
Fastnet[5]	16.13	28.93	35.45	55.12	22.88	21.74	24.37	15.55	13.33	30.28	20.91
LFnet[20]	14.39	26.44	37.98	50.52	21.16	23.13	25.66	15.33	14.47	28.77	19.16
OAD [22]	15.01	13.29	24.44	20.70	23.75	20.75	28.82	8.38	10.78	9.87	7.33
IGF[14]	8.34	9.96	20.83	15.79	9.62	12.67	19.63	7.42	10.75	10.85	9.57
GAC [34]	13.64	12.57	15.16	13.85	15.80	18.81	21.71	7.41	11.79	13.88	10.18
PLC [18]	21.13	19.15	25.65	27.64	23.59	22.78	28.93	49.13	22.14	32.65	27.17
OAVC[2]	7.67	12.43	8.91	**12.25**	9.96	10.71	13.32	7.35	9.95	8.87	8.89
Ours	**7.59**	**9.72**	**7.50**	13.89	**8.85**	**8.77**	**12.83**	**6.51**	**7.85**	**6.89**	**5.87**

many noisy and complex occlusions in the central view image, the detection method may fail.

(a)   (b)   (c)   (d)
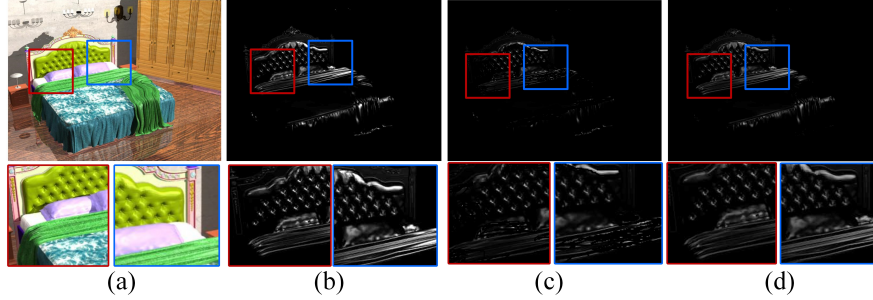
Fig. 6. The comparison results of the detection of the specular regions in the light field image *Bedroom*. (a) The central view image, (b) The ground truth, (c) Wang et al. [27], (d) Ours.

After the occluded points are determined, the accuracy of the depth estimation of the occluded points has a great influence on the final depth map. To quantitatively evaluate the performance of our occlusion detection and processing method, we calculate the badpix error of all comparison algorithms at the occluded points on the synthetic datasets [22, 24] according to the ground truth, and the results are shown in Table 1. From Table 1, we can find the error of the proposed algorithm is much lower than that of other comparison algorithms. OAD [22] can not handle the case of multiple occlusions, and GAC [34] is affected by the distribution of multiple occlusions. However, PLC [18] does not deal with occlusion in the depth estimation algorithm. The proposed method selects the un-occluded pixels in the angular sampling image using an adaptive threshold. It is not affected by different images and occlusion conditions.
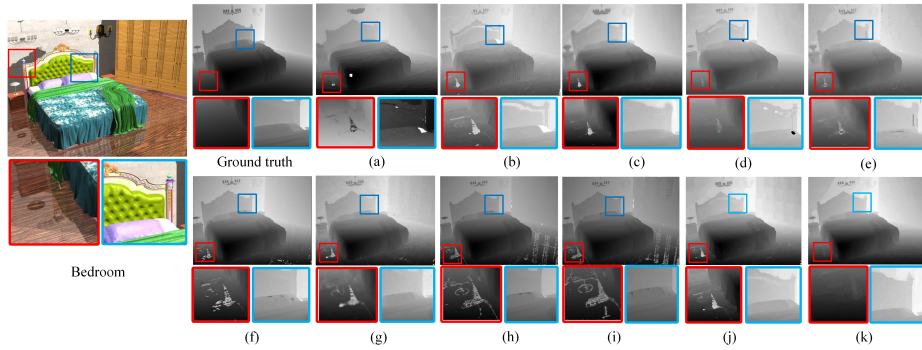
Fig. 7. The comparison results of the light field image *Bedroom*. (a) GAC [34],(b) OAD [22], (c) IGF [14], (d) OAVC [2], (e) PLC [18], (f) Epinet [15], (g) Fastnet [5], (h) LFnet [20], (i) OACC [23], (j) Ours without processing specular, (k) Ours.

