Robust Light Field Depth Estimation for Noisy Scene with Occlusion

Williem and In Kyu Park

Dept. of Information and Communication Engineering, Inha University

22112195@inha.edu, pik@inha.ac.kr

Abstract

Light field depth estimation is an essential part of many light field applications. Numerous algorithms have been developed using various light field characteristics. However, conventional methods fail when handling noisy scene with occlusion. To remedy this problem, we present a light field depth estimation method which is more robust to occlusion and less sensitive to noise. Novel data costs using angular entropy metric and adaptive defocus response are introduced. Integration of both data costs improves the occlusion and noise invariant capability significantly. Cost volume filtering and graph cut optimization are utilized to improve the accuracy of the depth map. Experimental results confirm that the proposed method is robust and achieves high quality depth maps in various scenes. The proposed method outperforms the state-of-the-art light field depth estimation methods in qualitative and quantitative evaluation.

1. Introduction

4D light field camera has become a potential technology in image acquisition due to its rich information captured at once. It does not capture the accumulated intensity of a pixel but captures the intensity for each light direction. Commercial light field cameras, such as Lytro [16] and Raytrix [18], trigger the consumer and researcher interests on light field because of its practicability compared to the conventional light field camera arrays [26]. A light field image allows wider application to explore than a conventional 2D image. Various applications have been presented in the recent literatures, such as refocusing [17], depth estimation [4, 15, 11, 19, 20, 22, 23], saliency detection [14], matting [5], calibration [2, 6, 7], editing [10], etc.

Depth estimation from a light field image has become a challenging and active problem for the last few years. Many researchers utilize various characteristics of light field (*e.g.* epipolar plane image, angular patch, and focal stack) to develop the algorithms. However, the state-of-the-art techniques mostly fail on occlusion because it breaks the photo consistency assumption. Chen *et al.* [4] introduced a
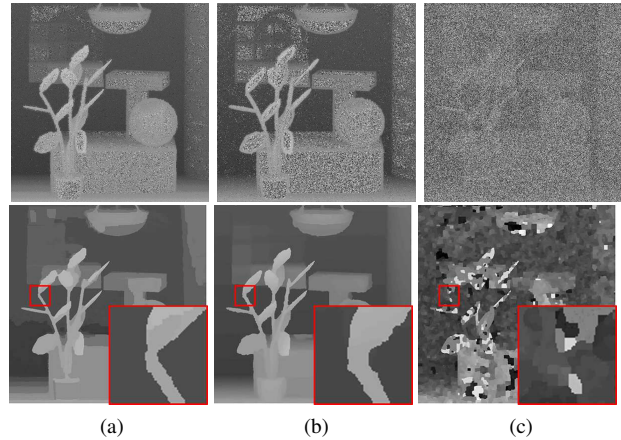

Figure 1: Comparison of disparity maps of various algorithms on a noisy light field image ($\sigma = 10$). (First row) Data cost only. (Second row) Data cost + global optimization. (a) Proposed data cost with less fattening effect; (b) Jeon's data cost [11]; (c) Chen's data cost [4].

method that is robust to occlusion but their method is sensitive to noise. Wang *et al.* [22] proposed an occlusion-aware depth estimation method but it is limited to a single occluder and highly depends on the edge detection result. It remains difficult for a depth estimation method to perform well on real data because of the occlusion and noise presence. Note that recent works mostly evaluate the results after the global optimization method is applied. Thus, the discrimination power of each data cost is not evaluated deeply since the final results depend on the individual optimization method.

In this paper, we introduce novel data costs based on our observation on the light field. Following the idea of [19], we utilize two different cues: correspondence and defocus cues. An angular entropy metric is proposed as the correspondence cue, which measures the pixel color randomness of the angular patch quantitatively. Adaptive defocus response is the modified version of the conventional defocus response [20] that is robust to occlusion. We perform cost volume filtering and graph cut for optimization. An extensive comparison between the proposed and the conventional data costs is done to measure the discrimination power of each data cost. In addition, to evaluate the proposed method

in a fair manner, we optimize the state-of-the-art data costs with the identical method. As seen in Figure 1, the proposed method achieves more accurate results in challenging scenes (with both occlusion and noise). Experimental results show that the proposed data costs significantly outperform the conventional approaches. The contribution of this paper is summarized as follows.

