Micro-Lens-Based Matching for Scene Recovery in Lenslet Cameras

Shuo Zhang, Hao Sheng, *Member, IEEE*, Da Yang, Jun Zhang, and Zhang Xiong

***Abstract*— Since a light-field camera is able to capture more information than a traditional camera, a lot of methods, such as depth estimation, image super-resolution, and view synthesis, are explored for recovering scene information. In this paper, we propose a novel framework for scene recovery based on lenslet-based light-field camera images. Instead of using traditional matching terms, we design a new micro-lens-based matching term to calculate structure information and recover several kinds of scene information simultaneously. On the one hand, inherent information in micro-lens images is selected to complement details in sub-aperture images. On the other hand, sub-aperture images are used to expand micro-lens images and synthesize new view images. A new micro-lens-based consistency metric is introduced for the matching term to handle occlusions in depth estimation and image reconstruction. The newly appeared and newly occluded areas in synthesized views are analyzed and recovered based on information from surrounding points. Experimental results show that the proposed depth estimation method outperforms state-of-the-art methods on both synthetic and lenslet-based light-field images, especially in low-texture and occlusion regions. Furthermore, the super-resolution and view synthesis methods are able to acquire view images with more details and less aliasing artifacts.**

***Index Terms*— Light field, depth estimation, reconstruction, super-resolution, view synthesis, micro-lens-based.**

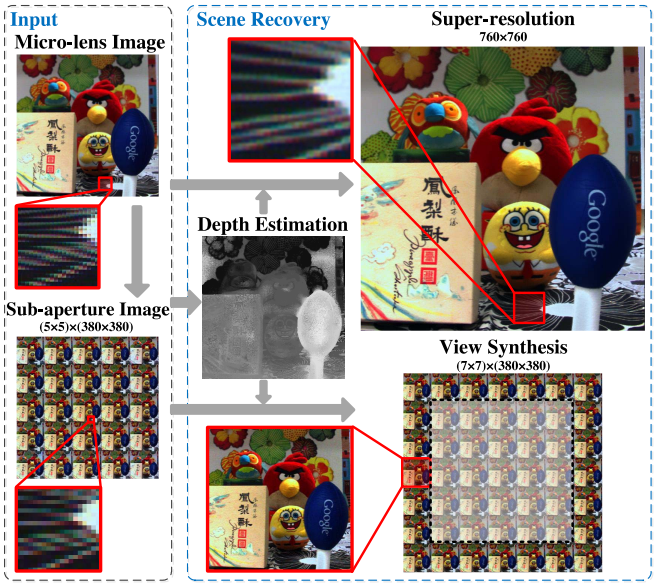


Fig. 1. The micro-lens images and the sub-aperture images extracted from the Lytro camera. The sub-aperture image has aliasing artifacts and low resolution due to the sparse sampling in angular domain. The proposed framework is able to estimate depth information, recover higher resolution sub-aperture images and synthesize new views simultaneously.

I. Introduction

LENSLET-BASED plenoptic cameras [1], *e.g.* Lytro [2] and Raytrix [3], have attracted a lot of attention recently. By inserting a micro-lens array between the main lens and the imaging plane, these cameras are able to capture the accumulated intensity of light together with its direction. Based on this design, multi-view images of the scene are captured simultaneously in one simple shot. The abundant information in light field images enables a wide range of applications, such as light field rendering [4], digital refocusing [2], super-resolution [5]–[7], 3D reconstruction [1], [8], saliency detection [9], [10] and material recognition [11].

However, the micro-lens array in plenoptic cameras also causes problems in light field image processing: 1) The local vignetting degrades the image quality, especially for border pixels in micro-lens images [12]. 2) As intensity values are recorded according to the arrangement of each micro-lens in the camera, adjacent pixels in sub-aperture images are extracted separately in each micro-lens. Therefore, sub-aperture images have aliasing artifacts due to sparse sampling in space [13], as shown in Fig. 1. 3)As plenoptic light field images record angular and spatial information, the sampling is sparse in both domains and the resolution of each sub-aperture image is relatively lower compared with traditional images. For example, the raw image resolution of Lytro Illum camera is 7728×5368, whereas the image resolution from one viewpoint is only 625×324. 4) Due to the hardware configuration, view

Manuscript received February 2, 2017; revised June 14, 2017 and August 28, 2017; accepted September 30, 2017. Date of publication October 17, 2017; date of current version December 5, 2017. This work was supported in part by the National Key R&D Program of China under Grant 2017YFB1002000, in part by the National Natural Science Foundation of China under Grant 61472019, 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 in part by the Open Fund of the State Key Laboratory of Software Development Environment under Grant SKLSDE-2017ZX-09. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrick Le Callet. *(Corresponding author: Hao Sheng.)*

S. Zhang, H. Sheng, D. Yang, and Z. Xiong are with the State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing 100191, China, and also with the Shenzhen Key Laboratory of Data Vitalization, Research Institute in Shenzhen, Beihang University, Shenzhen 518057, China (e-mail: shuo.zhang@buaa.edu.cn; shenghao@buaa.edu.cn; da.yang@buaa.edu.cn; xiongz@buaa.edu.cn).

J. Zhang is with the Department of Electrical Engineering and Computer Science, University of Wisconsin–Milwaukee, Milwaukee, WI 53201 USA (e-mail: junzhang@uwm.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.2017.2763823

images are captured with a narrow baseline, which is difficult to address in 3D reconstruction. The tradeoff between spatial and angular resolution is one of the major technical issues for light field cameras.

Besides the image quality problem, how to extract structure information from light field images is also crucial to most applications. In order to estimate depth information, plenty of methods have been developed based on sub-aperture images. The low resolution and highly aliasing sub-aperture images with the narrow baseline become a challenge for depth estimation. Some traditional problems in depth estimation, such as occluded, noisy and textureless regions, are still difficult to solve in light field images.

In order to recover scene information from light field images, methods super-resolution and view synthesis are proposed respectively. However, all these problems are solved independently and each method has its own requirements or assumptions which are difficult to satisfy. For example, images for some occlusion-aware depth estimation methods are assumed to have little noise or high angular resolution [14]–[16]. Some super-resolution and view synthesis methods are developed based on specifically designed depth maps [7], [12], [17]–[19]. In this case, reconstruction errors in these methods affect each other and each solution has to be adjusted differently depending on each application.

In this paper, we propose a novel framework to recover scene information from light field images, including depth estimation, image reconstruction, super-resolution and view expansion, as shown in Fig. 1. We solve these problems simultaneously by designing a new matching term and related reconstruction methods for lenslet-based plenoptic cameras. Specifically, the micro-lens image is treated as a unit to compare with the reference sub-aperture image. The structure information is recovered by analyzing the matching relationship between them. On one hand, micro-lens images are selected to complement details in sub-aperture images with higher resolution. On the other hand, sub-aperture images are used to expand micro-lens images and synthesize new view images from more viewpoints. Experimental results show that the proposed depth estimation method outperforms the state-of-the-art methods in both qualitative and quantitative evaluation, especially along occlusion boundaries and in low-texture regions. The reconstructed and super-resolved sub-aperture images using our method have fewer artifacts and achieve higher quality in various scenes than the other methods. Moreover, the synthesized novel views keep complete object boundaries and show state-of-the-art performance on a variety of real-world scenes. The main contributions of our work include:

1) A new light field image processing framework for scene recovery, including depth estimation, image reconstruction, super resolution and view expansion.
2) A novel micro-lens based matching term, which is designed for plenoptic light field camera images, to capture the scene structure information.
3) A series of metrics to handle occlusion problems in different applications in order to achieve accurate results along depth boundaries.

II. Related Work

In this section, we briefly review the corresponding state-of-the-art methods for depth estimation, image reconstruction, super-resolution and view synthesis.

A. Depth Estimation

The scene structure is important to light field image processing and ongoing efforts are being made to develop depth estimation methods. As sub-aperture images are similar to traditional multi-view images, the traditional sum of absolute differences (SAD), square differences (SSD) and gradient differences (GRAD) can be used for light field depth estimation [20]–[22]. Heber and Pock [23] introduced a novel matching term based on the Robust Principal Component Analysis (RPCA) to estimate depth information in light field images. Since the baseline of light field images is narrow, Epipolar Plane Images (EPIs) are also used for depth estimation. Wanner and Goldluecke [14], [24] proposed a structure tensor based method to measure slopes of lines in EPIs. Zhang *et al.* [25] proposed a spinning parallelogram operator (SPO) to locate lines in EPIs, which is robust to occlusions. According to the light field focal stack symmetry, Lin *et al.* [26] and Strecke *et al.* [27] designed new cost volumes based on focal stacks to extract scene depth information.

For noisy and aliased plenoptic camera images, Bishop and Favaro [13] used an anti-aliasing filter to avoid cross-talk image artifacts before depth estimation. They formulated a photo-consistency constraint designed for plenoptic cameras and performed matching directly on raw data. Tao *et al.* [20] proposed to combine defocus with correspondence to estimate smooth depth maps. Jeon *et al.* [21] designed a sub-pixel cost for lenslet light field images, where the final depth map was refined through discrete optimization and iterative quadratic fitting. However, these robust depth estimation methods designed for handling low-texture and noisy regions often produce inaccurate estimation along depth boundaries because of the over-smooth processing.

For occlusion problems, the concept of angular sampling images in [28] and [29], the same as the Surface Camera in [15], is proposed to analyze occlusions. Chen *et al.* [15] developed a bilateral consistency metric (BCM) to calculate occlusions by analyzing the surface cameras. Wang *et al.* [22] excluded occlusions by separating two regions in the angular domain and only using the region with fewer occlusions in their cost volume. Sheng *et al.* [29] modeled occlusions based on the central sub-aperture image and calculated the occlusion probabilities in angular sampling images. Strecke *et al.* [27] introduced partial focal stacks to deal with occlusions. Williem and Park [30] proposed different occlusion detection methods for defocus and corresponding cues to exclude occlusion points. Theoretically, these occlusion-aware methods chose to ignore information of possible occlusions in their depth calculation. In this case, for tiny structures, noisy or low-texture regions, the effective information, *e.g.* the higher contrast information in stereo matching methods, is kept too little to recover robust estimation.