4.2   Specular Regions Processing

In this subsection, we evaluate the performance of our specular region detection method compared with that of Wang et al. [27]. The detection results of image *Bedroom* are shown in Fig. 6. From Fig. 6, we can find the proposed method can detect more accurately. The reason is that the detection method of Wang et al. [27] used a threshold value strategy based on the initial depth, and the initial depth is obtained based on the assumption of the photo consistency, while the initial depth in the specular region had a large error. The proposed method uses superpixel segmentation to segment the the central view image into several tiny regions. According to the similarity of points in the same superpixel region, each point in the superpixel region is judged without the initial depth, thus improving the detection accuracy.

Table 2. The badpixel error of the points in the specular regions on synthetic datasets [24, 22].

	Buddha	Medieval	Papillon	Buddha2	Horses	Mona	StillLife	Bedroom	Livingroom	Outdoor	Plant
Epinet[15]	25.81	19.95	21.71	43.37	27.18	33.37	28.19	62.78	58.19	23.67	56.81
OACC[23]	22.11	18.98	20.13	40.17	25.57	29.19	26.76	58.18	53.27	18.87	52.98
Fastnet[5]	26.25	18.82	20.91	42.18	24.57	28.88	26.65	60.19	55.77	22.18	58.91
LFnet[20]	24.13	20.13	21.72	41.18	26.64	31.88	27.71	61.18	59.89	21.36	54.46
OAD [22]	22.13	15.57	13.37	28.97	15.75	18.16	16.56	42.88	32.18	19.97	43.17
IGF[14]	21.87	16.61	14.87	25.61	16.88	17.73	15.58	39.97	28.75	18.53	40.11
GAC [34]	23.33	14.76	15.57	26.73	15.79	18.81	14.34	38.87	26.65	17.73	38.87
PLC[18]	13.36	8.89	9.26	13.86	9.13	10.30	8.87	22.27	13.34	11.23	28.91
OAVC[2]	23.17	12.25	16.36	22.27	16.88	16.73	15.51	36.55	23.73	16.59	32.28
Ours	**9.73**	**7.92**	**7.68**	**10.40**	**7.98**	**9.57**	**10.33**	**16.85**	**9.78**	**8.81**	**18.45**

We also use the compared algorithms, the proposed method without processing the specular points and the proposed method to estimate the depth map of image *Bedroom*.

Table 3. The MSE errors of the estimated depth for the light field images from the synthetic datasets [24, 3]. The first set of images is from Wanner et al. [24], and the seconde set of images is from Honauer et al. [3] (Red means the best result in learning based methods, and blue means the best result in traditional methods.)

Image	Epinet[15]	OACC[23]	Fastnet[5]	LFnet[20]	OAD[22]	IGF[14]	GAC[34]	PLC[18]	OAVC[2]	Ours
Papillon	6.12	3.38	7.42	4.98	0.23	0.81	0.41	2.31	0.84	0.32
Medieval	2.28	1.61	3.37	0.50	0.18	1.03	0.67	1.27	0.88	0.09
StillLife	2.43	2.07	23.2	14.1	0.21	1.46	0.23	1.63	1.07	0.17
Mona	1.33	2.33	1.56	0.79	0.25	0.44	0.46	0.79	0.44	0.36
Buddha	0.36	0.78	0.42	0.33	0.38	0.59	0.29	0.64	0.36	0.19
Buddha2	6.64	4.31	5.06	6.06	1.18	0.64	1.02	2.42	1.29	0.53
Hourse	7.35	2.21	5.67	6.32	1.36	1.21	1.37	2.08	0.53	0.77
Average	3.79	2.24	6.67	4.72	0.68	0.88	0.64	1.59	0.77	0.35
Cotton	0.27	0.18	0.32	0.21	0.32	4.31	1.56	2.07	0.60	0.25
Boxes	6.09	3.32	4.39	3.99	9.85	12.32	9.42	10.31	6.99	8.18
Pyramids	0.010	0.004	0.018	0.004	0.021	0.039	0.098	0.21	0.04	0.013
Bakgmamon	2.62	2.69	1.52	3.65	3.71	5.83	6.08	4.67	3.84	2.94
Dots	2.52	1.01	3.17	1.63	3.01	3.89	5.82	3.87	16.6	2.61
Dino	0.17	0.08	0.19	0.09	1.14	0.12	0.31	0.41	0.27	0.54
Sideboard	0.83	0.54	0.75	0.53	2.30	0.13	1.02	0.99	1.05	0.32
Stripes	0.95	0.84	0.98	0.89	8.13	22.81	6.96	4.62	1.32	5.58
Average	1.68	1.08	1.42	1.37	3.56	6.18	3.91	3.39	3.84	2.55

