

# Geometric Occlusion Analysis in Depth Estimation Using Integral Guided Filter for Light-Field Image

Hao Sheng, *Member, IEEE*, Shuo Zhang, Xiaochun Cao, *Member, IEEE*, Yajun Fang, and Zhang Xiong

**Abstract**—Unlike traditional multi-view images, sampling in angular domain of light field images is distributed in different directions. Therefore, an angular sampling image (ASI), comprising of possible matching points extracted from each view, is available for each point. In this paper, we analyze the geometric relationship between ASIs and reference sub-aperture images, and then prove the occlusion boundary similarity. Based on the geometric relationship in extreme cases, we show that some points in ASIs have higher reliability than other points for depth calculation. An integral guided filter is then built based on the sub-aperture image to predict occlusion probabilities in ASIs. The filter is independent of ASIs and has no requirement for high angular resolution so that it is easy to apply to the cost volume calculation. We integrate the filter into our depth estimation framework and other state-of-the-art depth estimation frameworks. Experimental results demonstrate that the proposed filter is more effective to occluded point detection in ASIs than other methods. Results from different data sets show that our method outperforms the existing state-of-the-art depth estimation methods, especially along occlusion boundaries.

**Index Terms**—Geometric occlusion analysis, boundary similarity, integral guided filter, depth estimation, light field.

## I. INTRODUCTION

COMMERCIALLY available light field cameras, *e.g.* Lytro [1] and Raytrix [2], are popular in recent years. The micro-lens array, which is placed in the plenoptic camera, makes it possible to capture the scene from different directions in one shot. The abundant information of the scene in light field images reveals the scene structure, and allows a wide range of applications, such as super-resolution [3]–[6], digital

Manuscript received June 26, 2016; revised February 4, 2017 and May 10, 2017; accepted August 8, 2017. Date of publication August 25, 2017; date of current version September 15, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61472019, in part by the National Key Research and Development Program of China under Grant 2017YFC0806502, in part by the Macao Science and Technology Development Fund under Grant 138/2016/A3, in part by the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under Grant SKLSD-E-2017ZX-09, and in part by the HAWKEYE Group. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gene Cheung. (*Corresponding author:* Shuo Zhang.)

H. Sheng, S. Zhang, and Z. Xiong are with the State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: shenghao@buaa.edu.cn; shuo.zhang@buaa.edu.cn; xiongz@buaa.edu.cn).

X. Cao is with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China (e-mail: caoxiaochun@iee.ac.cn).

Y. Fang is with the Computer Science and Artificial Intelligent Laboratory, Intelligent Transportation Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139-4307 USA (e-mail: yajufang@csail.mit.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2017.2745100

refocusing [7], saliency detection [8], [9], 3D reconstruction [10]–[13] and material recognition [14].

As the structure information is one of the most important cues for these applications [4], [5], [15], it is crucial to precisely estimate depth information from light field images. Therefore, different kinds of depth estimation methods have been developed based on different image formats, such as slope estimation in Epipolar Plane Image (EPI) [6], [16], stereo matching in sub-aperture images [17], [18] and focusness estimation in focal stacks [13], [19]. However, some traditional difficulties in depth estimation problems, such as occluded, noisy and textureless regions, are still difficult to solve in light field images.

As light field image has a dense sampling in angular domain, occlusion becomes an important issue for light field depth estimation. An angular sampling image (ASI) [20], which contains the possible imaging points in all sub-aperture images, has been used for occlusion analyses in depth estimation [17], [21], [22]. If the scene is Lambertian and the point is not occluded, its ASI at the correct depth is supposed to be color consistent. However, when the point is occluded in a part of the sub-aperture images, points from these images are actually from the occlusions. Therefore, how to identify possible occlusion points in ASIs is crucial for occlusion-aware depth estimation.

In this paper, the scenes are assumed to be Lambertian and we focus on the occlusion problem in light field depth estimation. We first analyze the geometric relationship between the ASI of an occluded point and the reference sub-aperture image based on the known depth. We then prove that if the depth of occlusion is constant in the ASI, the occlusion boundary in the ASI is similar to the occlusion boundary in the reference image, as shown in Fig. 1. The boundary similarity provides a new point of view to deal with occlusions in ASIs.

Based on the geometric relationship in extreme cases, we prove that some points are non-occluded in the ASI no matter what actual positions they are located at. The position relationship in the spatial domain is explored to guide the occlusion estimation in the angular domain. Specifically, we build an integral guided filter based on the reference sub-aperture image to estimate the probabilities of occlusions in ASIs. The filter is independent of ASIs so that it has no requirement for the angular resolution and can be implemented for all possible depth labels. We then integrate the integral guided filter in different cost volume calculation to exclude occlusions. Experimental results prove that the proposed filter is able to detect possible occlusions in ASIs and help the matching cost calculation. Results from different datasets show

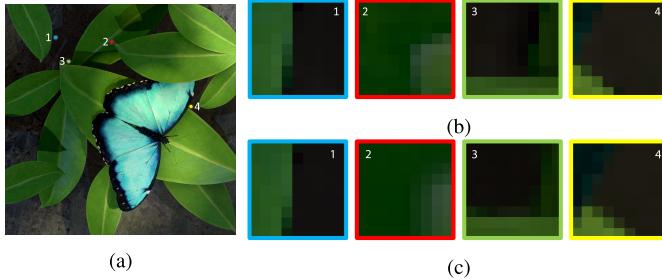


Fig. 1. The scene image and the angular sampling image (ASI) of occluded points. The occlusion boundary in the ASI is similar to the occlusion boundary in the reference scene image. (a) Scene image. (b) Scene image closeup. (c) ASI at the right depth label.

our method outperforms the existing state-of-the-art depth estimation methods, especially along occlusion boundaries.

The construction of the integral guided filter has been partly published in [20]. In this paper, we focus on the proof of the corresponding geometric relationship and the theorems of the proposed filter. Experimental results are complemented in terms of noises, occlusion detection and applicability. The main contributions of the paper are:

- 1) We analyze the geometric relationship between angular sampling images of occluded points and reference images based on depth information.
- 2) We mathematically prove that in the ASI, if the depth of occlusions is constant, the occlusion boundary is similar to the occlusion boundary in the central reference image.
- 3) We show the geometric relationship in extreme cases and develop an integral guided filter to estimate the occlusions probabilities in ASIs.
- 4) We integrate the integral guided filter into different depth estimation frameworks to estimate accurate depth maps.

## II. RELATED WORK

A lot of efforts have been made to obtain depth information from light field images. These methods are divided into three categories, multi-view based, epipolar plane image (EPI) based and focal stack based methods.

### A. Depth Estimation for Light Field Images

1) *Multi-View Based Methods*: Taking the pixels underneath the same position of each micro-lens, sub-aperture images, which capture scenes from different angles, are obtained from light field images. Therefore, traditional stereo matching based methods [23], [24], especially the small baseline 3D reconstruction [25]–[27], can be used for light field depth estimation. Yu *et al.* [18] developed a novel line-assisted graph-cut algorithm that encoded 3D line constraints into light field stereo matching. Chen *et al.* [17] introduced a bilateral consistency metric for light field stereo matching to handle significant occlusions. Heber and Pock [28] introduced a novel matching term based on the robust Principal Component Analysis (RPCA), which measured the difficulties of each view to be warped to a common warping center. Jeon *et al.* [12] designed a sub-pixel accurate cost that combined the sum of absolute differences (SAD) and the sum

of gradient differences (GRAD) for lenslet light field camera images. These methods calculated the consistency of each point in sub-aperture images and solved the occlusion problem by selecting appropriate sub-aperture images.

2) *EPI Based Methods*: Taking advantage of the continuous space in the angular direction, some methods based on EPIs have been developed. Wanner *et al.* [6] and Wanner and Goldluecke [11] developed a structure tensor based approach to measure each pixel's direction in EPIs. Their method did not try to match pixels at different locations, hence the processing speed is fast. Tasic and Berkner [29] proposed a local method for ray detection and dense depth estimation using the normalized second derivative of the Ray Gaussian kernel. The method detected occlusions by analyzing overlapping rays and their ordering. Zhang *et al.* [16] proposed a spinning parallelogram operator (SPO) to estimate the orientations of the lines in EPIs. The distribution distance they used is robust and is able to find correct depth labels even in occluded and noisy regions. Based on the EPI patches, some learning based methods are also presented to deal with the depth estimation problems in EPIs. Johannsen *et al.* [30] analyzed the coefficients of the EPI patches based on their depth-based dictionary and estimated depth for multiple depth layers. Heber and Pock [31] trained a Convolutional Neural Network (CNN) for orientation estimation in EPIs and implicitly learned the viewpoint selection in order to reduce occlusion artifacts.

3) *Focal Stack Based Methods*: Since scene images can be calculated to refocus at different positions [7] based on light field images, focal stack based methods are also developed to estimate depth information. Tao *et al.* [13] developed the defocus depth cues by computing the spatial variance after angular integration. Lin *et al.* [19] proposed to use symmetry analyses on the focal stack to find the in-focus slice for non-occluding points even with noise and undersampling. Strecke *et al.* [32] introduced partial focal stacks to deal with occlusions based on the focal stack symmetry. As the defocus cue is more robust to noisy scene compared to the correspondence cue, some methods combined the two cues for robust depth estimation [13], [19], [21].

### B. Occlusion Problems in Light Field Depth Estimation

For occlusions in multi-view, Kang *et al.* [33] introduced a cost function based on view selection to separate the matching term into two parts, and only the part with small variance is used. Vaish *et al.* [34] proposed two multi-view methods based on median and entropy to increase robustness to occlusions for images captured by large camera arrays.

As light field image has a dense sampling in angular domain, occlusion is becoming an important cue for most light field applications, especially for depth estimation. Angular sampling images (ASIs) [20], similar to the Surface Camera (SCam) [35] or the Surface Light Field [36], are shown in Fig 2. Recently, some research attempted to analyze the occlusion influences in light field images based on ASIs.

Chen *et al.* [17] developed a bilateral consistency metric (BCM) to estimate the occlusion points in ASI based on the statistic of the ASI in every possible depth label.