- Keen observation on the light field angular patch and the refocus image.

- Novel angular entropy metric and adaptive defocus response for occlusion and noise invariant light field depth estimation.

- Intensive evaluation of the existing cost functions for light field depth estimation.

2. Related Works

Depth estimation using light field images has been investigated for last a few years. Wanner and Goldluecke [23] measured the local line orientation in the epipolar plane image (EPI) to estimate the depth. They utilized the structure tensor to calculate the orientation with its reliability and introduced the variational method to optimize the depth information. However, their method was not robust because of the dependency on the angular line. Tao *et al.* [19] combined correspondence and defocus cues to obtain accurate depth. They utilized the variance in the angular patch as the correspondence data cost and the sharpness value in the generated refocus image as the defocus data cost. It was extended by Tao *et al.* [20] by adding a shading constraint as the regularization term and by modifying the original correspondence and defocus measure. Instead of variance based correspondence data cost, they employed the standard multi-view stereo data costs (sum of absolute differences). In addition, the defocus data cost was designed as the average of intensity difference between patches in the refocus and center pinhole images. Jeon *et al.* [11] proposed the method based on the phase shift theorem to deal with narrow baseline multi-view images. They utilized both the sum of absolute differences and gradient differences as the data costs. Although those methods could obtain accurate depth information, they would fail in the presence of occlusion.

Chen *et al.* [4] adopted the bilateral consistency metric on the angular patch as the data cost. It was shown that the data cost was robust to handle occlusion but it is sensitive to noise. Recently, Wang *et al.* [22] assumed that the edge orientation in angular and spatial patches were invariant. They separated the angular patch into two regions based on the edge orientation and utilized conventional correspondence and defocus data costs on each region to find the minimum cost. In addition, occlusion aware regularization term was introduced in [22]. However, their method is limited to a

single large occluder in an angular patch and the performance is affected by how well the angular patch is divided. Lin *et al.* [15] analyzed the color symmetry in light field focal stack. Their work introduced the novel infocus and consistency measure that were integrated with traditional depth estimation data costs. However, there was no extensive comparison for each data cost independently without global optimization.

Several works in multi-view stereo matching have already addressed the occlusion problem. Kolmogorov and Zabih [13] utilized the visibility constraint to model the occlusion which was optimized by graph cut. Instead of adding new term, Wei and Quan [25] handled the occlusion cost in the smoothness term. Bleyer *et al.* [1] proposed a soft segmentation method to apply the occlusion model in [25]. Those methods observed the visibility of a pixel in corresponding images to design the occlusion cost. However, it remains difficult to address the method in a huge number of views, such as light field. Kang *et al.* [12] utilized a shiftable windows to refine the data cost in occluded pixels. The method could be applied for the conventional defocus cost [20] but it might have ambiguity between occluder and occluded pixels. Vaish *et al.* [21] proposed the binned entropy data cost to reconstruct occluded surface. They measured the entropy value of a binned 3D color histogram that could lead to incorrect depth estimation, especially in smooth surfaces.

In this paper, we propose a novel depth estimation algorithm that is robust to occlusion by modelling the occlusion in the data costs directly. None of visibility constraint or edge orientation is required in the proposed data costs. In addition, the data costs are less sensitive to noise compared to the conventional ones.

3. Light Field Depth Estimation for Noisy Scene with Occlusion

3.1. Light Field Images

We observe new characteristics from light field images which are useful for designing the data cost. To measure the data cost for each depth candidate, we need to generate the angular patch for each pixel and the refocus image. Thus, each pixel in light field $L(x, y, u, v)$ is remapped to sheared light field image $L_\alpha(x, y, u, v)$ based on the depth label candidate α as follows.

$$L_\alpha(x, y, u, v) = L(x + \nabla_x(u, \alpha), y + \nabla_y(v, \alpha), u, v) \quad (1)$$

$$\nabla_x(u, \alpha) = (u - u_c)\alpha k \quad ; \quad \nabla_y(v, \alpha) = (v - v_c)\alpha k \quad (2)$$