The results are shown in Fig.7. From Fig. 7, it can be found that there are some holes in the depth maps of all the algorithms except the proposed method. Learning based methods also do not work well with specular regions. The main reason is that the learning based methods do not consider the effect of specular regions and the dataset used for training contains fewer specular regions. That means the performance of deep learning methods will be affected by the image characteristics of the training dataset, which has the generalization problems. The experimental results again prove the effectiveness of the proposed method in the specular reflection region.

In order to further test the proposed energy function for estimating the depth of the specular region, we calculate the badpix error of the estimated depth in the specular region using all depth estimation algorithms. The specular regions are detected using the proposed detection algorithm. The results are shown in Table 2. From Table 2, it can be found that compared with other estimation algorithms, PLC [18] and our method have smaller badpix errors. However, the average error of PLC [18] is higher than our method. This is mainly because PLC [18] is based on the differences between the specular components in multiple views, but the baseline between the sub-aperture images of the light field camera is too short, which results in a small difference. In addition, PLC [18] can not deal with saturated specular points. Our method uses the proposed energy function to obtain the depth of the specular region based on the depth of adjacent non-specular regions. It also can estimate the depth of saturated specular point.

Table 4. The badpix error of the estimated depth for the light field images from the synthetic datasets [24, 3]. The first set of images is from Wanner et al. [24], and the seconde set of images is from Honauer et al. [3] (Red means the best result in learning based methods, and blue means the best result in traditional methods.)

Image	Epinet[15]	OACC[23]	Fastnet[5]	LFnet[20]	OAD[22]	IGF[14]	GAC[34]	PLC[18]	OAVC[2]	Ours
Papillon	35.6	23.6	31.8	34.8	6.35	15.2	7.17	22.7	14.4	5.58
Medieval	18.8	15.4	16.6	11.7	2.22	8.04	2.13	2.49	10.9	0.78
StillLife	11.4	12.7	11.5	11.7	1.72	8.14	2.37	8.67	5.97	2.11
Mona	10.8	9.56	8.89	10.8	4.92	9.79	5.13	9.79	6.01	4.75
Buddha	1.55	3.33	2.23	2.02	7.21	6.85	2.34	3.09	1.78	1.68
Buddha2	34.8	28.6	31.4	34.2	14.4	9.35	10.2	19.1	11.7	8.68
Hourse	16.4	13.7	15.8	16.2	17.7	9.03	6.38	29.2	5.45	7.54
Average	18.5	15.3	16.9	17.4	8.92	9.48	5.12	13.7	8.03	4.44
Cotton	0.51	0.49	0.71	0.27	7.61	7.76	4.37	7.19	2.55	3.38
Boxes	6.09	13.31	18.70	11.04	9.85	12.32	9.42	24.06	16.1	12.36
Pyramids	0.29	0.11	0.61	0.19	0.34	0.41	0.57	0.44	0.83	0.22
Bakgmamon	2.29	4.81	3.75	3.13	3.49	5.15	5.21	5.48	3.12	3.15
Dots	5.36	1.65	21.06	1.43	7.49	7.98	12.41	8.84	69.1	5.85
Dino	1.29	0.97	2.41	0.85	14.91	1.94	2.18	4.08	3.94	2.65
Sideboard	4.81	3.35	7.03	2.87	18.49	2.13	9.29	8.65	12.4	7.96
Stripes	2.46	2.92	9.44	2.93	18.41	29.31	14.99	5.12	29.31	6.88
Average	2.89	3.45	7.96	2.89	10.07	8.38	7.31	7.98	17.17	5.31

4.3   Depth Map

To further evaluate the performance of the proposed algorithm, we use the different comparison algorithms to estimate the depth map on the synthetic datasets [24, 3]. We compute the MSE and badpix error to quantitatively analyze the estimated depth map. The results are shown in Table 3 and Table 4. From Table 3 and Table 4, the proposed method achieves the best result compared with those traditional algorithms. Moreover, although the deep learning methods obtain less error in Honauer et al. [3], which is the trianing dataset of these methods, they obtain large errors on Wannar's dataset [24]. To make a qualitative comparison, Fig. 8 shows the estimated depth maps of three light field images. As shown in Fig. 8, the results of the proposed method have clearer boundaries. The depth map results obtained by other compared traditional methods have more noise and fuzzy boundaries. The learning based methods get poor results in the three light field images because their methods are not trained on this dataset [24]. That means the deep learning methods have generalization problems in some other datasets.