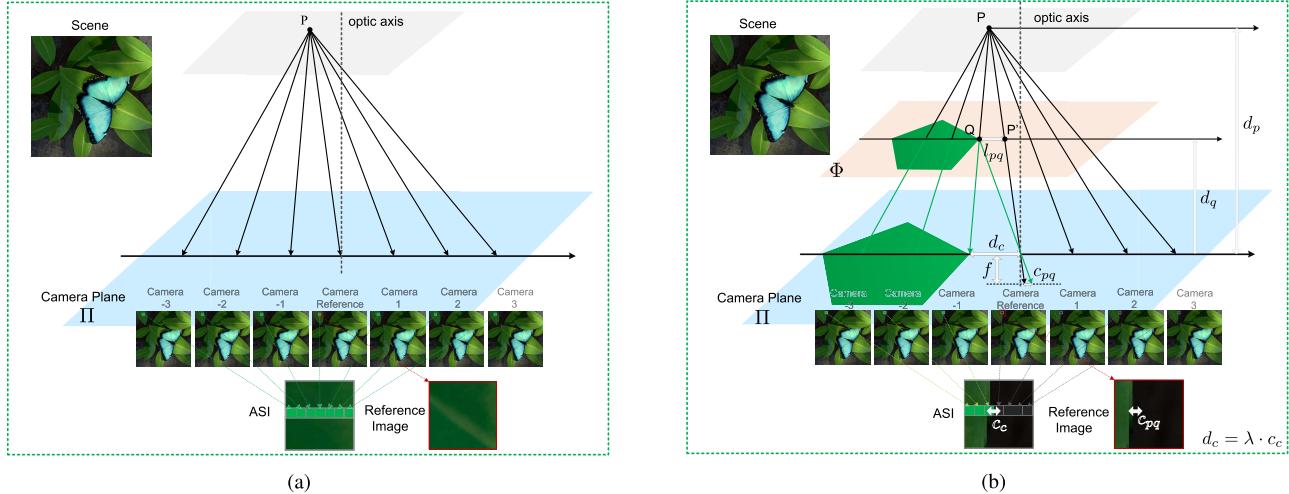


Fig. 2. The angular sampling images and the geometric relationship with the reference image. (a) Point  $P$  shows consistent color in different cameras and its ASI is constant when it is non-occluded. (b) Point  $Q$  occluded Point  $P$  in Camera  $-1, -2$  and  $-3$ . The ASI shows the similar occlusion boundary as in the reference scene image.

Williem and Park [22] developed an angular entropy metric and an adaptive defocus response for the correspondence and defocus cue respectively to deal with the occlusion and noise problems. However, these occlusion-aware methods estimate the probabilities of occlusions based on the calculated stereo or defocus responses in every depth label, which need dense sampling in the angular domain and also increase the estimation uncertainty, especially in textureless regions.

Wang *et al.* [21] used the reference view images to help exclude the occlusions in ASIs. The main idea is that the line separating two regions in the angular domain has the same orientation as the occlusion edge in the spatial domain. Therefore, they divided the ASI into two regions and used the means and variances in these two regions to remove the possible occluded region. However, they cannot handle the complex occlusion situations because the occlusion cannot be divided linearly. In our research, the occlusion relationship is also analyzed in the spatial domain and used in the angular domain, but it is more general for different occlusion situations.

In our paper, we demonstrate that the geometric relationship in the reference view image can be used to calculate the occlusion probabilities in angular patches. Unlike the other occlusion-aware methods, the occlusion probability is calculated only once for each point and applied to all possible depth labels. Therefore, the proposed integral guided filter is independent of depth labels and angular patches, and is easily integrated into other depth estimation frameworks.

### III. GEOMETRIC OCCLUSION RELATIONSHIP

In this paper, the 2D plane [6], [21] is used to parametrize the 4D light field. Light intersects the main lens plane at point  $(u, v)$  and intersects the imaging plane at point  $(x, y)$ . The light field is then expressed as  $L(x, y, u, v)$ , where  $(u, v)$  is the coordinate of the views, and  $(x, y)$  is the coordinate of the pixel in different sub-aperture images.

In this section, we first show the ASIs of general and occluded points in Lambertian scenes. Then we assume the depth is known and analyze the ASIs of the occluded points.

The geometrical occlusion relationship between the reference image and the ASI is also proved, which is used to estimate possible occlusions in the next section.

#### A. Angular Sampling Image

For every point in the scene, the corresponding imaging point can be found in different views. If we gather the corresponding imaging point of each view in order, we obtain angular sampling images (ASIs) [20]. Theoretically, if a point is observed in every view in the light field, it is identified as an unoccluded point and its ASI shows the constant color. As in Fig. 2, if no objects occlude the light from the point to the cameras, the imaging point in each camera has the same color from point  $P$ . Therefore, the points in the ASI at the correct depth label have the lowest variance compared with other possible ASIs. According to the basic theory of multi-view stereo matching, the depth with the ASI that has the lowest variance is regarded as the true depth.

If a point is visible in some views but occluded by other objects in other views, the ASI includes textures due to the front occlusions. As shown in the right side of Fig. 2, some views capture occlusions instead of  $P$ . In this case, the ASI is not consistent and includes intensities from occlusions. Therefore, we cannot acquire the correct depth label by simply calculating the variance of all points in ASIs.

In order to identify occlusions from ASIs, previous methods rely on statistical analyses in every possible ASI, such as consistency [17], variance [21], median and entropy [34]. Here, we propose that the distribution of occlusion points in the angular domain is related to its shape in the spatial domain.

#### B. Geometric Occlusion Relationship

In this section, we assume that the depth of the occlusions and the occluded points are known and analyze the detailed geometric relationship between the ASI and the reference image. Although the accurate depth values are not available before depth estimation, we will show in the next section that

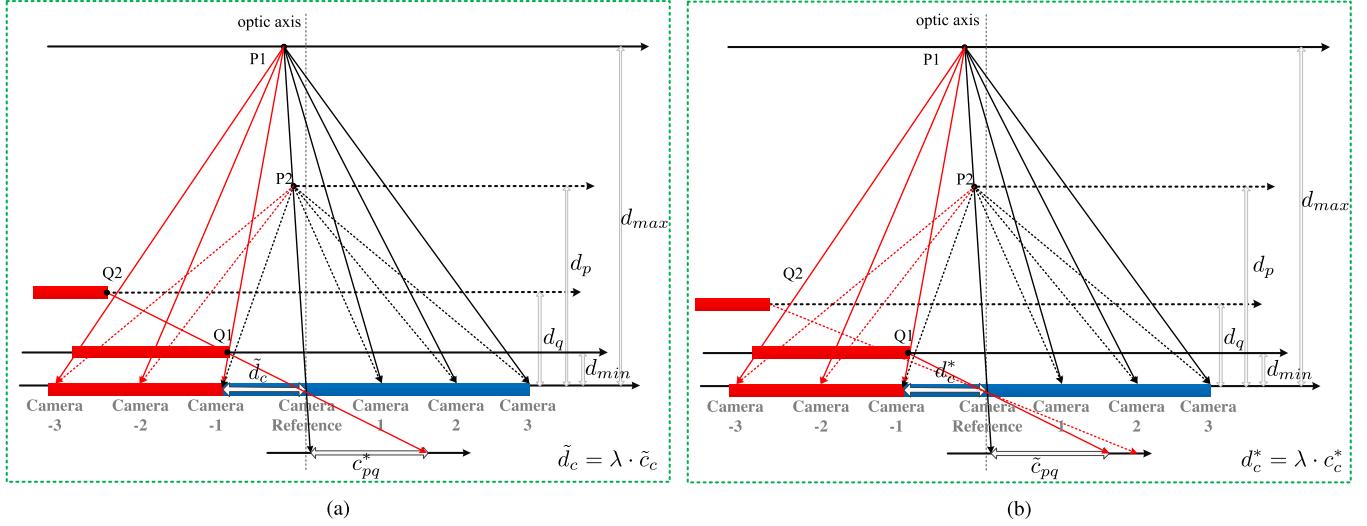


Fig. 3. The explanation for points under different occlusion assumptions. (a) When  $c_{pq}$  is fixed, if the distance from the occlusion to the central point in ASI  $c_c$  is less than  $\tilde{c}_c$ , the possible occlusion point can be ignored. (b) When  $c_c$  is fixed, if the distance from  $Q$  to  $P$  in the reference sub-aperture image  $c_{pq}$  is longer than  $\tilde{c}_{pq}$ , the point  $Q$  cannot occlude point  $P$  in the camera.

the geometric relationship in extreme cases can help to exclude occlusions in ASIs.

*Theorem 1:* Let  $d_p$  be the distance from the reference point  $P$  to the camera plane  $\Pi$ ;  $d_q$  be the distance from the occlusion point  $Q$  to the camera plane  $\Pi$  and  $d_q < d_p$ ;  $c_{pq}$  be the pixel distance from point  $P$  to the occlusion point  $Q$  in the reference sub-aperture image;  $c_c$  be the pixel distance from the occlusion  $Q$  to central reference point in angular sampling image;  $k$  is a constant that related to the configuration of the light field camera, which is independent of  $P$  and  $Q$ . Then we have:

$$\frac{c_{pq}}{c_c} = k \cdot \left( \frac{1}{d_q} - \frac{1}{d_p} \right) \quad (1)$$

For a proof, we introduce more variables as shown in Fig. 2:  $l_{pq}$  - the distance from  $P'$  to  $Q$  within plane  $\Phi$ , where  $\Phi$  is parallel to the camera plane  $\Pi$ .

$d_c$  - the distance from the optical axis of the reference view to the optical axis of the view where  $Q$  occludes  $P$  in the plane  $\Pi$ .

$f$  - the focal length of the reference camera.

$\lambda$  - the proportion between pixel distance  $c_c$  in ASIs and the actual camera distance  $d_c$ , where  $d_c = \lambda \cdot c_c$ .

According to the similar triangles, we have:

$$\frac{l_{pq}}{d_c} = \frac{d_p - d_q}{d_p} \quad (2)$$

In a similar way, according to the imaging model of the reference camera, the relationship between  $c_{pq}$  and  $l_{pq}$  is calculated as:

$$\frac{l_{pq}}{c_{pq}} = \frac{d_q}{f} \quad (3)$$

If we combine Eq. 2 and Eq. 3 and set  $k = f \cdot \lambda$ , the relationship between  $c_c$  and  $c_{pq}$  in Theorem 1 is proved. If  $d_q$  decreases,  $c_c$  decreases when other variables are fixed. This means occlusion  $Q$  near the camera is more likely to

occlude other objects. This description makes sense because we see nothing if one object is located just in front of our eyes. Similarly, if  $d_p$  increases,  $c_c$  decreases when other variables are fixed, which means further points are more likely to be occluded by the front points.