B. Image Reconstruction

Since the micro-lens array is placed between the main lens and the image sensor in plenoptic cameras, the raw light field images are composed of small micro-lens images. Current methods for image reconstruction include two steps: 1) Parameterize the 4D light field images and extract sub-aperture images from raw images; 2) Reconstruct sub-aperture images with more details based on structure information.

Due to manufacturing defects, the micro-lens array has a small rotation angle and does not perfectly align with image sensor coordinates [12]. In the calibration and sampling methods, Cho *et al.* [12] presented step-by-step procedures to achieve accurate calibration including Gamma correction, demosaicking, rotation estimation and center pixel estimation. Using checkerboards with known size, Bok *et al.* [31] proposed to calibrate the camera based on raw light field images. In [32], a 15-parameter camera model was proposed to represent the physics of lenslet-based cameras and procedures including decoding, calibration and rectification were used.

In the image reconstruction and super-resolution methods, Bishop *et al.* [5] proposed to iteratively estimate depth maps and reconstruct sub-aperture images. Cho *et al.* [12] proposed several interpolation techniques for pixel resampling in light field images, including barycentric interpolation, refinement according to structure information and a learning-based interpolation. Mitra and Veeraraghavan [17] provided a framework for light field denoising, angular and spatial super-resolution. They modeled light field patches using a Gaussian mixture model (GMM) and then reconstructed images based on the estimated disparity. Wanner and Goldluecke [7], [24] filled in additional pixels by interpolating lines in EPIs based on each pixel's direction. Without structure information, state-of-the-art super-resolution methods [33] for traditional images can be applied directly to sub-aperture images. For light field images, Yoon *et al.* [34] developed a deep convolutional neural network (CNN) to up-sample angular resolution as well as spatial resolution. They developed the spatial and angular networks independently and fine-tuned via end-to-end training. The structure information used in these methods is often estimated especially to recover scene images without artifacts. Moreover, these methods did not consider the influence of occlusions so that the reconstruction errors increase along occlusion boundaries.

In our paper, the dataset from Lytro cameras in the experiments are provided after calibration or using basic calibration methods as [12]. We focus on image reconstruction and super-resolution using structure information. The learning-based methods, which recover high-resolution images without structure information, can be used for further image super-resolution.

C. View Synthesis

Besides the methods for angular super-resolution in [17], [24], and [34], some methods are designed especially for light field view synthesis. Zhang *et al.* [18] presented a phase-based method to reconstruct 4D light field from a micro-baseline stereo pair. They iteratively optimized disparity maps to minimize differences between warped view image and ground truth input. Pujades *et al.* [35] developed a physics-based generative model to synthesize novel views from a set of input images. They showed that their Bayesian model improves the quality of novel views, in particular, if the scene geometry estimation is accurate. Kalantari *et al.* [19] proposed a learning-based approach to synthesize internal views using only four corner sub-aperture views. They developed two sequential convolutional neural networks to estimate disparities and novel view images respectively.

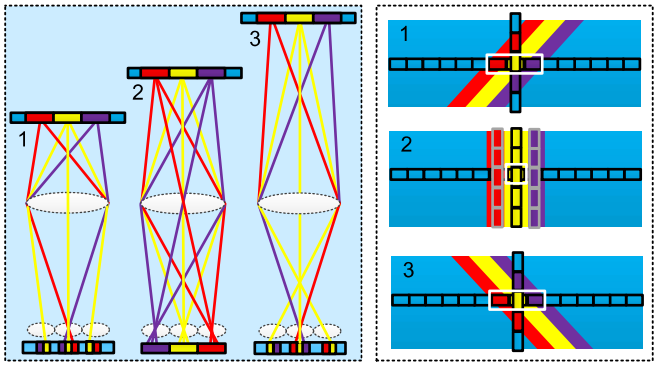


Fig. 2. The optical diagram in lenslet cameras and the corresponding epipolar plane images (EPIs). The size and direction of the region in the sub-aperture image (white border in the horizontal line in the EPI), which corresponds to the micro-lens image (the vertical line in the EPI), changes when the depth changes.

Most of the methods synthesize novel views from a sparse set of input views and the novel view points are located inside the input scope, which is a view interpolation problem. However, besides the sparse sampling in angular domain, the small baseline of the current light field camera is still difficult to address in most applications. In order to expand the baseline of the light field camera and improve the quality of 3D reconstruction, we focus on view extrapolation problems in light field images. For view extrapolation, the novel view images contain more uncertain information and are more difficult to synthesize, such as regions that are visible (invisible) in new views but invisible (visible) in input views. Most of the current methods cannot produce high-quality expanded view images, especially along occlusion boundaries.

III. Micro-Lens-Based Matching

In this paper, the 4D light field parameterization [36] is used and the light field is denoted as $L(x, y, u, v)$, where (x, y) are coordinates of pixels in spatial domain and (u, v) are coordinates of sub-aperture images in angular domain. The micro-lens image is calculated as $M = L(x^*, y^*, u, v)$ and the reference sub-aperture image is $V = L(x, y, u^*, v^*)$. In order to explain the correspondence relationship, the Epipolar Plane Image (EPI) [37] is introduced. The horizontal EPI in Fig. 2 is calculated by fixing two coordinates in different planes and changing the others as $E = L(x, y^*, u, v^*)$. We assume that all the scenes are Lambertian in the rest of the paper.

In this section, we propose to treat each micro-lens image as one unit and design a new matching term, which evaluates

similarities between micro-lens images and sub-aperture images. Instead of using sub-aperture images in traditional stereo matching methods, we choose to use micro-lens images because: 1) Points in micro-lens image are recorded adjacently in plenoptic cameras, which have less aliasing problems. 2) Micro-lens images are corresponding to one region in sub-aperture images, which provide various information and are robust for depth estimation. 3) Compared with the window-based matching, the proposed method is easier because we only calculate the correspondence of one pixel in one view image. 4) The proposed term is intuitional, which can be used directly for further image reconstruction and view synthesis.

A. Micro-Lens-Based Matching Term

We first consider the matching term for points without occlusions. As shown in Fig. 2, vertical lines in EPIs correspond to horizontal lines in micro-lens images, and horizontal lines in EPIs correspond to horizontal lines in sub-aperture images. The disparity of the point, *i.e.* the slope of the line, decides the size of the region in the central sub-aperture that matches the micro-lens image. Based on these observations, we design a new matching term to estimate the structure information.

The new matching term measures the square differences between the micro-lens image $M_{x,y}$ and the region in the reference sub-aperture image $V_{x,y}$ as:

$$C_{x,y}(l) = \sum_{i,j} \min((M_{x,y}(i,j) - V^l_{x,y}(i,j))^2, \tau), \qquad (1)$$

where (x, y) indicates the pixel position in the reference sub-aperture image and (i, j) is the pixel position in the micro-lens image. τ is a truncation value of a robust function, which has been widely used in traditional stereo matching methods [38], [39]. Truncating the matching cost limits wrong matches and can help the depth estimation for occluded regions. The region in reference sub-aperture image that has the minimum matching cost is supposed to match the micro-lens image and the corresponding depth label indicates structure information.

Suppose that the disparity $l = \Delta x/\Delta u$, the points in the selected region $V^l_{x,y}(i, j)$ that correspond to $M_{x,y}(i, j)$ are calculated as:

$$V^l_{x,y}(i,j) = V(x + l \times i, y + l \times j), \qquad (2)$$

where (i, j) corresponds to the pixel position in micro-lens images and the reference central point is regarded as the origin of the coordinate system. Suppose that the size of the micro-lens image is $h \times h$, which is 9×9 for most synthesized light field images. The size of the corresponding region in the sub-aperture image is denoted as $(h \times l) \times (h \times l)$. It is worth noting that if the scene is behind the focusing plane of the main lens, the micro-lens image is inverted compared with the scene view.

B. Comparison With Traditional Matching Terms

In traditional stereo matching, the consistency of one point in all sub-aperture images is calculated. The proposed matching term provides a new point of view to estimate depth for lenslet camera images, which treats the micro-lens image as a unit instead of one specific point. As shown in Fig. 2, the traditional stereo matching methods find slant lines that have same colors in EPIs, whereas we find horizontal lines which correspond to vertical lines in EPIs. In our approach, different matching points are chosen in different matching view images and their square differences are added as the matching cost of the reference point.

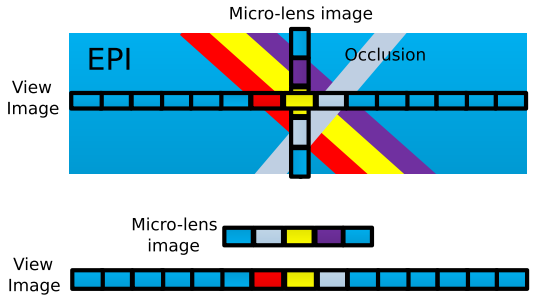


Fig. 3. The synthesized epipolar plane image with occlusions, where the occlusion point breaks the lines that lie behind it. The region in the sub-aperture image cannot perfectly match the micro-lens image.

The proposed method is similar to window-based matching methods because the correspondence of one whole region is used. The difference is that in one view, we only find the correspondence of one point, whereas the traditional window-based methods find the correspondence of one fixed window. Our computational complexity is lower than window-based matching methods. The size of the matching window in the sub-aperture image is determined by the possible depth values. As our local cost is calculated based on multiple points so that it has a good performance in textureless regions, which is similar to the traditional window-based matching term.

C. Occlusion Handling

For an unoccluded point, its micro-lens image is well matched with a specific region in the sub-aperture image as shown before. However, if points in micro-lens image are occluded in a complex scene, we cannot find a perfect match in sub-aperture images. As shown in Fig. 3, the object in the front occludes the object in the back so that the corresponding region is different from the micro-lens image. Another problem for micro-lens-based matching is the vignetting effect, which is caused by the micro-lens array in plenoptic cameras. Border pixels in micro-lens are darker than the central one because of the manufacturing technology and cause matching problems. Most of the methods [12], [20], [21] choose to ignore all border pixels and only use central pixels to calculate the matching term. As a result, the angular resolution becomes lower.