**Table 5.** The running time of all the estimation algorithms. (unit:s)

	Epinet[15]	OACC[23]	Fastnet[5]	LFnet[20]	OAD[22]	IGF[14]	GAC[34]	PLC[18]	OAVC[2]	Ours (w/o specular)	Ours
Time	3.51	0.049	1.71	2.77	362.38	425.93	635.17	398.42	105.58	315.33	371.88
Language	Python/CUDA	Python/CUDA	Python/CUDA	Python/CUDA	Matlab/C++	Matlab/C++	Matlab/C++	Matlab/C++	Python/CUDA	Matlab/C++	Matlab/C++
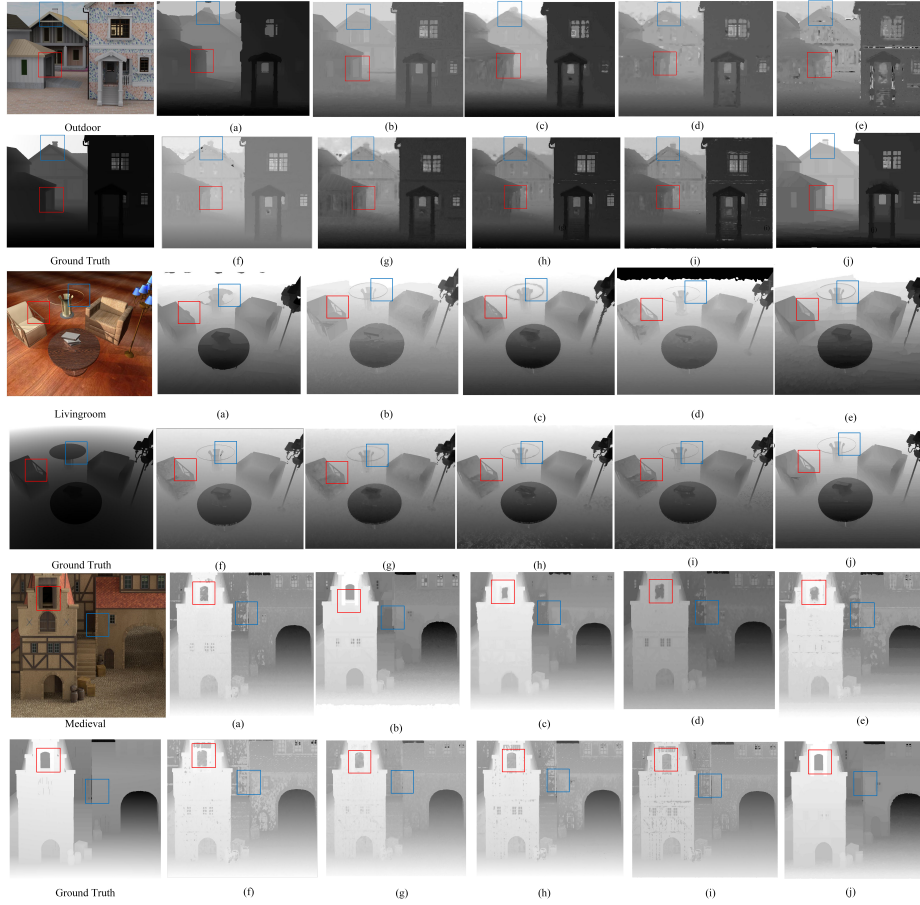
Fig. 8. Depth estimation for synthetic light field images. The first two images are from the dataset [24], and the third image is from the dataset [3]. (a) GAC [34],(b) OAD [22], (c) IGF [14], (d) OAVC [2], (e) PLC [18], (f) Epinet [15], (g) LFnet [20], (h) Fastnet [5], (i) OACC [23], (j) Ours.