where (x, y) and (u, v) are the spatial and angular coordinates, respectively. The center pinhole image position is denoted as (u_c, v_c). ∇_x and ∇_y are the shift value in x and y direction with the unit disparity label k. The shift value increases as the distance between light field subaperture image and the center pinhole image increases. We can
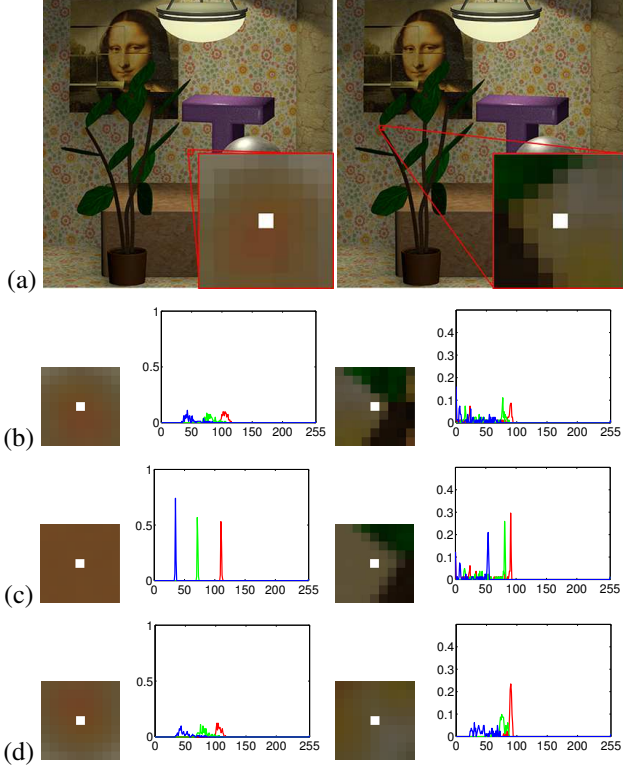
(a)

(b)

(c)

(d)

Figure 2: Angular patch analysis. (a) The center pinhole image with a spatial patch; (b) Angular patch and its histogram ($\alpha = 1$); (c) Angular patch and its histogram ($\alpha = 21$); (d) Angular patch and its histogram ($\alpha = 41$); (First column) Non-occluded pixel (Entropy costs are $3.09, 0.99, 3.15$, respectively) ; (Second column) Multi-occluded pixel (Entropy costs are $3.42, 2.34, 3.05$, respectively). Ground truth α is 21

generate an angular patch for each pixel (x, y) by extracting the pixels in the angular images from the sheared light field. Refocus image $\bar{L}_\alpha$ is generated by averaging the angular patch for all pixels.

3.2. Light Field Stereo Matching

The proposed light field depth estimation is modeled on MAP-MRF framework [3] as follows.

$$
\begin{aligned}
E = \sum_p & E_{unary}(p, \alpha(p)) + \\
& \lambda \sum_p \sum_{q \in N(p)} E_{binary}(p, q, \alpha(p), \alpha(q))
\end{aligned}
\tag{3}
$$

where $\alpha(p)$ and $N(p)$ are the depth label and the neighborhood pixels of pixel p, respectively. $E_{unary}(p, \alpha(p))$ is the data cost that measures how proper the label α of a given pixel p is. $E_{binary}(p, q, \alpha(p), \alpha(q))$ is the smoothness cost that forces the consistency between neighborhood pixels. λ is the weighting factor.

We propose two novel data costs for correspondence and defocus cues. For the correspondence response $C(p, \alpha(p))$, we measure the pixel color randomness in the angular patch by calculating the angular entropy metric. Then, we calculate the adaptive defocus response $D(p, \alpha(p))$ to obtain robust performance in the presence of occlusion. Each data cost is normalized and integrated to become a final data cost. The final data and smoothness costs are defined as follows.

$$
E_{unary}(p, \alpha(p)) = C(p, \alpha(p)) + D(p, \alpha(p)) \tag{4}
$$

$$
\begin{aligned}
E_{binary}&(p, q, \alpha(p), \alpha(q)) = \\
& \nabla I(p, q) \min(|\alpha(p) - \alpha(q)|, \tau)
\end{aligned}
\tag{5}
$$

where $\nabla I(p, q)$ is the intensity difference between pixel p and q. τ is the threshold value. Every slice in the final data cost volume is filtered with edge-preserving filter [8, 9]. Then, we perform graph cut to optimize the energy function [3]. The detail of each data cost is described in the following subsections.