In order to exclude occluded points in depth estimation, the assumption that occlusions have different colors from occluders is widely used in occlusion-aware depth estimation [15], [16]. Based on this, we propose to estimate the reliability of each pixel in micro-lens images by introducing a new micro-lens-based consistency metric (MCM):

$$W_{x,y}(i,j) = \exp\left(-\frac{(M_{x,y}(i,j) - M_{x,y}(c_i, c_j))^2}{\sigma^2}\right), \qquad (3)$$

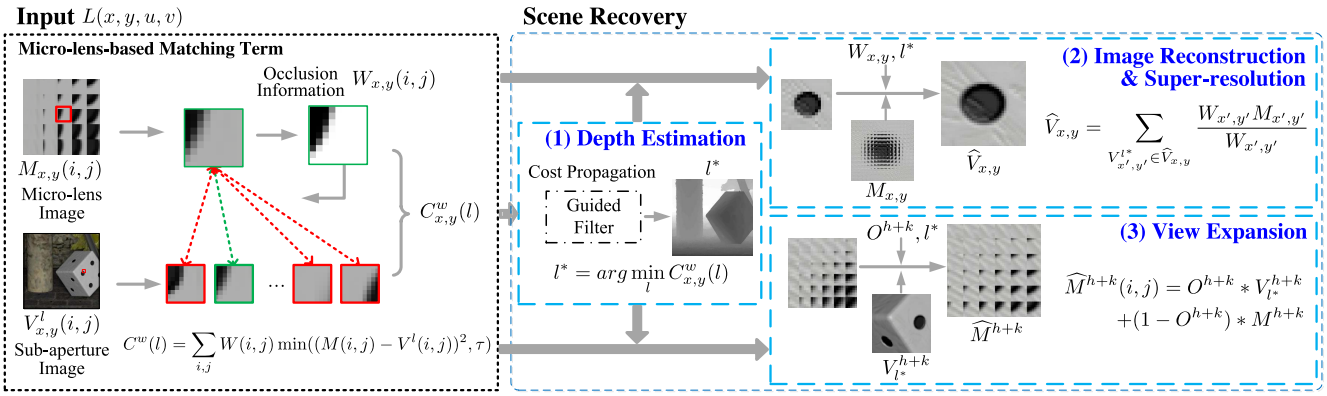


Fig. 4. The proposed framework for scene recovery. The structure information is first estimated based on the micro-lens-based matching term, where the occlusion information is introduced. Then micro-lens images are used to fill in details of sub-aperture images and sub-aperture images are used to reconstruct micro-lens images with expanded view-points.

where $M_{x,y}$ is the corresponding micro-lens image. $M_{x,y}(c_i, c_j)$ is the central reference point in the micro-lens image. The points $M_{x,y}(i, j)$ in the micro-lens image that have high contrast compared with the reference central point $M_{x,y}(c_i, c_j)$ are assigned low weights in MCM. These points are more likely to have different depth values with the reference central point and correspond to different regions in sub-aperture images.

Finally, we combine the estimated MCM and the micro-lens-based matching term to compute the cost volume as:

$$C^w(l) = \sum_{i,j} W(i, j) \min((M(i, j) - V^l(i, j))^2, \tau). \quad (4)$$

Using the MCM, wrong matching costs of the reference point in Eq. (1) are excluded due to the low weights and the cost of the right depth should still be the minimum in Eq. (4).

D. Comparison for Occlusion Handling

Depth maps of occluded regions are usually difficult to recover using traditional stereo matching methods. As mentioned, ASIs and SCam, which are constructed by extracting possible matching points in all sub-aperture images, are introduced for depth estimation. In the ASI, the color is constant for unoccluded points and the consistency is broken when the point is occluded. The occlusion-aware methods [15], [16], [29] proposed different metrics to weight occlusion probabilities in ASIs. The points which are treated as possible occlusions are excluded in the matching calculation. In the proposed method, the MCM excludes points that have high contrast with the reference point in micro-lens images and only the similar points are kept to calculate the matching cost. The truncation parameter τ in Eq. (4) further prevents wrong matches from possible occlusions. Notice that points in micro-lens images are different in the scene, whereas points in ASIs correspond to one same point in the scene at the true depth.

If objects have tiny structures in space, it is difficult to recover depth information accurately in traditional depth estimation methods. The shape of the object will be dilated when it occludes others and eroded when it is occluded by surrounding objects. For these occlusion-aware methods, these occlusions are too difficult to be separated from the matching points in the ASI because the matching points may be a small proportion. Finally, discontinuous structures are shown in their depth maps. However, in the proposed method, matching costs from similar surrounding points in micro-lens images are combined and are more robust to estimate correct depth labels.

For textureless regions, points with effective information are too little for obtaining correct depth labels in the previous occlusion-aware methods. In our method, if the region is textureless, all the points in the micro-lens image are used for the cost calculation. Similar with the advantages of traditional window-based methods, our method achieves robust depth estimation because more information from similar surrounding points is combined.

IV. Scene Recovery

Once the correspondence relationship is established, scene information can be recovered. As shown in Fig. 4, structure information of the scene is first estimated based on the matching volumes. The corresponding micro-lens images are then used to fill in details of sub-aperture images with higher resolution. From another point of view, the corresponding sub-aperture images are used to reconstruct the micro-lens images with larger view scope, where the synthesized views from expanded viewpoints are available.

A. Depth Estimation

In order to obtain a smooth depth map, we use an edge-preserving guided filter [40] to propagate reliable information to textureless regions. The guided filter is implemented for the cost volume in each depth label as in [41] to smooth the costs. The filter-based method is more efficient than the graph-based optimization and can be easily parallelized. Finally, the depth map is calculated using the winner-take-all strategy as:

$$l^*_{x,y} = arg\min_l C^w_{x,y}(l). \quad (5)$$

Algorithm 1 shows the complete depth estimation algorithm.

Algorithm 1 Depth Estimation

Require: Light Field Image $L(x, y, u, v)$
1: Extract the central reference image $V = L(x, y, u^*, v^*)$
2: **for** all points (x, y) in V **do**
3: Extract the micro-lens image $M_{x,y} = L(x^*, y^*, u, v)$;
4: Compute the micro-lens-based consistency metric $W_{x,y}$ based on Eq. (3);
5: **for** all $l \in [l_{min}, l_{max}]$ **do**
6: Compute $V^l_{x,y}$ in the reference image in Eq. (2);
7: Compute the matching cost $C^w_{x,y}(l)$ based on Eq. (4);
8: **end for**
9: **end for**
10: **for** all $l \in [l_{min}, l_{max}]$ **do**
11: Apply the guided filter to the cost volume $C^w(l)$;
12: **end for**
13: Compute depth label as: $l^*_{x,y} = arg\min_l C^w_{x,y}(l)$;

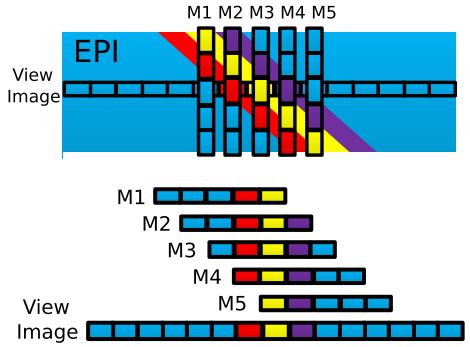


Fig. 5. The synthetical EPI and micro-lens images for image reconstruction. Once the correspondence relationship is calculated, the region in the sub-aperture image is reconstructed using multiple micro-lens images.

B. Image Reconstrution and Super-Resolution

As shown in Fig. 5, regions in sub-aperture images can be reconstructed using multiple micro-lens images based on estimated depth labels. Specifically, the reconstructed image $\widehat{V}$ is calculated as the weighted average of the related micro-lens images:

$$\widehat{V}_{x,y} = \frac{\sum_{V^{l^*}_{x',y'}(i,j)\in\widehat{V}_{x,y}} W_{x',y'}(i,j) M_{x',y'}(i,j)}{\sum_{V^{l^*}_{x',y'}(i,j)\in\widehat{V}_{x,y}} W_{x',y'}(i,j)}, \quad (6)$$

where $V^{l^*}_{x',y'}(i,j)$ is the pixel that $M_{x',y'}(i,j)$ matches at the estimated depth label l^*. The pixels in the micro-lens images, which are located in the corresponding position $\widehat{V}_{x,y}$ are collected to recover the image. In order to deal with occlusion errors, the $W_{x,y}(i,j)$ defined in Eq. (3) is applied to measure the confidence of each pixel in image reconstruction. The pixels that have higher confidence contribute more to the reconstructed images. If the pixel is considered to be possibly occluded in the micro-lens image, it is assigned a smaller weight in the reconstruction process.

Since micro-lens images may be located at the sub-pixel position in the reconstructed images $\widehat{V}$, the super-resolved sub-aperture image is easily acquired according to Eq. (6). As analyzed in [17], the maximum super-resolution of the image is related to its disparity and the angular resolution. If the point moves more quickly between the neighboring sub-aperture images, more micro-lens images are used to reconstruct the sub-aperture image. If the point moves more slowly, the larger region in one micro-lens image is used to reconstruct the sub-aperture image. Suppose that the size of the micro-lens image is $h \times h$, the disparity of one point is l and the super-resolution image is k times larger than the original image. If the pixel in the super-resolution image can find its corresponding value in the light field image, its disparity should have $|l| > |(k-1)/(h-1)|$. If the disparity is too small, the weight of the pixel may have $\sum_{V^{l^*}_{x',y'}(i,j)\in\widehat{V}_{x,y}} W_{x',y'}(i,j) = 0$ in Eq.(6) and no information is available for reconstruction. In order to have a large super-resolved image, the disparity should be large enough to contain effective information. However, traditional light field images in plenoptic cameras often have a narrow baseline and the angular resolution is less than 9×9. Therefore, we choose $k = 2\,or\,3$ so that most points in the images satisfy $|l| > 0.25$. Finally, the pixels that have no matched points from the micro-lens images are recovered using the traditional inpainting method [42]. The complete super-resolution algorithm is shown in Algorithm 2.