We also accomplish a lot of experiments to evaluate the performance of our method in real scenes. The estimated depth maps of four images with occlusion and specular regions from the real datasets [11, 10] as shown in Fig. 9. The results of the first two images demonstrate our method is effective in occlusion modeling, whether for single occluders or multiple occluders. The results of the last two images show that our method is effective for depth estimation of the specular regions. Although PLC [18] performs better in the specular regions, the method gets too smooth occlusion boundary. The other traditional methods and deep learning methods all produce some holes in the specular regions. The experiment results verify the proposed method can handle occlusions and specular regions effectively.
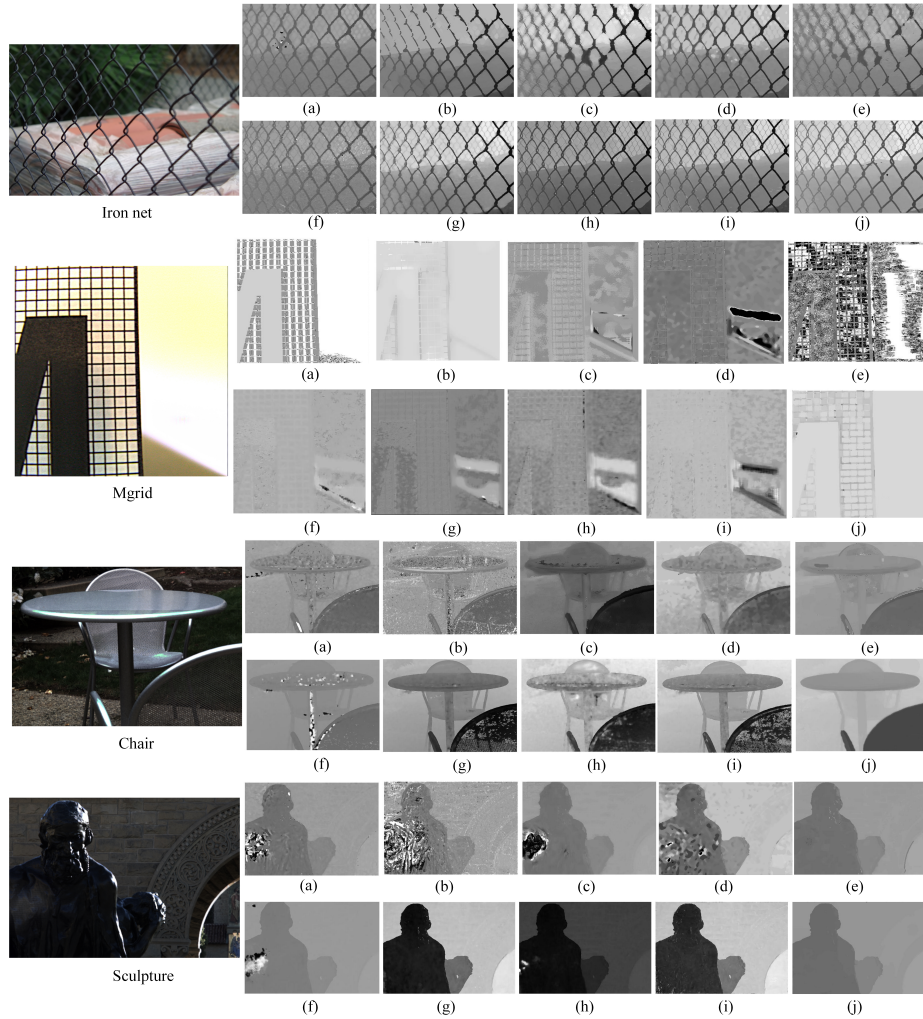
Fig. 9. Comparisons of depth maps from the real-scene images. (a) GAC [34], (b) OAD [22], (c) IGF [14], (d) OAVC [2], (e) PLC [18],(f) Epinet [15], (g) LFnet [20], (h) Fastnet [5], (i) OACC [23], (j) Ours.

4.4 Computational Time

The proposed method is implemented on a notebook computer with an Intel i5 2.50GHz CPU and 16GB RAM. The running platform is Matlab R2016b. The running time and running platforms of all comparison algorithms are shown in Table 5. Depth estimation of the specular region is performed by minimizing the energy function using the weighted least square method, which is a time-consuming step in the complete algorithm. We also show the running time with or without specular processing in Table 5.