3.3. Angular Entropy

Conventional correspondence data costs are designed to measure the similarity between pixels in the angular patch, but without considering the occlusion. When an occluder affects the angular patch, the photo consistency assumption is not satisfied for the pixels in the angular patch. However, a majority of pixels are still photo consistent. Therefore, we design a novel occlusion-aware correspondence data cost to capture this property by utilizing the intensity probability of the dominant pixels.

The first column in Figure 2 shows the angular patch of a pixel and its intensity histograms for several depth candidates. Without occlusion, the angular patch with the correct depth value ($\alpha = 21$) has uniform color and the intensity histogram has sharper and higher peaks as shown in Figure 2(b). Based on the observation, we measure the entropy in the angular patch, which is called *angular entropy metric*, to evaluate the randomness of photo consistency. Since light field has much more views than the conventional multi-view stereo setup, the angular patch has enough pixels to compute the entropy reliably. The angular entropy metric H is formulated as follows.

$$
H(p, \alpha) = -\sum_i h(i) \log(h(i)) \tag{6}
$$

where $h(i)$ is the probability of intensity i in the angular patch $A(p, \alpha)$. In our approach, the entropy metric is computed for each color channel independently. To integrate the costs from three channels, we cooperate two methods, max pooling C_{max} and averaging C_{avg}, which are formulated as follows.

$$
C_{max}(p, \alpha) = max(H_R(p, \alpha), H_G(p, \alpha), H_B(p, \alpha)) \tag{7}
$$

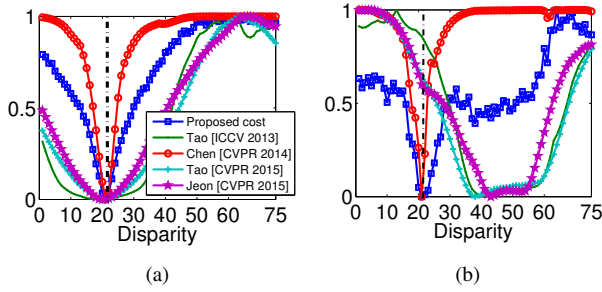(a)   (b)

Figure 3: Data cost curve comparison. (a) Non-occluded pixel in the first column of Figure 2; (b) Occluded pixel in the second column of Figure 2.

$$C_{avg}(p, \alpha) = \frac{H_R(p, \alpha) + H_G(p, \alpha) + H_B(p, \alpha)}{3} \quad (8)$$

where $\{R, G, B\}$ denotes the color channels. The max pooling C_{max} achieves better result when there is an object with a dominant color channel (*e.g.* red object has high intensity in the red channel and approximately zero intensity in the green and blue channels). Otherwise, the averaging C_{avg} performs better. To deal with various imaging conditions, the final data cost $C(p, \alpha)$ is designed as follows.

$$C(p, \alpha) = \beta C_{max}(p, \alpha) + (1 - \beta)C_{avg}(p, \alpha) \quad (9)$$

where $\beta \in [0 \sim 1]$ is the weight parameter. Figure 3 (a) shows the comparison of the data cost curves for the angular patch in the first column of Figure 2.

The angular entropy metric is also robust to the occlusion because it relies on the intensity probability of the dominant pixels. As long as the non-occluded pixels prevail in the angular patch, the metric gives low response. The second column in Figure 2 shows the angular patches when the occluders exist. Note that the proposed data cost yields the minimum response although there are multi-occluders in the angular patch.

Figure 3 (b) shows the comparison of the data cost curves of the proposed angular entropy metric and the state-of-the-art correspondence data costs. It is shown that the proposed data cost achieves the minimum cost together with Chen's bilateral data cost [4]. However, Chen's data cost is highly sensitive to noise. We compare both data costs on noisy data, which is shown in Figure 4 that Chen's data cost does not produce any meaningful disparity. On the contrary, the angular metric produces fairly promising disparity map on the noisy occluded regions.