### C. Occlusion Boundary Similarity

As the resolution of ASIs is low, here we assume that the occlusion points in one ASI have the same depth  $d_q$ . The other points in the ASI come from the same point  $P$  and also have the same depth  $d_p$ . According to Theorem 1, it is easy to prove that:

*Theorem 2:* If occlusions in one angular sampling image have the same depth value, the occlusion boundary in the angular sampling image and the occlusion boundary in the reference image are similar. The ratio of the two boundaries is  $k \cdot (\frac{1}{d_q} - \frac{1}{d_p})$ .

Some examples are shown in Fig. 1. The boundary similarity provides another point of view to analyze occlusion points in ASIs. Theorem 1 shows that if we know the depth of the occlusion ( $d_q$ ) and the measuring point ( $d_p$ ), the occlusion points in ASIs are calculable. Although the depth of the scene is unknown and it is impossible to calculate correct occlusion points before depth estimation, we can roughly estimate occlusion points based on different depth assumptions according to the boundary similarity.

## IV. INTEGRAL GUIDED FILTER

As we mentioned in the last section, occlusions can be estimated in ASIs if the depth of the scene is given. However, due to the fact that the depth of the whole scene is unknown, the problem is ill-posed. In this section, we propose to solve this problem by estimating the occlusion relationship in extreme cases and build an integral guided filter for depth estimation.

### A. Geometric Analysis in Extreme Cases

As we analyzed in the last section, the point tends to occlude the other points behind it if it locates near the camera. Likewise, the point is easier to be occluded by the other points if it is far from the camera. In this section, we assume that the farthest and nearest depth in the scene is known and calculate the possible occlusions relationships based on Theorem 1. The diagram in Fig. 3 illustrates the possible occlusion cases in the 1-D situation.

**Theorem 3:** If  $c_{pq}$  is fixed and  $d_{max}, d_{min}$  are the farthest and nearest points in the scene, the minimum distance  $c_c$  of the point in the angular sampling image that  $Q$  may occlude  $P$  is  $\tilde{c}_c$ :

$$\tilde{c}_c = c_{pq} / (k \cdot (\frac{1}{d_{min}} - \frac{1}{d_{max}})). \quad (4)$$

When  $Q$  occludes  $P$  with  $c_{pq}$  fixed and  $d_p, d_q \in (d_{min}, d_{max})$ , we have  $c_c \geq \tilde{c}_c$  according to Theorem 1. If  $c_c < \tilde{c}_c$ ,  $Q$  cannot occlude  $P$  so that the possible occlusion can be ignored. On the contrary, if  $c_c \geq \tilde{c}_c$ , the point may be occluded. For example as in Fig. 3,  $P1$  is occluded by  $Q1$  in Camera  $-1$  when the points locate in the farthest and nearest place, respectively. When  $c_{pq}$  is fixed, neither the possible  $Q2$  nor  $Q1$  can occlude  $P1$  or  $P2$  when  $c_c < \tilde{c}_c$ . This means, for a possible occlusion point that is  $c_{pq}$  away from  $P$  in the reference view, only the point in the ASI with  $c_c \geq \tilde{c}_c$  may be occluded.

**Theorem 4:** If  $c_c$  is fixed and  $d_{max}, d_{min}$  are the farthest and nearest points in the scene, the maximum distance  $c_{pq}$  in the reference image that  $P$  may be occluded is  $\tilde{c}_{pq}$ :

$$\tilde{c}_{pq} = c_c \cdot k \cdot (\frac{1}{d_{min}} - \frac{1}{d_{max}}), \quad (5)$$

When  $Q$  occludes  $P$  with  $c_c$  fixed and  $d_p, d_q \in (d_{min}, d_{max})$ , we have  $c_{pq} \leq \tilde{c}_{pq}$  according to Theorem 1. If  $c_{pq} > \tilde{c}_{pq}$ ,  $Q$  cannot occlude point  $P$  so that the possible occlusion  $Q$  can be ignored. Conversely, if  $c_{pq} \leq \tilde{c}_{pq}$ , the point  $P$  may be invisible in camera because of the occlusion  $Q$ . The examples are shown in Fig. 3.  $P1$  is occluded by  $Q1$  in Camera  $-1$  when the points are located at the farthest and nearest place, respectively. When  $c_c$  is fixed,  $Q2$  cannot occlude  $P1$  or  $P2$  in Camera  $-1$  because  $c_{pq} > \tilde{c}_{pq}$ . This means, for a point that is  $c_c$  from the central point in the ASI, we only need to consider possible occlusions in the reference view whose distance satisfies  $c_{pq} \leq \tilde{c}_{pq}$ .

Based on these analyses in extreme cases with the estimated maximum and minimum depth in the scene, we find that some points cannot be occluded in ASIs no matter what actual positions they are located at. Therefore, we propose to estimate the probabilities of occlusions in ASIs based on the reference sub-aperture images.

### B. Integral Guided Filter

Following the construction of the guided filter [37], we propose an integral guided filter to predict occlusions in ASIs based on the guidance of the sub-aperture image.

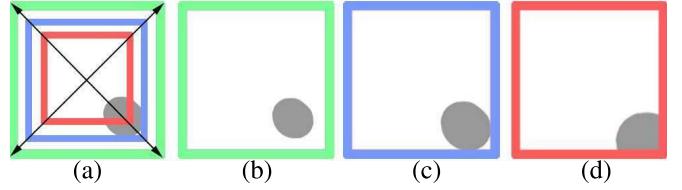


Fig. 4. The angular sampling images with different possible occlusion locations. (a) The occlusion in the reference sub-aperture image. (b), (c) and (d) shows the possible occlusion locations in ASIs with different depth labels.

We first define the filter  $W_r$ , which reflects the reliability of points in the ASI as:

$$W_r(x, y) = e^{(-W_a(x, y)/\sigma_w^2)}, \quad (6)$$

where  $\sigma_w$  is used to adjust the color sensitivity. The integral window  $W_a$  measures the probabilities that points are occluded in ASIs and is obtained by:

$$W_a(x, y) = \sum_0^x \sum_0^y W_o(\text{sign}(x)x, \text{sign}(y)y), \quad (7)$$

where  $x, y$  is the coordinates of the points in the window. The origin of the coordinate system is set to the center reference point for easier understanding.  $W_o$  measures the probability that the point is occluded based on the reference sub-aperture image and is defined as the color distances between the reference point and the surround points:

$$W_o(x, y) = |I(x, y) - I(x^*, y^*)|^2, \quad (8)$$

where  $I(x, y)$  is the point in a defined window centered at the reference point  $I(x^*, y^*)$  in the central sub-aperture image  $I$ . The window size is defined as  $2\tilde{c}_{pq}$ .

Here we analyze why the proposed filter is able to exclude occlusions in ASIs. Firstly, as shown in Theorem 4, only the points in the sub-aperture image that satisfy  $c_{pq} < \tilde{c}_{pq}$  may occlude the reference point in the ASIs. If the angular resolution of the light field is  $c_c$ , we only need to consider the points around the reference point with the distance  $c_{pq} < \tilde{c}_{pq}$  in the reference image. Therefore, the window size is defined as  $2\tilde{c}_{pq}$  in Eq. 8.

Secondly, we explore how to identify the possible occlusion points in the reference sub-aperture image  $W_o$ . According to a widely used assumption [17], [21], the points which have high contrast with the reference point may be occlusions. For the texture regions, which also have high contrast, the estimation becomes more ambiguous because the window excludes a part of points. This is a common problem for occlusion-aware depth estimation methods. However, the rest points usually have enough information for an accurate estimation. The estimation can also be further optimized using explicit smoothness assumptions.

Thirdly, we calculate the occlusion probability in ASIs ( $W_a$ ) based on the occlusion probability in the reference sub-aperture image ( $W_o$ ). As the actual depth of the occluded point and occlusion is not exactly the maximum and minimum depth, the occlusion points in ASI is uncertain. According to

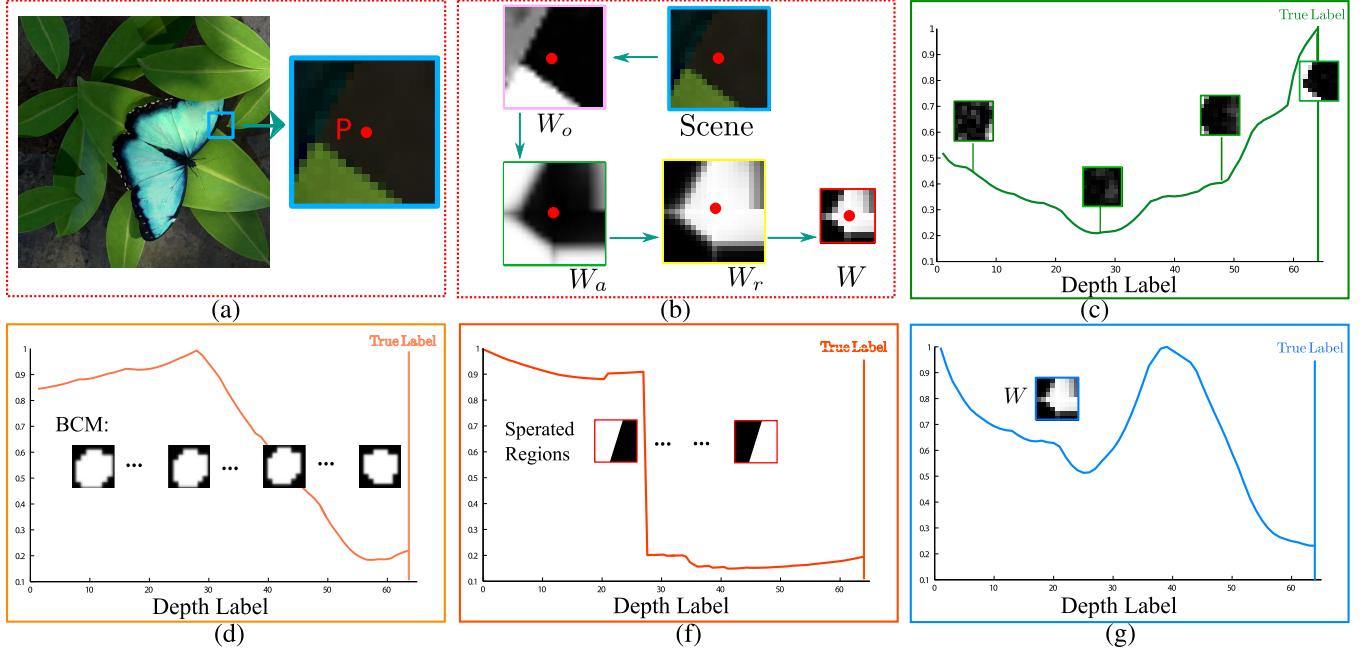


Fig. 5. The integral guided filter and the cost volume comparison. (a) The scene image and the reference point  $P$ , which is occluded in some views. (b) The proposed guided integral filter for  $P$ . The calculated filter  $W_o$ ,  $W_a$ ,  $W_r$  and  $W$  is shown accordingly. (c) The original cost volume  $\sum_j \rho(ASI_{j,d}, I_i)$  without occlusion exclusion. The corresponding ASIs in different depth labels are shown. Because of occlusions, it has large matching costs at the correct label. (d) The cost volume using BCM [17] and the corresponding weight is shown. BCM is calculated based on each possible ISI to exclude occlusions. (e) The cost volume using [21]. The possible ASIs are separated into two regions and only one region with small variance is used. The cost has a sudden change because the used region changes. Here the junction cannot be separated by a simple line and they fail to detect the correct depth label. (f) The cost volume  $\sum_j W_{i,j} \rho(ASI_{j,d}, I_i)$  calculated using the proposed filter  $W$ . The proposed method achieves the minimum cost at the correct depth label.