LFnet [20] needs a long time to load the model. OAVC [2] takes less time using Python with CUDA. From Table 5, we can see that compared with traditional algorithms, the proposed method takes less time. The time taken to estimate the depth of the specular region does not affect the running speed of the whole algorithm. In general, the proposed method is faster than most of these compared algorithms.

5   Conclusion and Limitation

In this paper, we present a novel depth estimation method for light field cameras to handle occlusion and specular reflection problems. Firstly, the central view image is segmented using superpixel segmentation algorithm. The occluded point is determined according to the minimum cost value of this point after adding a penalty term to the pixel in the angular sampling image whose color difference exceeds an adaptive threshold. We propose a voting method using label accumulation to select the un-occluded pixels for depth estimation of occlusion points. The specular regions are detected using the properties that the pixels in the same superpixel region have similar diffuse chromaticity. Finally, the depth of the specular region is estimated by minimizing the energy function. The proposed method can detect the occlusion and specular regions accurately and estimate the depth of these regions accurately. Compared with the previous works, experimental results demonstrate that the proposed method has better performance on both synthetic and real light-field datasets, especially in multi-occluder occlusions and specular regions.

The proposed method can accurately estimate the depth of NPCR. However, the depth of the specular region is estimated based on the adjacent superpixel regions. If the specular region is too large or distributed extensively around the scene, the proposed method can not handle such cases. Although we can consider adjusting the parameters of the superpixel segmentation, this would make the depth map too smooth. In addition, this will increase the running time of the proposed algorithm.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, J., Hou, J., Ni, Y., Chau, L.P.: Accurate light field depth estimation with superpixel regularization over partially occluded regions. IEEE Transactions on Image Processing **27**(10), 4889–4900 (2018). https://doi.org/10.1109/TIP.2018.2839524
2. Han, K., Xiang, W., Wang, E., Huang, T.: A novel occlusion-aware vote cost for light field depth estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(11), 8022–8035 (2022). https://doi.org/10.1109/TPAMI.2021.3105523

3. Honauer, K., Johannsen, O., Kondermann, D., Goldluecke, B.: A dataset and evaluation methodology for depth estimation on 4d light fields. In: Asian Conference on Computer Vision (2016)
4. Huang, C.T.: Empirical bayesian light-field stereo matching by robust pseudo random field modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(3), 552–565 (2019)
5. Huang, Z., Hu, X., Xue, Z., Xu, W., Yue, T.: Fast light-field disparity estimation with multi-disparity-scale cost aggregation. In: 2021 IEEE/CVF International Conference on Computer Vision, ICCV. pp. 6300–6309. IEEE (2021)
6. Jeon, H.G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y.W., Kweon, I.S.: Accurate depth map estimation from a lenslet light field camera. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1547–1555 (2015). https://doi.org/10.1109/CVPR.2015.7298762
7. Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., Gross, M.: Scene reconstruction from high spatio-angular resolution light fields. Acm Transactions on Graphics **32**(4), 1 (2013)
8. Kim, S., Ban, Y., Lee, S.: Face liveness detection using a light field camera. Sensors **14**(12), 22471–22499 (2014)
9. Kolmogorov, V., Zabih, R.: What energy functions can be minimizedvia graph cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(2), 147–159 (2004)
10. Mousnier, A., Vural, E., Guillemot, C.: Partial light field tomographic reconstruction from a fixed-camera focal stack. Computer Science (2015)
11. Raj, A.S., Lowney, M., Shah, R., Wetzstein, G.: Stanford light field archives. http://lightfields.stanford.edu/ (2016)
12. Ren, N., Levoy, M., Bredif, M., Duval, G., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Stanford University Cstr (2005)
13. Shafer, S.A.: Using color to separate reflection components. Color Research & Application **10**(4), 210–218 (1985)
14. Sheng, H., Zhang, S., Cao, X., Fang, Y., Xiong, Z.: Geometric occlusion analysis in depth estimation using integral guided filter for light-field image. IEEE Transactions on Image Processing **26**(12), 5758–5771 (2017). https://doi.org/10.1109/TIP.2017.2745100
15. Shin, C., Jeon, H.G., Yoon, Y., Kweon, I.S., Kim, S.J.: Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4748–4757 (2018). https://doi.org/10.1109/CVPR.2018.00499
16. Strutz, T.: Data fitting and uncertainty: A practical introduction to weighted least squares and beyond. Springer (2011)
17. Tao, M.W., Hadap, S., Malik, J., Ramamoorthi, R.: Depth from combining defocus and correspondence using light-field cameras. In: 2013 IEEE International Conference on Computer Vision. pp. 673–680 (2013). https://doi.org/10.1109/ICCV.2013.89
18. Tao, M.W., Su, J.C., Wang, T.C., Malik, J., Ramamoorthi, R.: Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. IEEE transactions on pattern analysis and machine intelligence **38**(6), 1155–1169 (2015)
19. Tao, M.W., Wang, T.C., Malik, J., Ramamoorthi, R.: Depth estimation for glossy surfaces with light-field cameras. In: European Conference on Computer Vision. pp. 533–547. Springer (2014)
20. Tsai, Y.J., Liu, Y.L., Ouhyoung, M., Chuang, Y.Y.: Attention-based view selection networks for light-field disparity estimation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12095–12103 (2020)
21. Wang, T.C., Efros, A.A., Ramamoorthi, R.: Occlusion-aware depth estimation using light-field cameras. In: IEEE International Conference on Computer Vision (2016)