3.4. Adaptive Defocus Response

Conventional defocus responses for the light field depth estimation are robust to noisy scene but fail on the occluded region [19, 20]. To solve the problem, we propose the *adaptive defocus response* that is robust to not only noise but

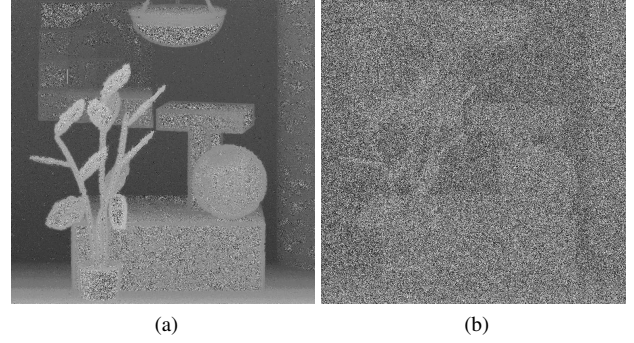(a)   (b)

Figure 4: Disparity maps of noisy light field image generated from the local data cost ($\sigma = 10$). (a) Proposed angular entropy metric; (b) Chen *et al.* [4].

also occlusion. We observe that the blurry artifact from the occluder in the refocus image causes the ambiguity in the conventional data costs. Figure 5 (a) and (c)~(f) show the spatial patches in the center pinhole image and refocus images, respectively. Conventional defocus data costs fail to produce optimal response on the patches. We compute the difference maps between the patches in the center image and refocus images to show clearer observation, as exemplified in Figure 5 (g)~(j). It is shown that the large patch in non-ground truth label ($\alpha = 35$) obtains smaller difference than the ground truth ($\alpha = 21$).

Based on the difference map observation, the idea of the adaptive defocus response is developed to find the minimum response among the neighborhood regions. Instead of measuring the response in the whole region (15×15) which is affected by the blurry artifact, we look for a subregion without blur, *i.e.* a subregion which is not affected by the occluder. To find the clear subregion, the original patch (15×15) is divided into 9 subpatches (5×5). Then, we measure the defocus response $D_c(p, \alpha)$ of each subpatch $N_c(p)$ independently as follows.

$$D_c(p, \alpha) = \frac{1}{|N_c(p)|} \sum_{q \in N_c(p)} |\bar{L}_\alpha(q) - P(q)| \quad (10)$$

where c is the index of subpatch and P is the center pinhole image. The adaptive defocus response is computed as the minimum patch response at the subpatch $c^\star$ (*i.e.* $c^\star = \min_c D_c(p, \alpha)$).

However, the initial cost still leads to the ambiguity between occluder and occluded regions as shown in Figure 5 (b). To discriminate the data cost between two cases, we introduce the additional color similarity constraint D_{col}. The constraint is the difference between the mean color of the minimum subpatch and the center pixel color, which is formulated as follows.

$$D_{col}(p, \alpha) = |\{\frac{1}{|N_{c^\star}(p)|} \sum_{q \in N_{c^\star}(p)} \bar{L}_\alpha(q)\} - P(p)| \quad (11)$$

(a)　　　　　　　　　　(b)
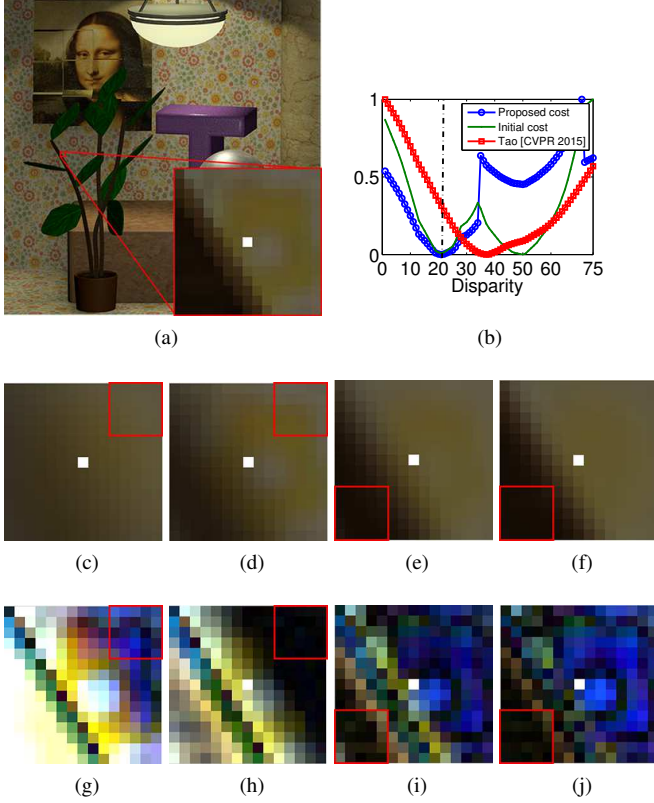

(c)　　　　(d)　　　　(e)　　　　(f)

(g)　　　　(h)　　　　(i)　　　　(j)

Figure 5: Defocus cost analysis. (a) The center pinhole image with a spatial patch; (b) Data cost curve comparison; (c)∼(f) Spatial patch from refocus image ($\alpha = 1, 21, 35, 41$); (g)∼(j) Different map of patches in (c)∼(f) We multiply the different map by 10 for better visualization. Red box shows the minimum small patch. Ground truth α is 21.