Algorithm 2 Super Resolution

Require: Light Field Image $L(x, y, u, v)$, Depth Label l^*
1: **for** all points (x, y) in V **do**
2: Extract the micro-lens image $M_{x,y} = L(x^*, y^*, u, v)$;
3: Compute the micro-lens-based consistency metric $W_{x,y}$ based on Eq. (3);
4: Find the corresponding region position $V^{l^*}_{x,y}$ in the reconstructed image according to l^*;
5: Add all the corresponding $M_{x,y}$ together according to $V^{l^*}_{x,y}$ as in Eq. (6), the same with $W_{x,y}$;
6: **end for**
7: Reconstruct the super-resolved reference image $\widehat{V}$ based on Eq. (6);

We compare the proposed reconstruction method with the digital refocusing method in [2]. The refocusing method also uses micro-lens images to construct a new image, which is similar to our method. The difference is that they accumulate micro-lens images to reconstruct the scene image based on a fixed depth value. Therefore, regions that do not match micro-lens images are blurred during the addition process. The refocusing result is an image which focuses at a fixed depth with some regions in-focus and others defocused. On the other hand, our method uses micro-lens images to reconstruct regions based on their depth values. The matching cost is defined as the similarity between the two regions so that all the regions are replaced with the most similar micro-lens images. As a result, all-in-focus images with more details and higher resolution are acquired.

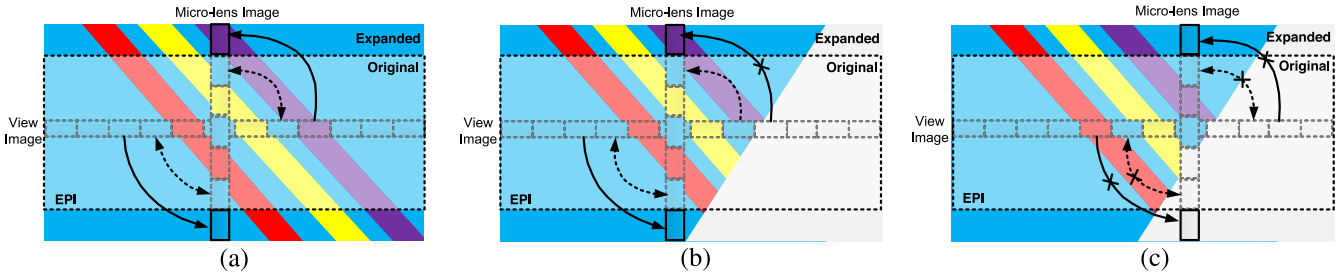


Fig. 6. The synthetical EPI images for view expansion. (a) Without occlusions, the micro-lens image with larger size can be easily calculated using the corresponding view image according to its disparity. (b) For newly-appeared regions, a part of points are occluded in existing views but visible in the new view. (c) For newly-occluded regions, a part of points are visible in existing views but occluded in the new view.

C. View Synthesis

In this paper, we focus on the view expansion problem. The original micro-lens image M^h, whose resolution is $h \times h$, is expanded by k pixels and a larger micro-lens image $\widehat{M}^{h+k}$ is obtained.

For micro-lens image M^h, the corresponding region in the sub-aperture image is calculated as $V_{l^*}^h$ based on the estimated depth l^*. If the surrounding points have the same depth value, a larger micro-lens image $\widehat{M}^{h+k}$ is easily calculated using $V_{l^*}^{h+k}$ in the sub-aperture image. However, when the points in one micro-lens image have different depth values, the expanded micro-lens image cannot be recovered accurately based on the corresponding sub-aperture image. As we can imagine, some areas will appear and some will be occluded by other objects in the new viewpoint. These newly-appeared and newly-occluded areas may be estimated with artifacts along the depth boundaries in the new view as in [18].

In order to recover points that are occluded in existing views but visible in the new view, as shown in Fig. 6, we propose to use the structure information. Since the disparity of the sub-aperture image is estimated, we extract the depth of the corresponding region $D_{l^*}^{h+k}$ according to Eq. (2). If the depth of the expanded points is closer to the camera than the reference point, these regions are regarded as newly-appeared regions:

$$O_1^{h+k}(i,j) = (D^{h+k}(i,j) - D(c_i,c_j)) > T_1, \tag{7}$$

where $D(c_i, c_j)$ is the disparities of the reference point and T_1 is set as 1 in our experiments.

On the other hand, we have to deal points that are visible in existing views but occluded in the new view, as in Fig. 6. We identify the newly-occluded regions by comparing the micro-lens image and the corresponding sub-aperture image as:

$$O_2^h(i,j) = (M^h(i,j) - V^h(i,j))^2 < T_2, \tag{8}$$

where T_2 is set as 100 in our experiments. If occlusions exist, the two regions are inconsistent and their difference is big. For the unknown expanded region, we predict that it has the same trend with surrounding points. Therefore, we expand the region predictor $O_2^h(i,j)$ to $O_2^{h+k}(i,j)$ using a simple nearest interpolation operation.

These newly-occluded or newly-appeared regions cannot be recovered precisely. We propose to fill in these regions using similar regions around them. Specifically, we expand the original micro-lens image M^h into a larger size M^{h+k} using the nearest interpolation. Then the final expanded micro-lens image is calculated as:

$$\widehat{M}^{h+k}(i,j) = O^{h+k} * V_{l^*}^{h+k} + (1 - O^{h+k}) * M^{h+k}, \tag{9}$$

where $O^{h+k} = O_1^{h+k} \& O_2^{h+k}$. In this case, if and only if O_1^{h+k} and O_2^{h+k} are both true, the expanded regions are replaced with the sub-aperture images. If not, the pixel is assigned the color similar to its nearest neighbor in the micro-lens image. By doing so, the background content is stretched to fill in the newly-appeared regions and the foreground content is assigned to the newly-occluded part. The complete algorithm is shown in Algorithm 3.

Algorithm 3 View Synthesis

Require: Light Field Image $L(x,y,u,v)$, Depth Label $l^*_{u^*,v^*}$
1: Extract the central reference image $V = L(x,y,u^*,v^*)$
2: **for** all points (x,y) in V **do**
3: Extract the micro-lens image $M_{x,y} = L(x^*,y^*,u,v)$
4: Compute $V_{l^*}^{h+k}$ in the reference image based on Eq. (2)
5: Compute the depth label $D_{l^*}^{h+k}$ based on Eq. (2)
6: Compute the newly-appeared region O_1^{h+k}
7: Compute the newly-occluded region O_2^{h+k}
8: Reconstruct the micro-lens image $\widehat{M}_{x,y}^{h+k}$ in Eq. (9)
9: **end for**

The experiments show that k can be set as 1 or 2 for traditional Lytro camera images. For micro-lens images, whose effective size is restricted to 7×7 or less because of manufacturing defects and vignetting effects, the proposed method can help to increase the angular resolution. The number of views is also highly increased because the pixels are added around the existing angular scope. The added view images and the expanded baseline can improve the quality of 3D reconstruction and is helpful for most of the light field applications.

V. Experiments

In this section, we show experimental results including depth estimation, image reconstruction, super-resolution and view expansion.

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

Image	Cube		StillLife		Buddha		Mona	
	All	Occ	All	Occ	All	Occ	All	Occ
Wanner *et al.* [14]	**0.92**	13.29	4.30	8.35	2.14	15.01	12.52	20.44
Chen *et al.* [15]	1.11	9.72	1.88	8.37	1.72	8.31	10.00	16.48
Wang *et al.* [22]	1.44	11.46	3.99	12.79	6.20	13.52	13.23	22.26
Jeon *et al.* [21]	1.04	9.89	2.68	12.59	3.12	16.99	9.13	18.47
Ours	0.97	**9.32**	**1.55**	**6.83**	**1.24**	**6.04**	**7.25**	**12.40**

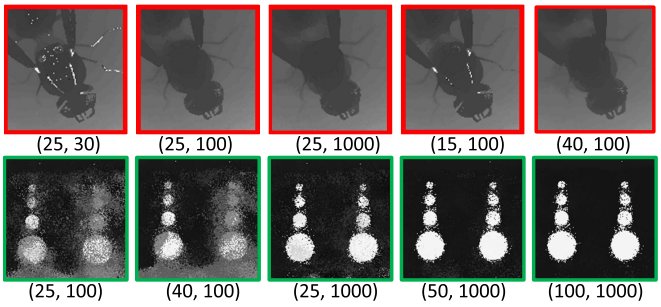


Fig. 7. The evluations for parameter τ and σ in Eq.(1) and Eq. (3). The value of τ and σ are illustrated in the bracket, respectively. The depth boundaries get ambiguous as σ and τ increases. However, if σ or τ is too small, depth boundaries become discontinuous. For the noisy scenes 'Dots' in the second row, once σ and τ drops below the noise level, the errors become very large. We choose $\sigma = 100$ and $\tau = 25$ in our experiments for most images.

A. Depth Estimatiom

We evaluate the performance of our depth estimation method on synthetic light field images from [22], [43], and [44] and Lytro Illum camera images from [22], which are provided after calibration. Results are compared with state-of-the-art depth estimation methods, including Wanner and Goldluecke [14], Zhang *et al.* [25], Johannsen *et al.* [45], Wang *et al.* [22], and Chen *et al.* [15]. The methods developed for Lytro camera images, such as Tao *et al.* [20] and Jeon *et al.* [21], are also evaluated for further comparisons.

As a quality measurement, we use the percentage of depth value below a relative error based on the ground truth as [14]:

$$BadPix(t) = \frac{\{|x \in M : |l(x) - gt(x)| > t\}|}{|M|}. \qquad (10)$$

We set $t = 5\% * L$ in our experiments, where L is the maximum disparity of the scene. The accuracy is calculated for the overall images (All) and occlusion regions (Occ) respectively. The $MSE * 100 = \frac{1}{|M|}\sum_{x \in M}(l(x) - gt(x))^2 * 100$ and $BadPix(0.07)$ metrics used for the new benchmark in [44] are also evaluated for the corresponding images.

1) Parameter Selection: In this subsection, we analyze the parameter τ and σ in Eq. (1) and Eq. (3) in Fig. 7. The scale σ determines the number of pixels for matching cost calculation. As σ decreases, the depth boundaries are more clear. However, if σ is too small, most of the information is lost during the calculation and causes discontinuous structures in the final depth maps. Similarly, τ controls the maximum cost value and is used to truncate the high matching cost from occlusions. We choose $\sigma = 100$ and $\tau = 25$ in our experiments for the traditional synthetical images.