Theorem 2, if the occlusions are calculated in the reference image, the possible distributions of occlusions in ASIs are shown in Fig. 4. As we analyzed, if the occlusion point is  $c_{pq}$  from the reference point in the sub-aperture image, it may occlude points that satisfy  $c_c \geq \tilde{c}_c$ . This means all the points that satisfy  $c_c \geq \tilde{c}_c$  in the ISI should be added its corresponding occlusion probability. Therefore, we add its occlusion possibility to points along the direction to the reference point in the ISI. From another point of view, for a point that is  $c_c$  from the central point in the ISI, the possible occlusions in the reference view, which satisfy  $c_{pq} \leq \tilde{c}_{pq}$ , are considered as possible occlusions. Therefore, all the probabilities of occlusions which are closer to the reference points are added to its final occlusion probability.

As  $c_c$  is calculated as the distance between the possible occlusion and the reference point, the occlusion probability should be added along the line between the point and the central reference point. However, in digital images, it is hard to implement the addition along the ray from the central point. Therefore, we propose to use an integral summation as an approximation to add all the probabilities as in Eq. 7. Specifically, the  $W_a$  extends the possible occlusion regions along the diagonal direction to the borders. The window detects all possible occlusions with different possible depth relationships.

Finally, the probability that the point is occluded in the ISI ( $W_a$ ) is then changed to the reliability of the point in the ISI ( $W_r$ ) using the exponential function. The points that have high reliabilities means we are more confident that they are not occluded by other points.  $W_r$  is then resampled to fit the

size of ASIs, which is formulated as the proposed integral guided filter  $W$ . The filter  $W$  is then used to filter ASIs in cost calculations. The example filter in each step is shown in Fig. 5.

## V. DEPTH ESTIMATION

Because the integral guided filter  $W$  is independent of the cost volumes calculation, it can be easily applied to different depth estimation frameworks. In this section, we integrate the proposed integral guided filter in the local cost calculation to reduce occlusions. Then the optimization methods, such as the edge-preserving filter [37] and multi-label optimization [38], are applied and smooth depth maps with clear edges are acquired.

### A. Local Cost Calculation

Here, we first show how to integrate the proposed filter  $W$  in a traditional stereo matching process. Specifically, for pixel  $i$  in the reference view  $I$ , the cost volume in different depth is defined as:

$$C_{i,d} = \sum_j W_{i,j} \cdot \rho(ASI_{j,d}, I_i), \quad (9)$$

where  $ASI_{j,d}$  is intensity of pixel  $j$  in the corresponding ISI extracted at the possible depth label  $d$ .  $\rho$  is a distance function and is defined as various kinds of measurements, such as the traditional sum of absolute differences (SAD), sum of square differences (SSD) and sum of gradient differences (GRAD). In our simple depth estimation framework, the distance is defined as  $\rho(a, b) = 1 - \exp(-(a-b)^2/\sigma_c^2)$ , where  $\sigma_c$  controls

the sensitivity of the function. Notice that  $W$  is independent of  $d$ , which means it is unchanged for all possible depth labels.

For the synthetic light field dataset that has many textures and little noise, the simple corresponding cue is enough. However, for images captured by plenoptic cameras, one traditional corresponding cue is not robust. Therefore, the defocus cue [13], [21] and the gradient difference [12] is also proposed to combine with the original corresponding cue for a more robust local cost calculation. Like the corresponding cue, the proposed filter can be easily applied to different cost calculations. In the experiment part, we show that the filter improves the depth estimation results based on different cost calculations.

Similar to Wang *et al.* [21], our integral guided filter can also be used for occluded point detection. Another opposite filter ( $W' = 1 - W$ ) is applied to the local depth cost calculation and compared with the one used filter  $W$ . The occlusions are predicted based on the differences between the two cost volumes. The accuracy of the occluded points detection measures the effectiveness of the proposed filter to some extent. Comparative results are shown in our experiments.

### B. Comparison for Occlusion Exclusion

Recently, the ASI is also studied in other research work [17], [21], [22] to handle the occlusion problem. Here, we compare the example using Bilateral Consistency Metric (BCM) [17], occlusion-aware method [21] and the proposed filter. The corresponding cost volumes of the same occluded point and the calculated weights in different methods are shown in Fig. 5. The corresponding ASI in every possible depth label is shown, where the ASI in the correct depth label contains occlusions.

Chen *et al.* [17] built the BCM to estimate the probability of each pixel in ASIs belonging to occlusions based on the ASIs themselves. Specifically, they assigned a small weight if the point has a high color and spatial variance with the reference point. Finally, only half pixels in ASIs with the high weight is used. The metric, shown in Fig. 5, changes at every possible depth label because the BCM is calculated for each depth label. Compared with the original cost volume, the cost volume using BCM shows that occlusions are partly removed in this example. However, their method needs a dense sampling in the angular domain to reliably measure the BCM. Furthermore, as we show in the experiment part, they delete all the points with high variances in each depth label so that the minimum label is ambiguous to find, especially in textureless and noisy regions.

Wang *et al.* [21] divided the ASI into two regions using a single line according to the orientation of the point in the reference sub-aperture image. The region with smaller variance in the ASI is picked for the following cost calculations. Their method is partly similar to ours, but it cannot handle complex occlusions. They also need a color consistency constraint for a ‘reversed’ patch when they deal with an incorrect depth. As shown in Fig. 5, the picked region rapidly changes in one specific depth label. For this example, the occlusion part cannot be separated by a single line, so that they fail to exclude occlusions in the ASI and the depth label is wrongly estimated.

By contrast, the proposed integral guided filter constructs a window only according to the reference sub-aperture image and the filter is then used for all possible depth labels. The occlusion estimation is independent of the cost calculation so that it is simple to apply in various scenes and different cost calculations. As shown in Fig. 5, the proposed filter is able to correctly detect and exclude occlusions in ASIs and our method achieves the minimum cost at the correct depth label.

### C. Depth Optimization

In order to propagate the correct information from reliable points to textureless points, we first refine the cost volume using the guided filter [37]:

$$C'_{i,d} = \sum_j W_{i,j}^g(I) C_{i,d}, \quad (10)$$

where  $W^g$  is the guided filter as in [39] and  $j$  is the pixel in the defined filter window. Instead of using the graph based optimization, the filter-based method is more efficient and can be easily parallelized. Moreover, because the possible occlusions have been excluded, the simple propagation method is enough for most examples. Finally, we assign the depth label with global minimum value as the correct depth label using the winner-take-all method. In the experiments, we use this simple optimization method for synthetic light field images to prove the effectiveness of the proposed integral guided filter in depth estimation.

Recently, more complicated optimization methods for light field images are developed especially for Lytro images. For example, Jeon *et al.* [12] used weighted median filter [40], graph cuts [38], and finally the iterative refinement [41] to acquire the final depth maps. Wang *et al.* [21] introduced a global regularization using a smoothness term and reduced the strength of the constraint if an occlusion is likely to be between them. All the optimization methods are technically feasible to our depth estimation framework and can be chosen depending on different datasets, hardware restrictions and time requirement. The depth maps using more optimization methods for noisy Lytro images are also shown in our experiments.

## VI. EXPERIMENTAL RESULTS

In this section, we show experimental results based on different cost volumes and different optimization methods using different parameters, angular resolutions and datasets.

*Dataset:* The images we use in the experiments can be divided into two parts. The synthetic light field images are collected from Wanner *et al.* [42], Wang *et al.* [21] and Honauer *et al.* [43] with available ground truth. The real light field images captured by Lytro [12], [13] and Lytro Illum camera [21] are also evaluated.

*Compared Methods:* In this paper, the performance of our algorithm is compared with Chen *et al.* [17] (BCM), Wang *et al.* [21], Zhang *et al.* [16], Wanner and Goldluecke [11], Johannsen *et al.* [30] and MV [44], which have shown good performances on light field images recently. The methods designed for plenoptic light field camera images, e.g. Tao *et al.* [13], Jeon *et al.* [12], are also illustrated for comparisons.

TABLE I  
THE ERROR RATE COMPARED WITH GROUND TRUTH. (%)

Image	Wanner [11]		Chen [17]		Ours	
	All	Occ	All	Occ	All	Occ
Buddha(9)	2.41	15.01	1.72	8.31	<b>1.21</b>	<b>6.25</b>
Buddha(7)	2.58	15.68	2.14	8.34	<b>1.35</b>	<b>6.58</b>
Buddha(5)	3.72	22.72	2.54	9.11	<b>1.72</b>	<b>7.78</b>
Buddha(3)	-	-	11.37	19.43	<b>4.54</b>	<b>15.50</b>

**Evaluation Methods:** As a quality measurement, we use the percentage of depth value below a relative error (5% in our experiments) based on the ground truth, as [11]. The accuracy is calculated for the overall images (All) and occlusion regions (Occ) respectively. The  $MSE * 100$  and  $BadPix(0.07)$  metrics are also used to evaluate the depth maps in the benchmark [43]. The occlusion boundaries, which are calculated from the estimated depth map, are evaluated using precision-recall curves as in [21]. We also use the precision-recall curves to evaluate the occluded point detection results.