22. Wang, T.C., Efros, A.A., Ramamoorthi, R.: Depth estimation with occlusion modeling using light-field cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(11), 2170–2181 (2016). https://doi.org/10.1109/TPAMI.2016.2515615

23. Wang, Y., Wang, L., Liang, Z., Yang, J., An, W., Guo, Y.: Occlusion-aware cost constructor for light field depth estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19809–19818 (June 2022)

24. Wanner, S., Meister, S., Goldluecke, B.: Datasets and benchmarks for densely sampled 4d light fields. In: Vision, Modeling and Visualization. pp. 225–226 (2013)

25. Williem, Park, I.K., Lee, K.M.: Robust light field depth estimation using occlusion-noise aware data costs. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(10), 2484–2497 (2018). https://doi.org/10.1109/TPAMI.2017.2746858

26. Williem, W., Park, I.K.: Robust light field depth estimation for noisy scene with occlusion. In: Computer Vision and Pattern Recognition. pp. 4396–4404 (2016)

27. Xu, C., Wang, X., Wang, H., Zhang, Y.: Accurate image specular highlight removal based on light field imaging. In: 2015 Visual Communications and Image Processing (VCIP). pp. 1–4. IEEE (2015)

28. Yu, W., XiangYang, J., QiongHai, D.: Key technologies of light field capture for 3d reconstruction in microscopic scene. Science China-Information Sciences **53**(10), 1917–1930 (2010)

29. Zhang, S., Sheng, H., Li, C., Zhang, J., Xiong, Z.: Robust depth estimation for light field via spinning parallelogram operator. Computer Vision and Image Understanding **145**, 148–159 (2016). https://doi.org/https://doi.org/10.1016/j.cviu.2015.12.007

30. Zhang, Y., Lv, H., Liu, Y., Wang, H., Wang, X., Huang, Q., Xiang, X., Dai, Q.: Light-field depth estimation via epipolar plane image analysis and locally linear embedding. IEEE Transactions on Circuits and Systems for Video Technology **27**(4), 739–747 (2017)

31. Zhang, Y., Dai, W., Xu, M., Zou, J., Zhang, X., Xiong, H.: Depth estimation from light field using graph-based structure-aware analysis. IEEE Transactions on Circuits and Systems for Video Technology **30**(11), 4269–4283 (2020)

32. Zhao, W., Zhang, B., Xu, C., Duan, L., Wang, S.: Optical sectioning tomographic reconstruction of three-dimensional flame temperature distribution using single light field camera. IEEE (2) (2018)

33. Zhu, H., Wang, Q., Yu, J.: Light field imaging: models, calibrations, reconstructions, and applications. Frontiers Inf. Technol. Electron. Eng. **18**(9), 1236–1249 (2017)

34. Zhu, H., Wang, Q., Yu, J.: Occlusion-model guided antiocclusion depth estimation in light field. IEEE Journal of Selected Topics in Signal Processing **11**(7), 965–978 (2017). https://doi.org/10.1109/JSTSP.2017.2730818