For the noisy scenes, like the image in the second row in Fig. 7, once σ and τ drops below the noise level, the errors become very large. As analyzed in [38] and [39], the best parameters depend on the signal-to-noise-ratio (SNR). The suitable parameters should be large for images with low SNR and right above the noise level.

2) Occlusion Performance: In benchmark [44], the 'Backgammon' is designed to evaluate fine structures and occlusion boundaries. As shown in Fig. 8, our depth map shows clear thin structures and has fewer error pixels than the other methods. The depth estimation performance using synthetic light field images by Wanner and Goldluecke [14] is shown in Fig. 9. Our approach achieves better performance in maintaining small structures, like the antenna of an insect and the small hole. The numerical results are shown in Table I, where our method outperforms other occlusion-aware depth estimation methods in the occluded regions. Note that these numerical results cannot accurately show the advantage of our method for tiny structures. More detailed comparisons using synthetic light field images [22] are shown in Fig. 10. Our results show accurate boundaries and correct shapes of the small structures. Specifically, we obtain the accurate boundary of the twig, the correct shape of the decoration below the lampshade and the thin shapes of the lamp in different images.

Although Wang *et al.* [22] and Chen *et al.* [15] developed their methods for handling heavy occlusions, they still fail in these small occlusions. The reason is that they both tried to exclude occlusions based on the matching cost statistical analyses as in Sec. III-D. When the estimated point is spatially isolated, it is difficult to exclude occlusions since most of the surrounding points come from different depth levels. On the contrary, the micro-lens-based matching combines the costs of points that have similar color with the reference point and acquires complete structures.

3) Low-Texture Performance: Low-texture regions also cause matching problems in most depth estimation methods. In Fig. 8, the 'Stripes' is used to evaluate the influence of texture and contrast at occlusion boundaries [44]. Our method is able to keep more effective information than other depth estimation methods and shows reliable depth estimation in low-texture regions. In the bottom row in Fig. 9, the surface texture of the leaf is ambiguous and other methods propagate inaccurate depth information in these regions. By contrast, our method captures the unobvious texture on the leaves and performs well in textureless regions. In Fig. 10, our method recovers more accurate depth in ambiguous regions, such as the base of the table in the second image and the bright color in the head of the bed in the third image. Other methods smooth over these regions with background depth or generate

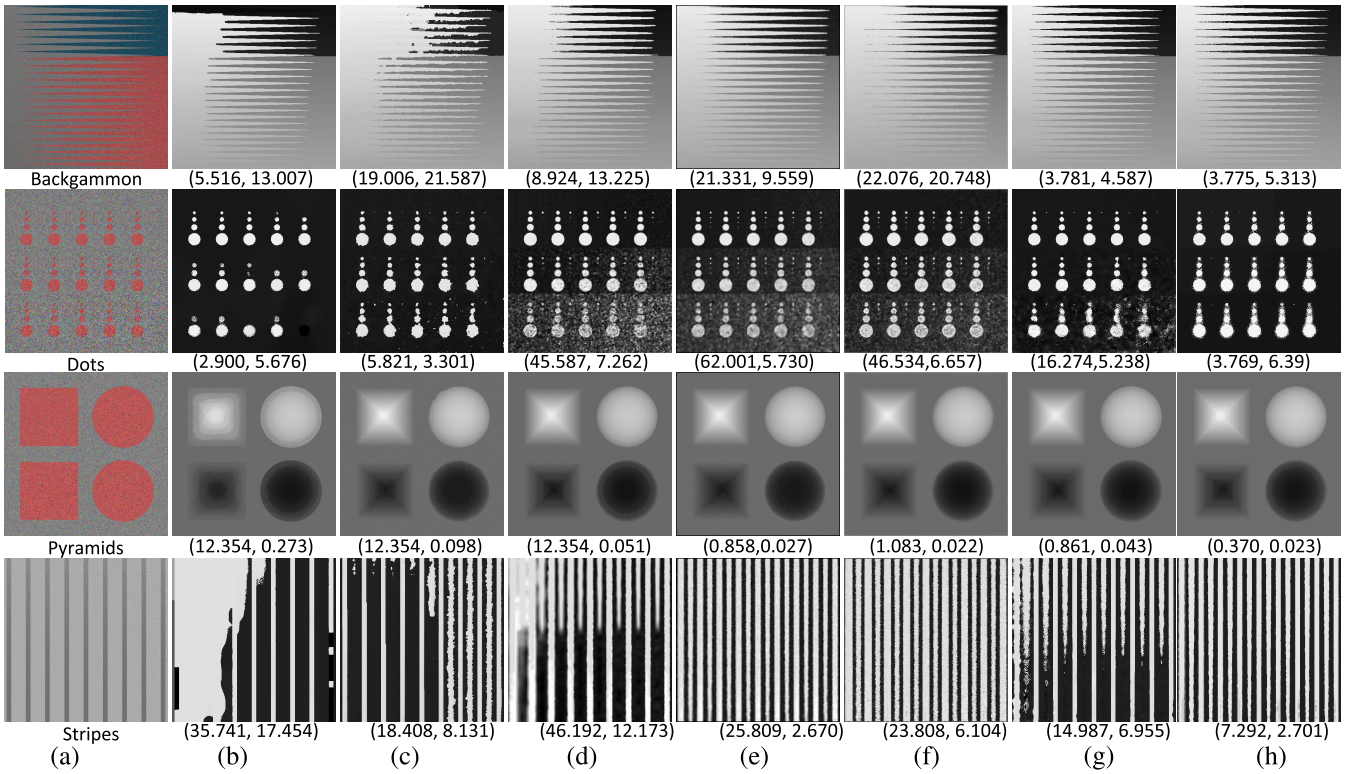


Fig. 8. The depth estimation results for benchmark [44] (a) view image (b) Jeon *et al.* [21] (c) Wang *et al.* [16] (d) MV [46] (e) Johannsen *et al.* [45] (f) Wanner and Goldluecke [47] (g) Zhang *et al.* [25] (h) Our results. The $BadPix$ and MSE metrics are also illustrated in the brackets, respectively. Our depth map shows clear thin structures in 'Backgammon', reliable depth estimation in low-texture regions in 'Stripes' and comparable results for noisy images 'Dots'.

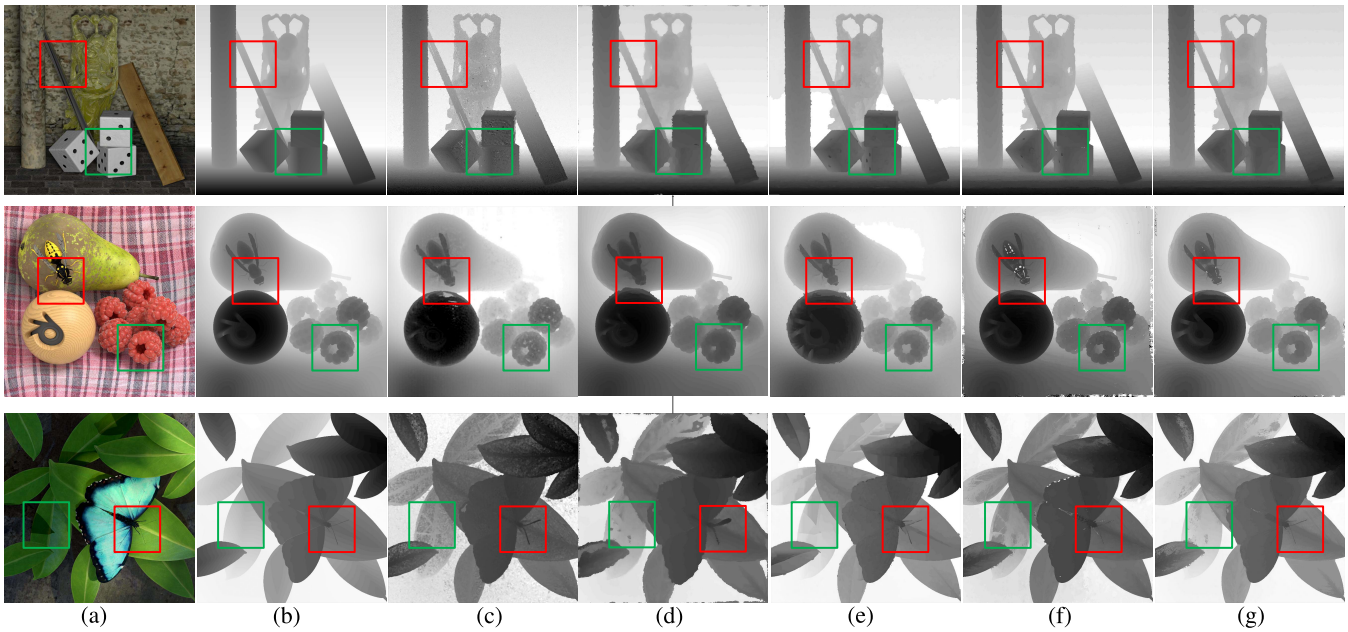


Fig. 9. The depth estimation results using synthetic light field images provided by Wanner and Goldluecke [14]. Our method achieves better performance for keeping the small structures, like the antenna of an insect and the small hole in the scenes. Moreover, our method also captures the unobvious textures on the leaves and achieves better depth estimation results in textureless regions. (a) Scene Image. (b) Ground Truth. (c) Wanner and Goldluecke [14]. (d) Jeon *et al.* [21]. (e) Wang *et al.* [22]. (f) Chen *et al.* [15]. (g) Ours.

inaccurate structures. The main reason is that we consider different pixels' correspondences in different views and reliable information is included as analyzed in Sec. III-B.