### A. Performance Analysis

**Optimal Parameter Selection:** In this subsection, we analyze the parameter  $\sigma_w$  and  $\sigma_c$  used in Eq. 6 and Eq. 9, respectively. The scale  $\sigma_w$  determines the sensitivity of the proposed filter. As the scales decreased, the filter is more sensitive to occlusion. If the image has a lot of noise,  $\sigma_w$  needs to be larger in order to better distinguish occlusions from noisy regions. The scale  $\sigma_c$  determines the sensitivity of the cost calculations, which is also used in the traditional cost settings [23]. Similarly, if  $\sigma_c$  is assigned a smaller value, it is better for occlusion exclusion but is more sensitive to noises. The optimal parameter is then found by testing a number of different parameters on different datasets. In our implementation,  $\sigma_w$  is set to 0.5, and  $\sigma_c$  is set to 0.1.

**Angular Resolution:** We implement the proposed method for synthetic light field images with different angular resolutions. The numerical results, compared with the state-of-art algorithms, are illustrated in Table I. It should be noted that the other methods are not effective if they use a small set of views. The metric Chen *et al.* [17] used to estimate the occlusions is calculated on the ASI for every possible depth label, which is not reliable if the angular resolution is too low. By contrast, our method relies on the reference view image to predict the occlusions, which is independent of the angular resolution.

The visual results, shown in Fig. 6, verify that we acquire the depth map with clear edges even with  $5 \times 5$  views. This quality is quite important for light field images captured by plenoptic cameras, which have fewer available views for depth estimation because of the hardware restrictions.

**Noisy Scenes:** We evaluate the robustness of our method to noisy scenes using the image ‘Dots’ [43] in the new benchmark. The scene is designed to show the effects of camera noises. Gaussian noise with growing variances is added in the row-major order in the image.

As the proposed filter is calculated only based on the reference image, we propose to denoise the reference sub-

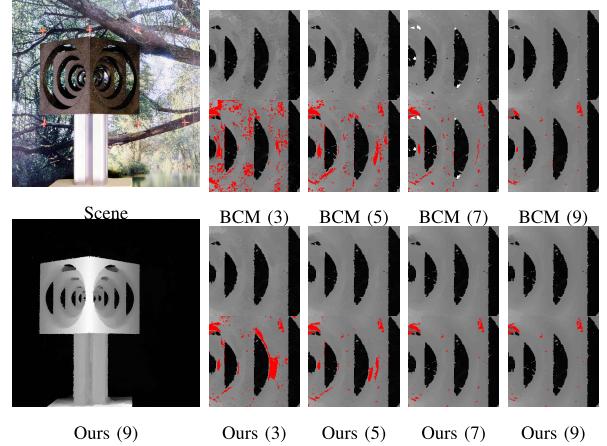


Fig. 6. Depth estimation from image “Cube” with different angular resolutions and the error pixels are labeled red. The proposed method achieves high accuracy even with a low angular resolution.

aperture image first using the guided filter [37] in order to calculate the correct filter to exclude occlusions. The original light field image with noises are used to calculate the stereo matching costs. Fig. 8 shows after the denoising operation for the reference view image, the proposed filter is more robust to noise and still able to exclude occlusions.

### B. Occluded Point Detection Comparisons

In order to show the effectiveness of the proposed integral guided filter in ASIs, we develop a new method to evaluate the occluded point detection results. According to the ground truth depth of the surrounding points, the point is labeled as an occluded point if it is calculated to be occluded in one or more views in light field images. The example ground truth of the occluded points is shown in Fig. 7. If the objects are far away from the occlusions, the number of occluded points increases around the depth discontinuous boundaries. These occluded points may have depth estimation errors as they are partly occluded in different views. If the occluded points are detected accurately during the cost calculation, the depth estimation will be more robust for occlusions.

As analyzed in Sec. V-A, we use the points that are defined as possible occlusions ( $W'$ ) in ASIs to calculate another matching cost volume. The occluded point detection method, same as [21], is used for different occlusion aware methods. The results are shown in Fig. 7, where our method achieves better performance than the other methods, which means in the local cost calculation, our method is more effective for excluding occlusions.

### C. Depth Estimation Comparisons

The numerical results for synthetic light field images with  $9 \times 9$  views [42] are illustrated in Table II. Our method outperforms the existing state-of-the-art depth estimation methods, especially in occluded regions.

**1) Occlusion Performance:** The detailed comparisons on depth discontinuity regions with state-of-the-art occlusion-aware methods are shown in Fig. 8, Fig. 9 and Fig. 10. Our method shows more accurate depth estimation compared

TABLE II  
THE ERROR RATE OF THE ESTIMATED DEPTH COMPARED WITH GROUND TRUTH. (%)

Image	Wanner <i>et al.</i> [11]		Chen <i>et al.</i> [17]		Wang <i>et al.</i> [21]		Ours	
	overall	occlusion	overall	occlusion	overall	occlusion	overall	occlusion
Cube(9)	0.92	13.29	1.11	9.72	1.44	11.46	<b>0.83</b>	<b>8.36</b>
Maria(9)	<b>2.07</b>	12.07	2.67	<b>10.56</b>	4.49	14.8	2.24	11.87
StillLife(9)	4.30	8.35	<b>1.51</b>	6.53	3.99	12.79	1.62	<b>6.23</b>
Papillon(9)	21.24	34.44	12.86	12.83	10.9	17.9	<b>7.31</b>	<b>12.55</b>

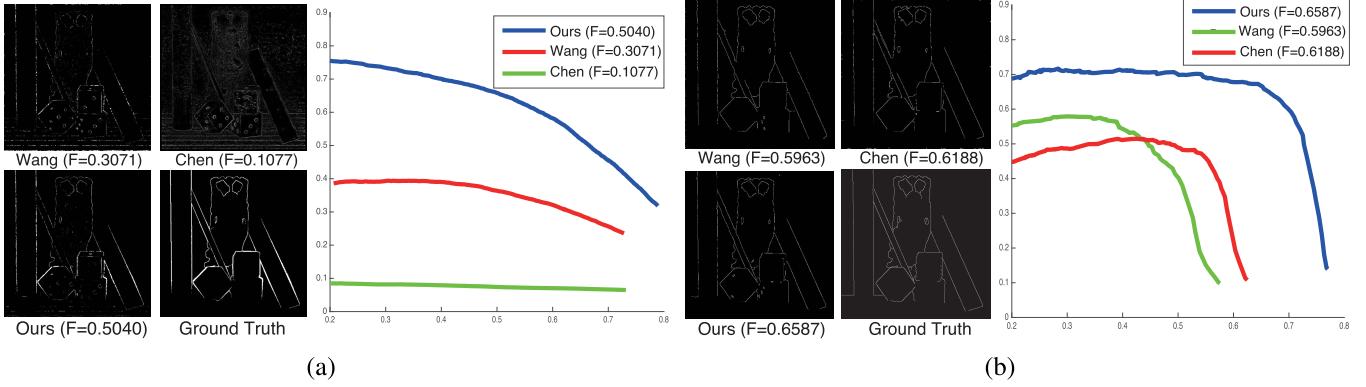


Fig. 7. (a) PR curves of occluded points detection results. The points are detected by comparing two cost volumes (with occlusion and without occlusion) and are calculated using different occlusion exclusion methods. (b) PR curves of occlusion boundary evaluation. The occlusion boundary is calculated using the estimated depth map. Our method achieves better performance than current state-of-the-art occlusion-aware methods in both results.

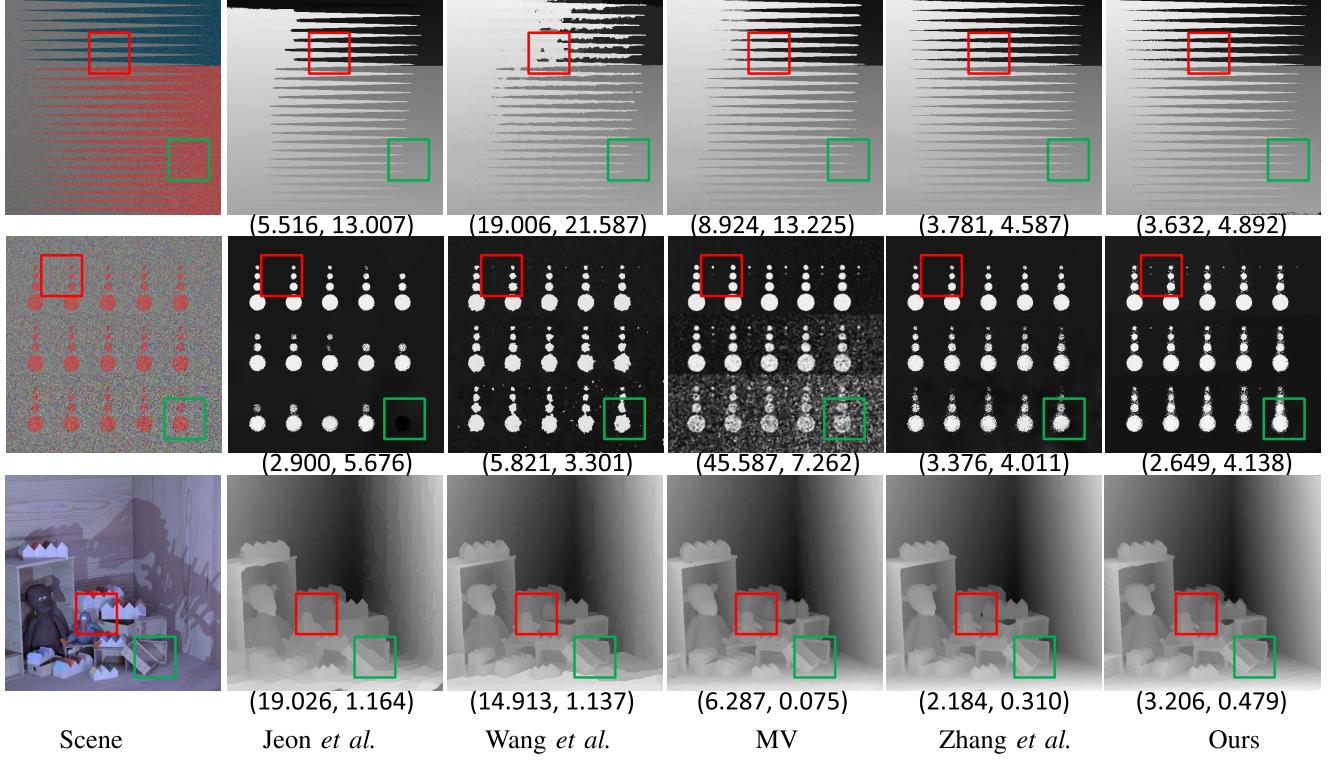


Fig. 8. The depth estimation results using synthetic light field images by Honauer *et al.* [43]. The *BadPix* and *MSE* metrics are also illustrated in the brackets, respectively. Our method captures the details along the depth edges when the occlusions relationship is complex (labeled regions). The second image is filled with noises. After denoising the guided reference image, our method achieves better results than the others.

with other methods, especially in occlusion regions. Examples of outdoor scenes captured by Lytro Illum cameras are shown in Fig. 11, in which we also achieve the best

performance. As in [21], the occlusion boundaries are also calculated using the estimated depth map and compared with the ground truth occlusion boundaries. The precision-recall

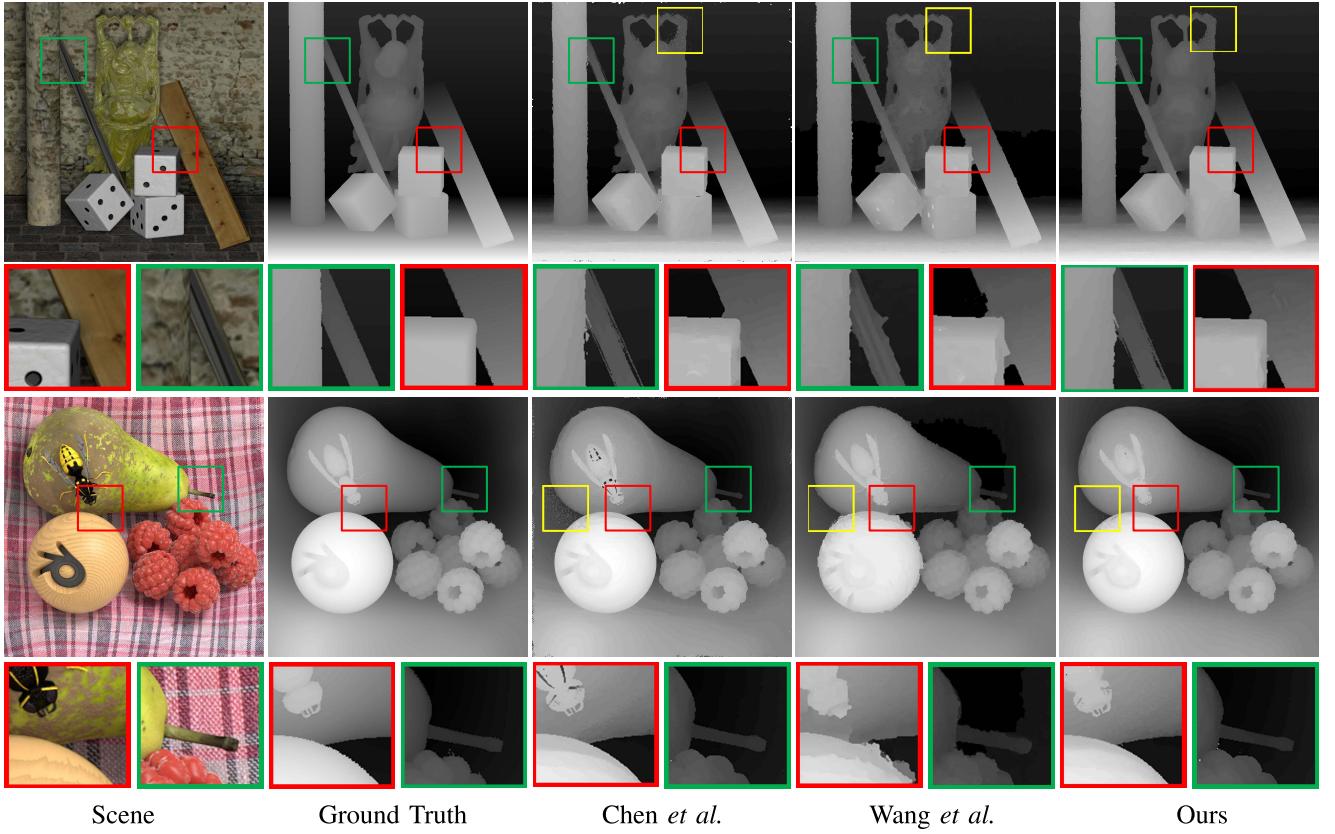


Fig. 9. Depth estimation for synthetic light field images by Wanner *et al.* [42]. Occlusion regions are amplified (red and green boxes) for better contrast, and we correctly capture the shape of the objects. Wang *et al.* [21] fails to capture the details along the depth edges when the occlusions relationship is complex. The initial depth maps of the textured background without any filters or optimization are partly shown in the yellow boxes. The initial estimation results of Chen *et al.* [17] show a lot of noise whereas our method estimates a lot of accurate depth labels in these regions.

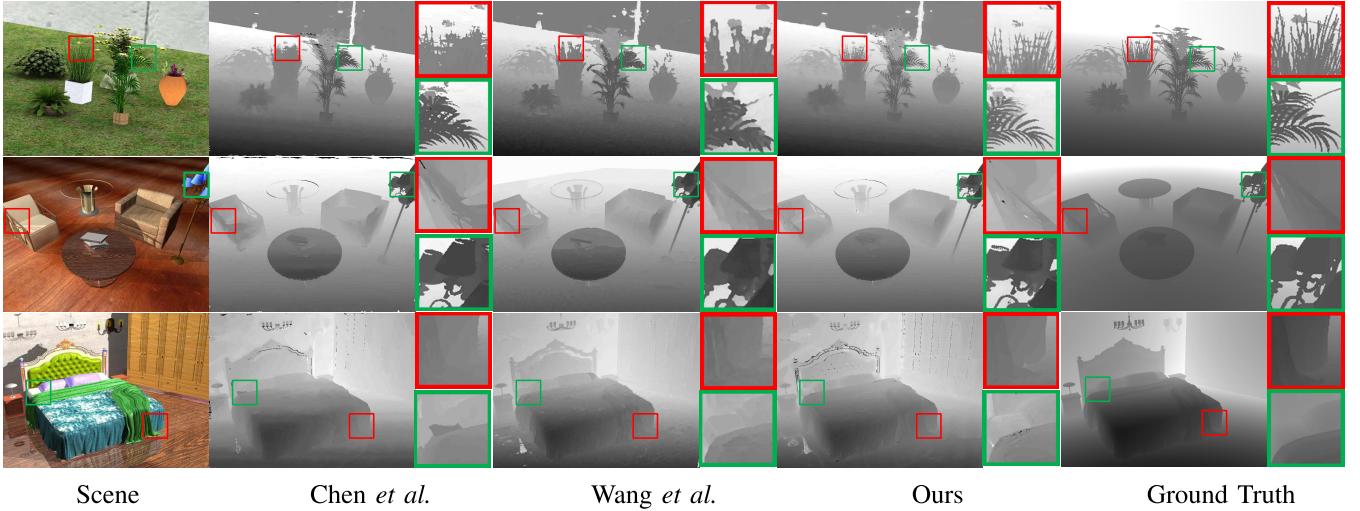


Fig. 10. Depth estimation result for synthetic light field images. Wang *et al.* [21] has many errors in complex occlusion environments, such as the outdoor scenes. Our method achieves the clearest shape of the leaves in the first example, which proves the ability to handle heavy occlusions. We also acquire the most smooth table and bed sheet in the second and third example, which proves that our method is able to keep useful information compared with other methods.

curves are shown in Fig. 7 and the proposed method achieves the best performance.

Chen *et al.* [17] calculated the occlusion possibilities of the ASIs based on the statistics of the ASIs themselves. Pixels that have a large contrast with the reference

point are all deleted in their cost calculation so that their method is able to acquire clear depth edges. However, their method cannot handle flat regions with little texture and is ambiguous to find the correct depth in complex occlusion relationship.

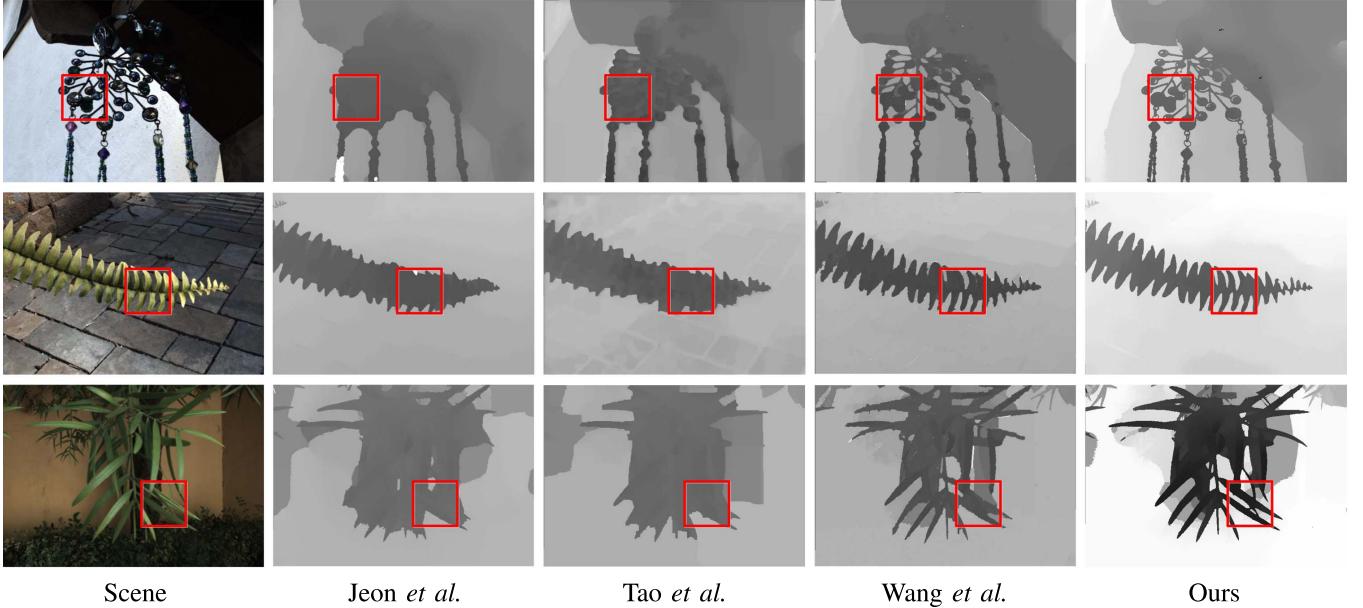


Fig. 11. Depth estimation of Lytro Illum images provided by Wang *et al.* [21]. Our method successfully captures the shapes of each object in the images, including the thin structures, while other methods fail or generate thickened results.