4) Overall Performance: For the overall performance, as illustrated in Table. I, our method achieves the state-of-the-art performance. Our performance for noisy regions

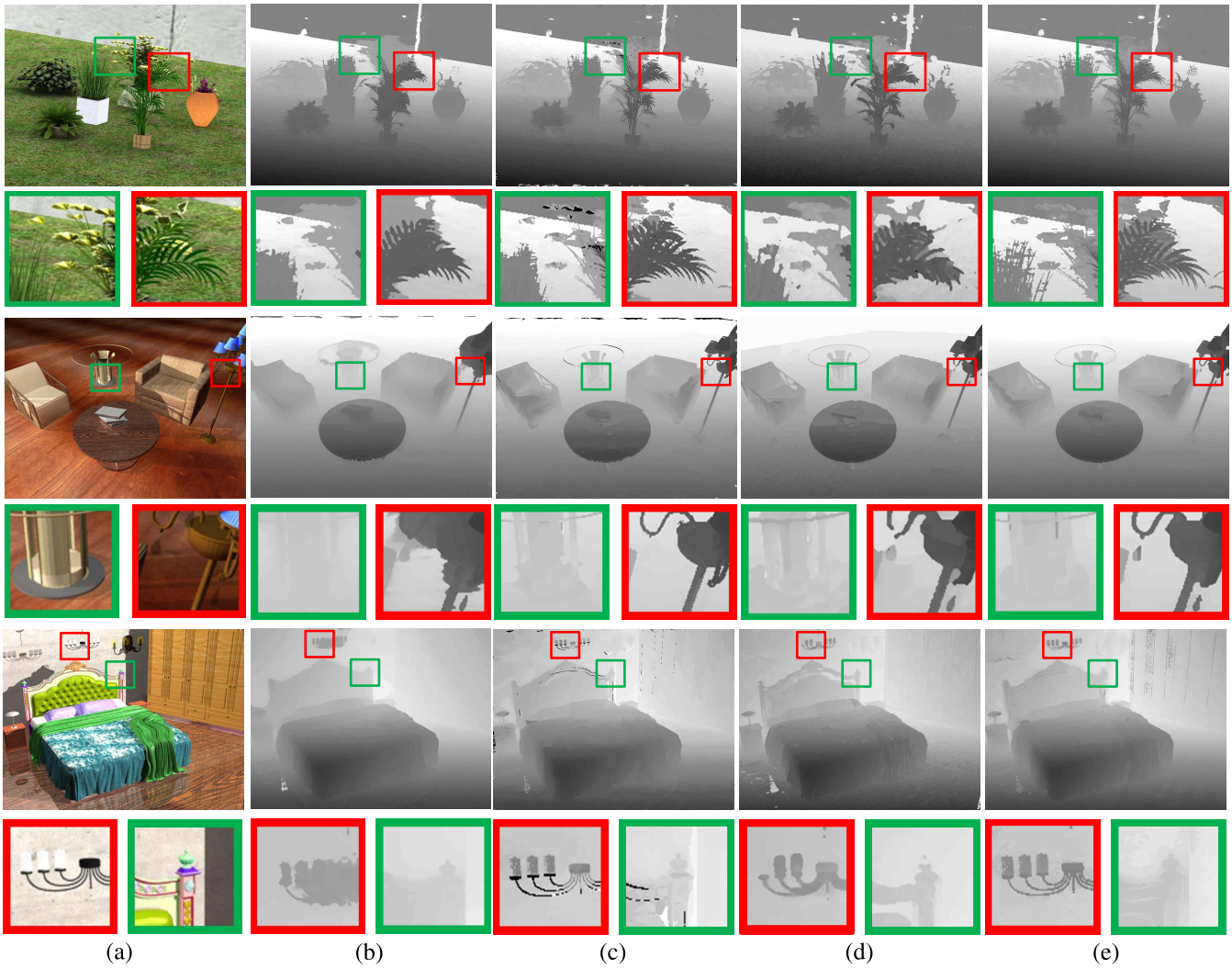


Fig. 10. Depth estimation results using synthetic light field images by Wang and Goldluecke [22]. In depth discontinuous regions, our results show accurate boundaries and correct shapes of the small structures, such as the boundary of the twig, the shape of the decoration below the lampshade and the shape of the lamp. In ambiguous regions, our method recovers more accurate depth values, such as the base of the table and the bright color in the head of the bed. Other methods smooth over these regions with background depth value or generate inaccurate structures. (a) Scene Image. (b) Jeon *et al.* [21]. (c) Chen *et al.* [15]. (d) Wang *et al.* [22]. (e) Ours.

and surface reconstruction is evaluated using 'Dots' and 'Pyramids' in Fig. 8, respectively, For noisy regions, we prefer more robust parameters. The depth of the noisy regions is reconstructed more completely in our method than others. For surface reconstruction, we set a small value as the radius of the filter in the cost propagation and achieve comparable results. The number of discrete disparity labels and the optimization methods which consider the surface modeling are the main factors that affect the surface reconstruction.

5) Lytro Images: The depth estimation results for Lytro Illum images are shown in Fig. 11. We compare our performance with Tao *et al.* [20] and Jeon *et al.* [21], who developed methods specifically for Lytro images. The real light field images contain more noise and artifacts than the synthetic light field images. Our method performs better around occlusion boundaries and thin objects, such as the circles in the hang decoration and thin branches. The other methods cannot obtain clear depth edges using these real light field images.

B. Image Reconstruction and Super-Resolution

In this experiments, some Lytro images are provided from Boominathan *et al.* [49], who developed a hybrid imaging system consisting of a standard LF camera and a high-resolution standard camera. Lytro images from Cho *et al.* [12] are calibrated using rotation estimation, center pixel estimation and bicubic interpolation as in their paper. Results are mainly compared with Cho *et al.* [12], Mitra and Veeraraghavan [17], Wanner and Goldluecke [24], Yoon *et al.* [34], and Dong *et al.* [33]. The learning-based methods, *e.g.* sparse coding [12] and deep convolutional networks [33], [34], can also be used as a post-processing procedure to improve image quality after our method. The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) [50] are used as the evaluation metrics.

1) Synthetic Images: The performance of the proposed method is first evaluated using synthetic light field images from [43]. We first downsample the sub-aperture images to 384×384 and then perform super-resolution. The results are

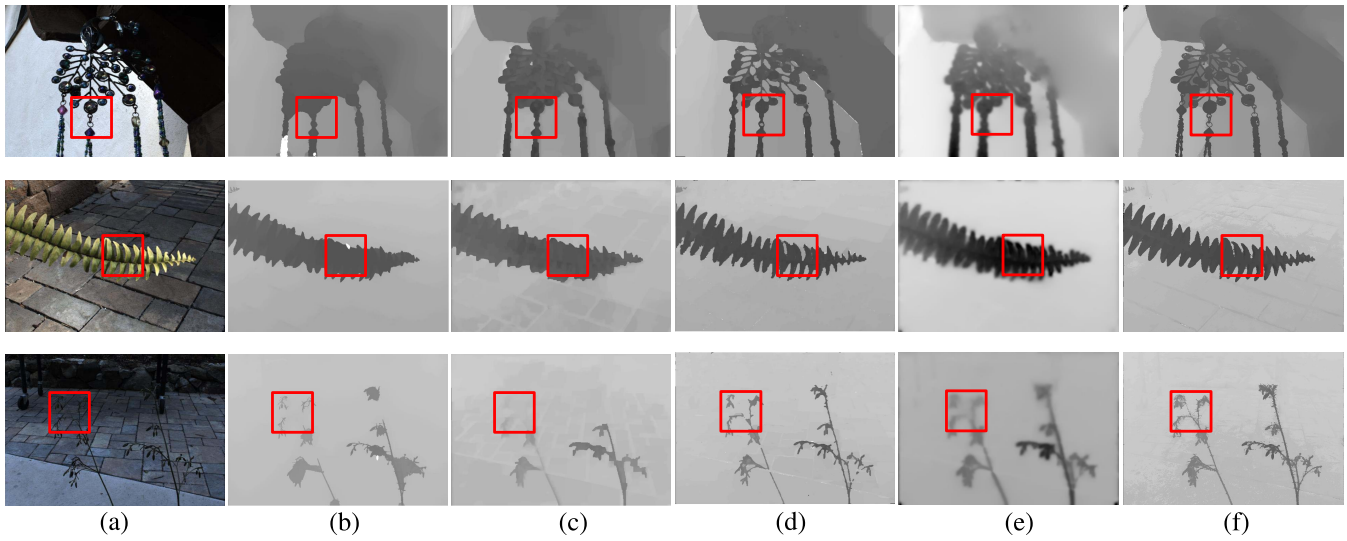


Fig. 11. The real scenes depth estimation results for Lytro Illum images. Our method performs better around occlusion boundaries compared with the other methods. For thin objects, such as the circles in the hang decoration and thin branches, our method is able to recover its accurate shape. (a) Scene image. (b) Jeon *et al.* [21]. (c) Tao *et al.* [20]. (d) Wang *et al.* [22]. (e) Tao *et al.* [48.] (f) Ours.

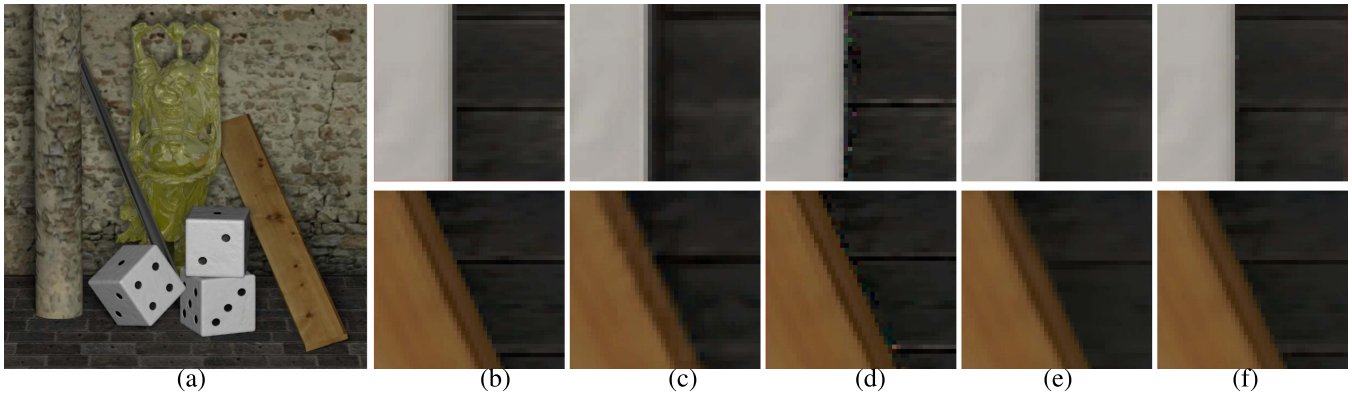


Fig. 12. The super-resolution results using different procedures of the proposed method. (a) Our super-resolved image. (b) Groun truth image. (c) Learning-based reconstruction [12]. (d) Cho *et al.* [12]. (e) Our method without occlusions handling. (f) Our method with the proposed MCM in Sec. III-C. Our reconsruction method keeps more accurate details especially along depth boudaries.