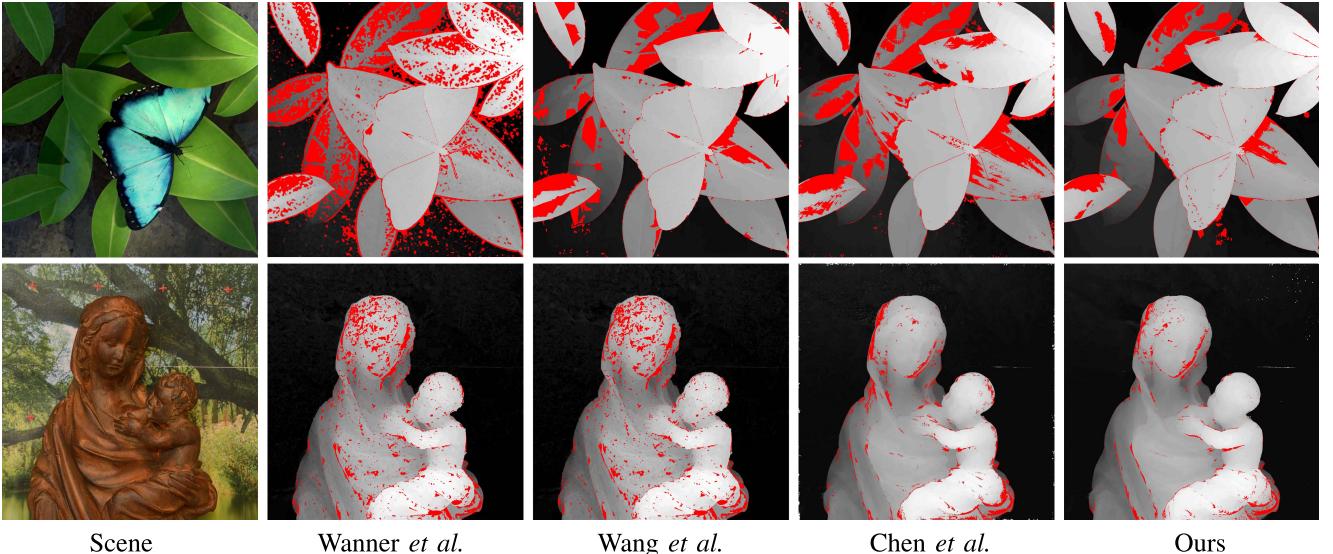


Fig. 12. Depth estimation and errors maps for synthetic light field images provided by Wanner *et al.* [42]. The pixels with large errors compared with ground truth are labeled red for distinct comparison. These images have lots of ambiguous regions along with many occlusions. Chen *et al.* [17] has many errors in ambiguous regions. Our algorithm achieves comparable results in ambiguous regions and acquires higher overall accuracy.

Wang *et al.* [21] achieved good performance when the depth edge is straight. The reason is they separated the ASIs into two regions and the half region that may contain occlusions is not used. However, as clearly shown in the images, their method fails to handle occlusions when the occlusions relationship is complex. Zhang *et al.* [16] are ranked first in both the average *BadPix*(0.07) and average *MSE* part currently in the new benchmark [43]. The spinning parallelogram operator (SPO) is implemented in EPIs to estimate the orientations of the lines. The SPO keeps effective information when the points are occluded and correct depth maps are then acquired after the depth propagation. However, as they do not explore

the occlusion relationship, the occlusion still exists in their depth estimation. For complex occlusion situations, the occlusion information in SPO cannot be corrected by the simple propagation.

We calculate the occlusion probability according to the reference view. Therefore, no matter whether the occlusion boundary is straight or not, the shape of the filter is determined by its shape in the reference image. For image ‘Backgammon’ in the first row of Fig. 8, which is designed to assess the interplay of fine structures, occlusion boundaries and disparity differences, our results achieve the state-of-the-art performances. For complex occlusion situations, such as the

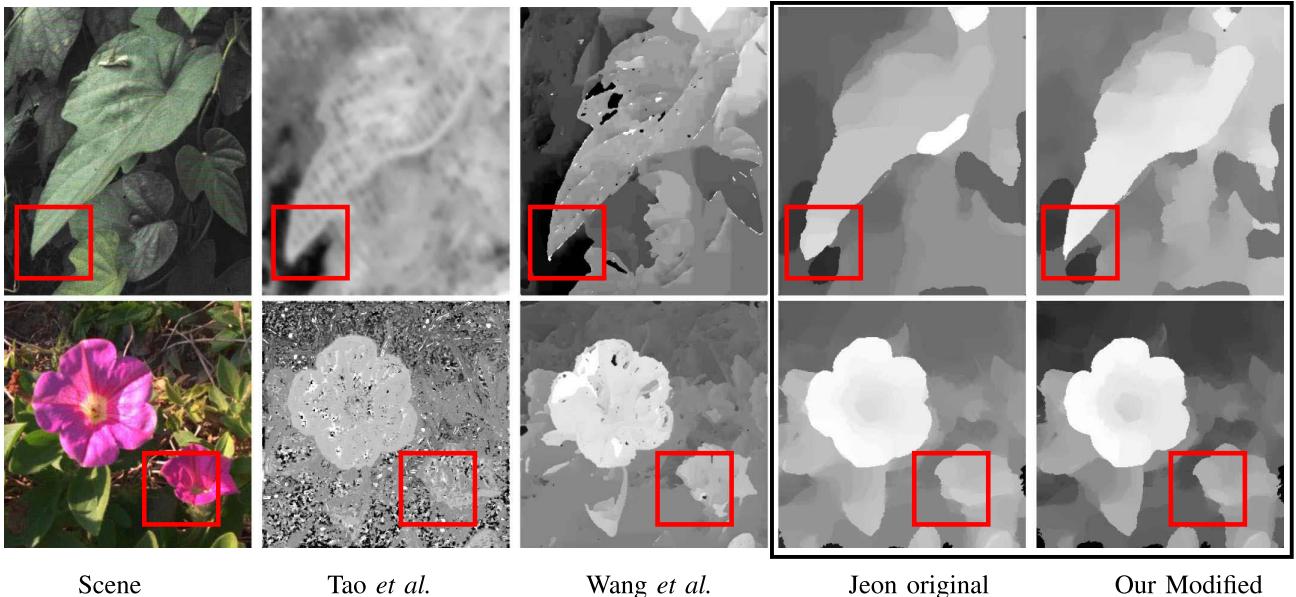


Fig. 13. Depth estimation using images from Lytro camera provided by Tao *et al.* [13] and Jeon *et al.* [12]. Images are noisy due to the hardware restriction. The proposed filter is added in the depth estimation framework by Jeon *et al.* [12]. Compared with their original results, depth maps with our filter show clear structures along the depth edges. The proposed filter is proved to be effective to exclude occlusions in their cost calculation.

insect in the second image in Fig. 9 and the leaves in the first image in Fig. 10, our method is still able to keep the correct depth boundaries.

2) *Overall Performance*: Fig. 12 compares results for synthetic light field image “Papillon” and “Maria” with ground truth. These images contain a lot of ambiguous regions along with many occlusions. The proposed method achieves comparable results in textureless regions compared with the state-of-the-art occlusion-aware methods. We also show the initial depth estimation results of the background parts in the yellow boxes in Fig. 9. Without any filters or optimization methods, we still estimate quite a lot accurate depth labels and the depth maps contain fewer noises than the others in these regions.

Compared with Wanner and Goldluecke [11], our algorithm recovers better results on occlusion regions and textureless regions. Chen *et al.* [17] has many errors in textureless and complex texture regions because the half of the pixels in ASIs with higher contrast than the others are deleted. Wang *et al.* [21] made an assumption that occlusions only come from the occlusion candidates so that for non-occlusion candidate points use the whole ASIs to calculate the cost volumes.

In contrast, for textureless regions, we pick more than half pixels in ASIs for cost calculations. For complex texture regions, as we pick the same region in every possible ATI to calculate the cost volume, we still keep some effective information in the cost volume to calculate the correct depth. Therefore, the proposed method is able to acquire a higher overall accuracy compared with the state-of-the-art methods.

#### D. Applicability Analysis

As the proposed filter is independent of the cost calculations, we add the integral guided filter into different frameworks to show its capability for occlusions exclusion.

We first add the filter in the cost volume proposed by Jeon *et al.* [12], where the sum of absolute differences (SAD) and the sum of gradient differences (GRAD) are used. The  $\sigma_w$  is set as 0.5 and other parameters are kept unchanged as they used in their framework. For a clear comparison, the optimization method is also kept the same. Comparison results are depicted in Fig. 13, which shows that the proposed filter is able to improve their depth estimation performances in occlusion regions.

Another example of Wang *et al.* [21] is given in Fig. 14, which is designed based on the combination of corresponding cue and defocus cue. Their occlusion exclusion method is replaced using the proposed filter and other procedures are kept. The initial occlusion prediction results from the local depth estimation (without optimization) are illustrated and our results show more accurate depth edges compared with their method. Our final depth maps also show clear edges than their original algorithms. The experiments show our method is applicable to various depth estimation frameworks to improve the accuracy of the occlusion boundaries.

#### E. Runtime Analysis

We implement the algorithm using Matlab with a computer equipped with Intel i7 3.60GHz CPU and 8GB RAM. The proposed integral guided filter is calculated independent of the cost volume computation, which is approximately 33 seconds for the synthetic light field images with  $768 \times 768$  spatial resolution and  $9 \times 9$  angular resolution. The runtime of the depth estimation is mainly determined by the matching cost calculation method. The filter is only used in the cost volume calculation where multiple operations with each pixel in ASIs are added. In our experiment, 27 seconds are added using the proposed filter compared with the original cost volume computation for the synthetic images (64 depth labels).