TABLE II
QUANTITATIVE EVALUATION FOR IMAGE SUPER-RESOLUTION

Image	Buddha		Mona	
	PSNR	SSIM	PSNR	SSIM
Mitra [17]	32.37	0.91	34.53	0.95
Wanner [24]	33.83	0.91	**36.84**	0.94
Cho [12]	31.92	0.90	31.77	0.91
Ours	**38.27**	**0.97**	34.44	**0.97**

compared with original 768 × 768 images. The quantitative results, illustrated in Table II, show that our method achieves the higher PSNR and SSIM scores than other methods.

The qualitative results of our method with different procedures are shown in Fig. 12. Without the proposed MCM in Sec. III-C, the occlusion boundary is recovered ambiguously, *e.g.* the blurred texture of the floor. By adding MCM, our method is the only one that is able to keep accurate edge details, including smooth boundaries and clear textures. The reason is that the proposed MCM measures occlusion probabilities and selects reliable pixels from micro-lens images for reconstruction. Among the state-of-the-art methods, the recovered images by Wanner and Goldluecke [24] have obvious artifacts along the depth boundaries because they do not consider the reconstruction along occlusion boundaries. The images calculated by the learning-based method [12] are over-smooth because they are reconstructed without structure information.

2) Lytro Images: We then evaluate real light field images provided by [12] and [49]. The sub-aperture image reconstruction results are mainly compared with those of Cho *et al.* [12], who used multiple sub-aperture images and the learning-based method for image reconstruction. As shown in Fig. 13, our reconstruction results are more accurate compared with both their multi-view reconstruction method and learning-based method. Images from Cho *et al.* [12] have obvious reconstruction errors because they do not consider the matching pixels for the reconstructed regions.

More super-resolution results are shown in Fig. 14, whose resolution is two to three times higher than original

Fig. 13. The sub-aperture image reconstruction results for Lytro images. (a) Original sub-aperture image. (b) SRCNN [33] (c) Multi-view reconstruction [12]. (d) Learning-based reconstruction [12]. (e) Our method for image reconstruction. Our results keep more accurate details than the results of multi-view reconstruction and learning-based reconstruction [12] methods.

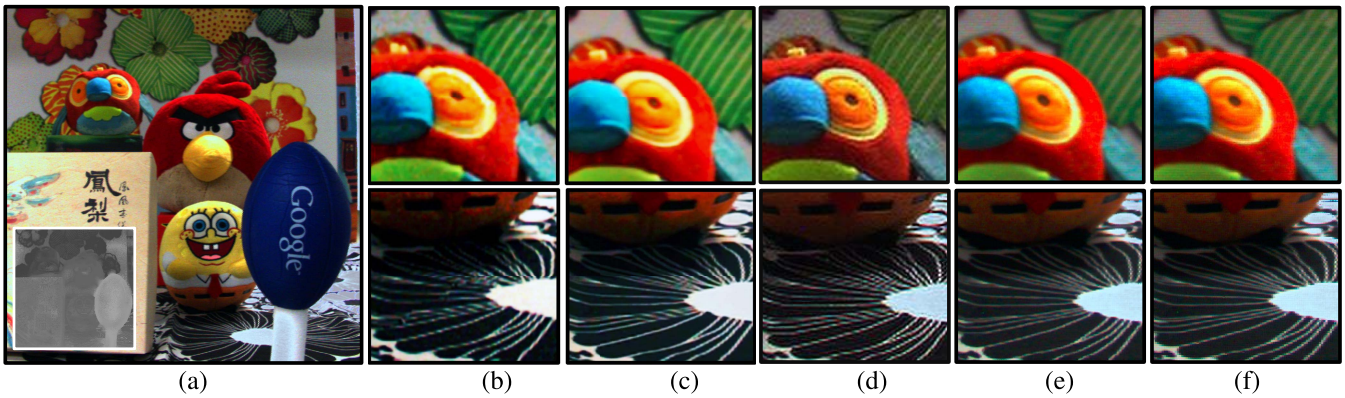


Fig. 14. The super-resolution results for Lytro images. (a) The proposed super-resolution result, which is two times larger than the original sub-aperture images and the estimated depth map. (b) Cho *et al.* [12] (c) Mitra and Veeraraghavan [17] (d) Yoon *et al.* [34] (e) and (f) The proposed super-resolution results, which is two times and three times larger than the original sub-aperture images, respectively. The super-resolution image with larger size keeps more details and shows more accurate reconstruction results compared with other state-of-art methods.

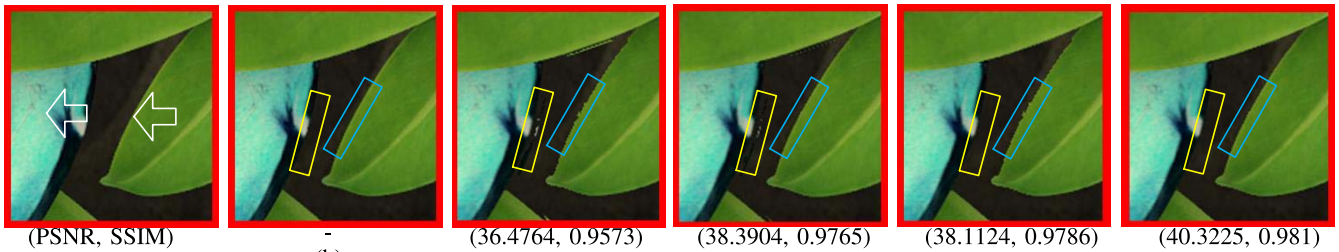


Fig. 15. The view expansion results using different occlusion assumptions, where PSNR and SSIM is illustrated in the bracket, respectively. (a) The reference view. (b) The ground truth expanded left view. The foreground objects are moving towards the left side of the image in the expanded view. (c) shows the initial view expansion result. (d) and (e) shows the view expansion results only using Eq. (8) or Eq. (7) respectively. (f) Results using both Eq. (7) and Eq. (8) in Eq. (9). The combined results in (f) show the most correct view expansion result and have the highest PSNR and SSIM.

sub-aperture images. As we can observe, the super-resolution images with higher resolution provide more details. Images from Cho *et al.* [12] have obvious reconstruction errors along the depth edges. The results of Mitra and Veeraraghavan [17] have less noise because of the denoising process. Compared with their results, our images show the sharpest edges and contain more accurate details. The aliasing artifacts, such as the undistinguishable lines on the tablecloth, are also reduced using our method.

C. View Synthesis

The view synthesis results are evaluated using synthetic and Lytro Illum camera images. We use central 7×7 sub-aperture images to calculate structure information and then synthesize novel images with 9×9 angular resolution. Results are compared with ground truth view images, where SSIM and PSNR are used for numerical comparison.

We compare our method with Kalantari *et al.* [19], who designed two sequential convolutional neural networks to