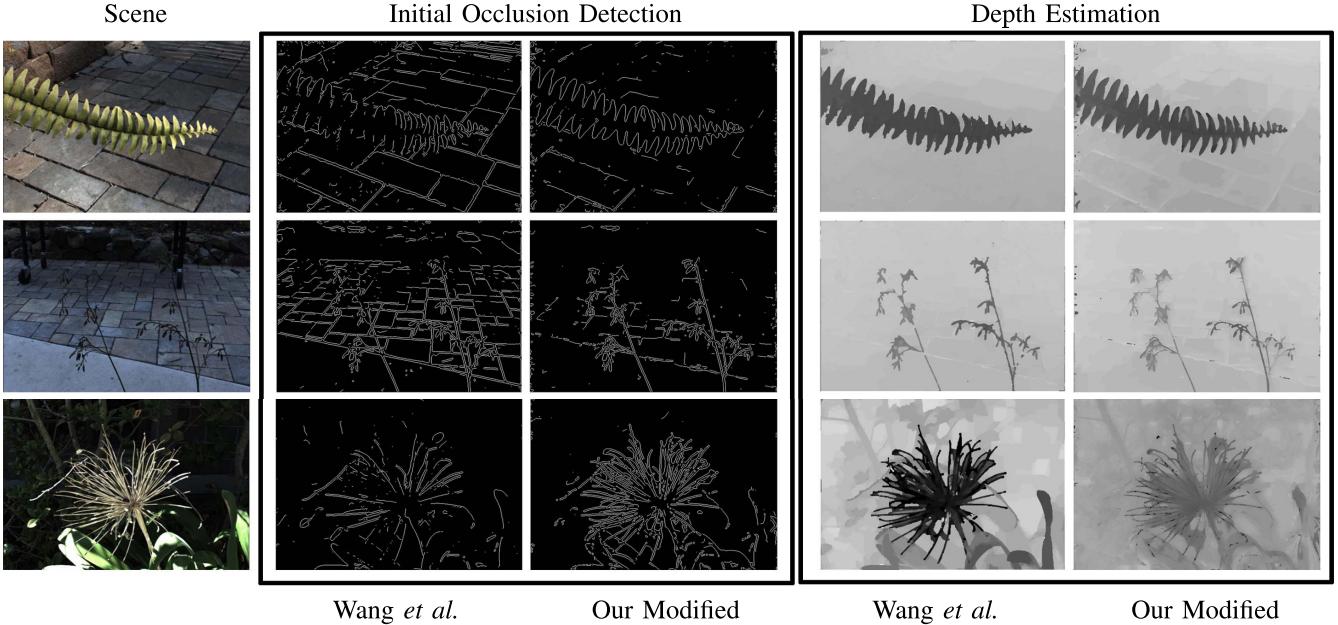


Fig. 14. The occlusion detection results from the local depth estimation (without optimization) and the final depth estimation. The proposed filter is added in the depth estimation framework by Jeon *et al.* [12]. The initial occlusion prediction results show that our method is able to detect occlusion points in ASIs for cost calculation. Compared with their original results, the final depth map using our filter shows accurate structures along the depth edges.

## VII. CONCLUSION

Taking into account the special structure of the light field data, we analyzed the geometric relationship between the ASIs and the reference sub-aperture images. We also proved the occlusion boundary similarity and showed that in extreme cases, some points cannot be occluded in ASIs. A novel integral guided filter is then proposed to handle occlusions in the angular domain based on the reference image in the spatial domain. This filter is used to exclude the occluded points in different stereo matching frameworks.

Compared with the state-of-the-art light field depth estimation methods, the proposed method achieves the sharpest edges in occlusion regions and good performances in ambiguous regions in different datasets. The algorithm works for light field images with low angular resolution and is robust for noisy images. Moreover, the proposed filter is implemented to different depth estimation frameworks and achieves more clear depth edges than the original algorithms.

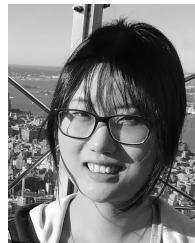
## REFERENCES

- [1] R. Ng, *Lytro Redefines Photography With Light Field Cameras*. Accessed on May 10, 2017. [Online]. Available: <http://www.lytro.com>
- [2] C. Perw and L. Wietzke, *The Next Generation of Photography*. Accessed on May 10, 2017. [Online]. Available: <http://www.raytrix.de>
- [3] T. E. Bishop, S. Zanetti, and P. Favaro, “Light field superresolution,” in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2009, pp. 1–9.
- [4] T. E. Bishop and P. Favaro, “The light field camera: Extended depth of field, aliasing, and superresolution,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 972–986, May 2012.
- [5] S. Wanner and B. Goldluecke, “Spatial and angular variational super-resolution of 4D light fields,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 608–621.
- [6] S. Wanner and B. Goldluecke, “Variational light field analysis for disparity estimation and super-resolution,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar. 2014.
- [7] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” *Comput. Sci. Tech. Rep.*, vol. 2, no. 11, pp. 1–11, 2005.
- [8] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, “Saliency detection on light field,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2806–2813.
- [9] H. Sheng, S. Zhang, X. Liu, and Z. Xiong, “Relative location for light field saliency detection,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2016, pp. 1631–1635.
- [10] E. H. Adelson and J. Y. A. Wang, “Single lens stereo with a plenoptic camera,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992.
- [11] S. Wanner and B. Goldluecke, “Globally consistent depth labeling of 4D light fields,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 41–48.
- [12] H.-G. Jeon *et al.*, “Accurate depth map estimation from a lenslet light field camera,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1547–1555.
- [13] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, “Depth from combining defocus and correspondence using light-field cameras,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 673–680.
- [14] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, “A 4D light-field dataset and CNN architectures for material recognition,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 121–138.
- [15] S. Wanner, C. Straehle, and B. Goldluecke, “Globally consistent multi-label assignment on the ray space of 4D light fields,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1011–1018.
- [16] S. Zhang, H. Sheng, C. Li, J. Zhang, and Z. Xiong, “Robust depth estimation for light field via spinning parallelogram operator,” *Comput. Vis. Image Understand.*, vol. 145, pp. 148–159, Apr. 2016.
- [17] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu, “Light field stereo matching using bilateral statistics of surface cameras,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1518–1525.
- [18] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu, “Line assisted light field triangulation and stereo matching,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2792–2799.
- [19] H. Lin, C. Chen, S. Bing Kang, and J. Yu, “Depth recovery from light field using focal stack symmetry,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3451–3459.
- [20] S. Hao, Z. Shuo, Z. Gengliang, and X. Zhang, “Guided integral filter for light field stereo matching,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 852–856.
- [21] T.-C. Wang, A. Efros, and R. Ramamoorthi, “Depth estimation with occlusion modeling using light-field cameras,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2170–2181, Nov. 2016.

- [22] W. Williem and I. K. Park, "Robust light field depth estimation for noisy scene with occlusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4396–4404.
- [23] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.
- [24] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2006, pp. 519–528.
- [25] F. Yu and D. Gallup, "3D reconstruction from accidental motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3986–3993.
- [26] H. Ha, S. Im, J. Park, H.-G. Jeon, and I. So Kweon, "High-quality depth from uncalibrated small motion clip," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5413–5421.
- [27] Z. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 166–176, Jun. 2015.
- [28] S. Heber and T. Pock, "Shape from light field meets robust PCA," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 751–767.
- [29] I. Tosic and K. Berkner, "Light field scale-depth space transform for dense depth estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2014, pp. 441–448.
- [30] O. Johannsen, A. Sulc, and B. Goldluecke, "What sparse light field coding reveals about scene structure," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3262–3270.
- [31] S. Heber and T. Pock, "Convolutional networks for shape from light field," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3746–3754.
- [32] M. Strecke, A. Alperovich, and B. Goldluecke, "Accurate depth and normal maps from occlusion-aware focal stack symmetry," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 2814–2822.
- [33] S. B. Kang, R. Szeliski, and J. Chai, "Handling occlusions in dense multi-view stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2001, pp. I-103–I-110.
- [34] V. Vaish, M. Levoy, R. Szeliski, C. L. Zitnick, and S. B. Kang, "Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2006, pp. 2331–2338.
- [35] J. Yu, L. McMillan, and S. Gortler, "Surface camera (SCAM) light field rendering," *Int. J. Image Graph.*, vol. 4, no. 4, pp. 605–625, 2004.
- [36] D. N. Wood *et al.*, "Surface light fields for 3D photography," in *Proc. 27th Annu. Conf. Comput. Graph. Interaction Techn.*, 2000, pp. 287–296.
- [37] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 1–14.
- [38] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2002, pp. 82–96.
- [39] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3017–3024.
- [40] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, "Constant time weighted median filtering for stereo matching and beyond," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 49–56.
- [41] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.
- [42] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4D light fields," in *Proc. Vis., Modeling Vis.*, 2013, pp. 225–226.
- [43] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2016, pp. 19–34.
- [44] Lab Code: Multi View Stereo With TGV-L1 Regularization. Accessed on May 10, 2017. [Online]. Available: <http://hci-lightfield.iwr.uni-heidelberg.de/benchmark/algorithm/mv>



**Hao Sheng** received the B.S. and Ph.D. degrees from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2003 and 2009, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, Beihang University. His research interests include computer vision, pattern recognition, and machine learning.



**Shuo Zhang** received the B.S. degree from the School of Information Engineering, Zhengzhou University, Henan, China, in 2012. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Beihang University, Beijing, China. From 2015 to 2016, she was a Visiting Student with the Computer Science And Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge. Her research interests include machine learning, computer vision, and computational photography.



**Xiaochun Cao** received the B.S. and M.S. degrees in computer science from Beihang University, Beijing, China, and the Ph.D. degree in computer science from the University of Central Florida, Orlando, FL, USA. He has been a Professor with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, since 2012.



**Yajun Fang** received the Ph.D. degree with the Department of Electrical Engineering and Computer Science, MIT. She is currently a Research Scientist with Intelligent Transportation and Research Center, MIT. Her research areas focus on machine vision, machine learning, big data analysis, and system theory. She is the Coordinator of the MIT Universal Village Program, and the Conference Chair of the International Conference on Universal Village.



**Zhang Xiong** received the B.S. degree from Harbin Engineering University in 1982 and the M.S. degree from Beihang University, Beijing, China, in 1985. He is currently a Professor with the School of Computer Science and Engineering, Beihang University. His research interests include computer vision, information security, and data vitalization.