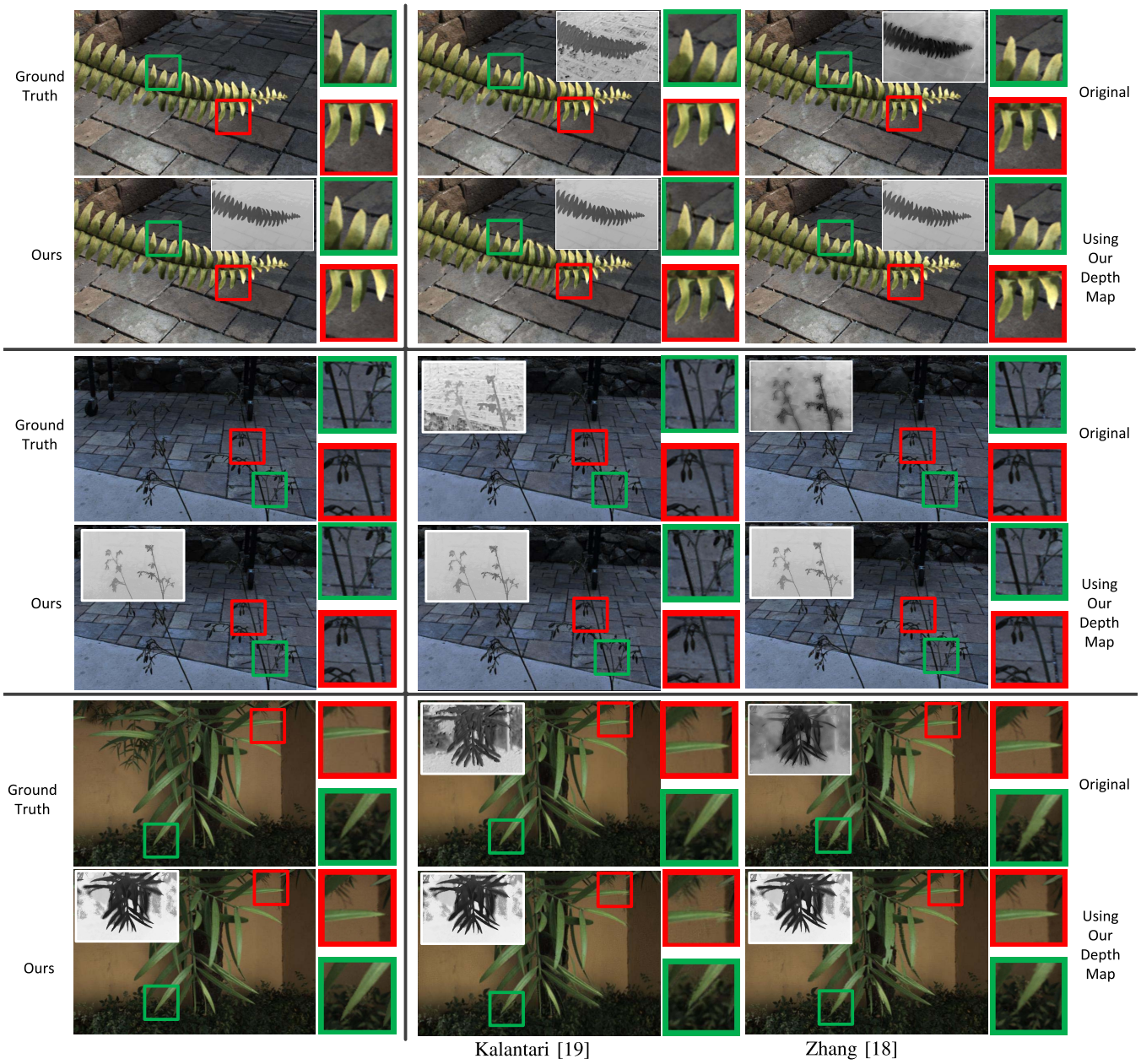


Fig. 16. The view expansion results for Lytro Illum images. The view synthesis results for Kalantari *et al.* [19] using original depth estimation methods and our depth maps are illustrated respectively. The average PSNR and SSIM of the synthesized views is (34.9140, 0.9387) and (34.183, 0.9282), respectively. Similarly, results for Zhang *et al.* [18] is (31.1466, 0.9136) and (26.1154, 0.8986) using their depth maps and our depth maps. Ours results are **(35.4215, 0.9532)**. Our method performs better around occlusion boundaries compared with the other methods, especially for thin objects and the textured background near occlusion boundaries.

synthesize new views. The comparison with Zhang *et al.* [18] is also presented, who developed a phase-based method for reconstructing 4D light field. Further comparisons using our unified depth maps are also illustrated to evaluate the view synthesis performance.

1) Occlusion Analyses: In this section, we first compare the results for newly-appeared and newly-occluded regions. As shown in Fig. 15, the synthesized images, which do not consider these regions, have expanded and discontinuous edges due to occlusions. When we only use the region detection method in Eq. (8), the newly-appeared regions cannot be recovered correctly. On the other hand, if we use Eq. (7), regions that are invisible in the reference view are filled with surrounding similar regions. Experiments show reasonable results compared with ground truth images. If we only use Eq. (7), regions that are occluded in new views are recovered with ambiguous edges. By contrast, if we use Eq. (8), the edges are recovered more accurately.

Combining the region detection method with the recovering method in Eq. (9), we fill in the newly-appeared region with the background content and fill in the newly-occluded part using the correct foreground region. The view expansion results show correct textures compared with the ground truth and achieve the highest PSNR and SSIM.

2) Comparisons: We compare our view synthesis performance with Kalantari *et al.* [19] and Zhang *et al.* [18]. The numerical results and the recovered images for Lytro Illum images are illustrated in Fig. 16.

The synthesized view images by Kalantari *et al.* [19] have obvious incorrect background textures around the depth edges, *e.g.* the edges of the floor in the first image. When our estimated depth map is used, the background textures become complete than original results. However, as they do not consider occlusions in their synthesis process, the edges of the foreground objects diffuse to other places. Zhang *et al.* [18] shows good details for foreground objects, which is mainly because depth is iteratively calculated to reduce differences between warped view images and ground truth. Their calculated depth maps change smoothly along depth boundaries. However, from the overall comparison, we found that the recovered background has serious distortions. The PSNR and SSIM of their results are also much lower than the other methods. Using our depth maps, they recover the foreground objects with incorrect shapes since they do not consider the occlusion influences.

By contrast, our method is able to recover the textured background near the occlusion boundaries, such as the complicated texture behind the plant in the first image and road textures which are occluded by the leaves in some view images. We also achieve higher PSNR and SSIM scores than other state-of-the-art methods.

D. Run-Time Analysis

On an i7-4790 3.60GHz machine implemented in MATLAB, our algorithm takes about 3 minutes on a Lytro Illum Image (7728×5368 pixels) for depth estimation, less than 1 minute for image super-resolution by a factor of 2 and 1 minute for view synthesis. Our depth estimation computational demand is comparable to [15] and less than [20], [22], [30]. The super-resolution and view synthesis is also a local method without any further optimizations, which is calculated based on the estimated depth information.

VI. Conclusion and Future Work

In this paper, we developed a new framework to recover scene information for light field images, including depth estimation, image reconstruction, super-resolution and view expansion. The micro-lens image is used as a unit in the proposed matching term and to reconstruct high-resolution scene images. The sub-aperture image is then calculated to expand micro-lens images and acquire novel expanded view images. In order to handle occlusions in the matching process, a micro-lens-based consistency metric is introduced to measure reliabilities of pixels. We verified our method on both synthetic and plenoptic light field images. Experimental results show that the proposed matching term is able to acquire accurate depth maps, especially for tiny structures and textureless regions. Furthermore, our reconstruction method and view expansion method outperforms state-of-the-art methods with more details and less aliasing artifacts.

The key component of our method is the micro-lens image for the matching term. However, the size of the corresponding region in the sub-aperture image varies based on its depth value, which leads to slightly different performances. When the disparity is zero, the micro-lens image corresponds to one point in the sub-aperture image and the proposed method is same with traditional stereo matching methods. We can try to evaluate the size of the region in the matching part and balance the differences between different depth labels, which is part of our future work. Moreover, we would like to investigate the possibility of using a deep convolutional neural network to find more accurate image patches with high resolution for image super-resolution and view expansion.

Acknowledgment

The authors would like to thank the support from HAWKEYE Group.

References

[1] 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.

[2] 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.

[3] C. Perwa and L. Wietzke. *The Next Generation of Photography*, Accessed: Aug. 28, 2017. [Online]. Available: http://www.raytrix.de

[4] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Proc. ACM 22nd Annu. Conf. Comput. Graph. Interactive Techn.*, 1995, pp. 39–46.

[5] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Apr. 2009, pp. 1–9.

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

[7] 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.

[8] M. Levoy, "Light fields and computational imaging," *Computer*, vol. 39, no. 8, pp. 46–55, Aug. 2006.

[9] 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.

[10] 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.

[11] 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.

[12] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 3280–3287.

[13] T. E. Bishop and P. Favaro, "Full-resolution depth map estimation from an aliased plenoptic light field," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2011, pp. 186–200.

[14] 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.

[15] 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.

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

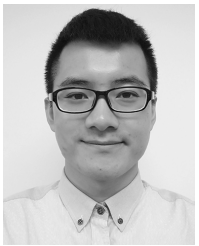
[17] K. Mitra and A. Veeraraghavan, "Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2012, pp. 22–28.

[18] Z. Zhang, Y. Liu, and Q. Dai, "Light field from micro-baseline image pair," in *Proc. CVPR*, Jun. 2015, pp. 3800–3809.

[19] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. Graph.*, vol. 35, no. 6, p. 193, 2016.
[20] 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.
[21] 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.
[22] 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.
[23] S. Heber and T. Pock, "Shape from light field meets robust PCA," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 751–767.
[24] 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.
[25] 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.
[26] 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.*, Dec. 2015, pp. 3451–3459.
[27] 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)*, vol. 1. Jun. 2017, p. 4.
[28] H. Sheng, S. Zhang, G. Zhu, and Z. Xiong, "Guided integral filter for light field stereo matching," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 852–856.
[29] H. Sheng, S. Zhang, X. Cao, Y. Fang, and Z. Xiong, "Geometric occlusion analysis in depth estimation using integral guided filter for light-field image," *IEEE Transaction Image Process. (TIP)*, vol. 26, no. 12, pp. 5758–5771, Dec. 2017.
[30] 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.
[31] Y. Bok, H.-G. Jeon, and I. S. Kweon, "Geometric calibration of micro-lens-based light-field cameras using line features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 287–300, Feb. 2014.
[32] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1027–1034.
[33] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.
[34] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. So Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2015, pp. 24–32.
[35] S. Pujades, F. Devernay, and B. Goldluecke, "Bayesian view synthesis and image-based rendering principles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3906–3913.
[36] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. ACM 23rd Annu. Conf. Comput. Graph. Interactive Techn.*, 1996, pp. 31–42.
[37] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.*, vol. 1, no. 1, pp. 7–55, 1987.
[38] 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.
[39] D. Scharstein and R. Szeliski, "Stereo matching with nonlinear diffusion," *Int. J. Comput. Vis.*, vol. 28, no. 2, pp. 155–174, 1998.
[40] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 1–14.
[41] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-, vol. filtering, for visual correspondence and beyond," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3017–3024.
[42] J. Dahl, P. C. Hansen, S. H. Jensen, and T. L. Jensen, "Algorithms and software for total variation image reconstruction via first-order methods," *Numer. Algorithms*, vol. 53, no. 1, pp. 67–92, 2010.
[43] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4D light fields," in *Proc. Vis., Modelling Visualization (VMV)*, 2013, pp. 225–226.
[44] 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.
[45] 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.
[46] *Lab Code: Multi View Stereo With TGV-L1 Regularization*, Accessed: Aug. 28, 2017. [Online]. Available: http://hci-lightfield.iwr.uni-heidelberg.de/benchmark/algorithm/mv
[47] S. Wanner and B. Goldluecke, "Reconstructing reflective and transparent surfaces from epipolar plane images," in *Proc. German Conf. Pattern Recognit.*, 2013, pp. 1–10.
[48] M. W. Tao, J.-C. Su, T.-C. Wang, J. Malik, and R. Ramamoorthi, "Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1155–1169, Jun. 2016.
[49] V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2014, pp. 1–10.
[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

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.

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.



Da Yang received the B.S. degree from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree. His research interests include computer vision and machine learning.

Jun Zhang received the B.S. degree from the Harbin Shipbuilding Engineering Institute, Harbin, China, in 1982, and the M.S. and Ph.D. degrees from the Rensselaer Polytechnic Institute in 1985 and 1988, respectively. He was admitted to the Graduate Program of the Radio Electronic Department, Tsinghua University. He joined the Faculty of Department of Electrical Engineering and Computer Science, University of Wisconsin–Milwaukee, where he is currently a Professor. His research interests include image processing and signal processing.

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.