Now, the final adaptive defocus response is formed as follows.

$$D(p, \alpha) = D_{c^\star}(p, \alpha) + \gamma \, D_{col}(p, \alpha) \qquad (12)$$

where $\gamma \, (= 0.1)$ is the influence parameter of the constraint. Figure 5 (b) shows the comparison of the data cost curves of the proposed adaptive defocus response and Tao's defocus data cost [20]. It is shown that the proposed method finds the correct disparity in the occluded region.

3.5. Data Cost Integration

Both data costs are combined together to accommodate individual strength. Figure 6 shows the effect of data cost integration for clean and noisy images. Note that the angular entropy metric is robust to occlusion region and less sensitive to noise, while the adaptive defocus response is robust to noise and less sensitive to occlusion. Therefore, the combination of both data costs yields an improved data cost that is robust to both occlusion and noise. The integration leads to smaller error as evaluated in the following section.
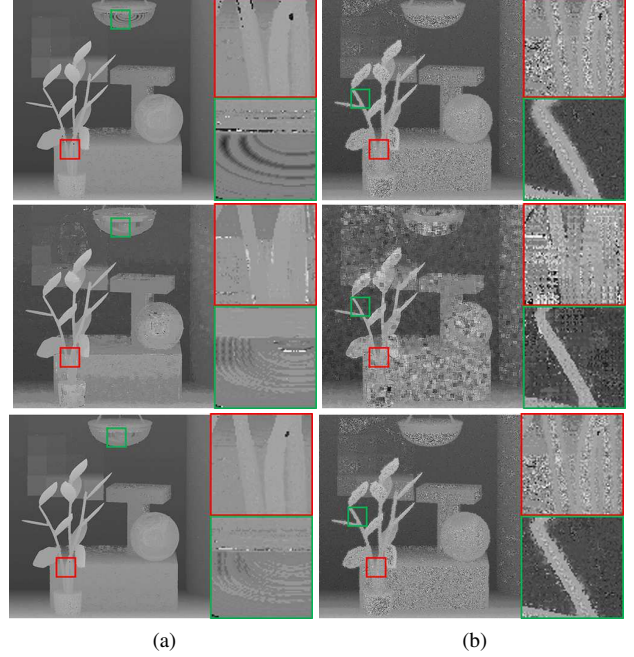

(a)　　　　　　　　　　(b)

Figure 6: Data cost integration analysis. (a) Clean images; (b) Noisy images ($\sigma = 10$); (First row) Disparity maps from angular entropy metric; (Second row) Disparity maps from adaptive defocus response; (Third row) Disparity maps from integrated data cost.

4. Experimental Results

The proposed algorithm is implemented on an Intel i7 4770 @ 3.4 GHz with 12GB RAM. We compare the performance of the proposed data costs with the recent light field depth estimation data costs. To this end, we use the code shared by the authors (Jeon *et al.* [11], Tao *et al.* [19], and Wang *et al.* [22]) and implement the other methods that are not available. For fair comparison, we first compare the depth estimation result without global optimization to identify the discriminate power of each data cost. Then, globally optimized depth is compared for a variety of challenging scenes.

4D light field benchmark is used for the synthetic dataset [24]. The real light field images are captured using the original Lytro and Lytro Illum [16]. To extract the 4D real light field image, we utilize the toolbox provided by Dansereau *et al.* [7]. We set the parameters as follows: $\lambda = 0.5$, $\beta = 0.5$, and $\tau = 10$. For the cost slice filtering, the parameter setting is $r = 5$ and $\epsilon = 0.0001$. The depth search range is 75 for all dataset.

Table 1 shows the comparison of the computational time for the data cost volume generation. We measure the runtime for each method with different image size. The proposed method has comparably fast computation. Note that our work is occlusion and noise aware depth estimation
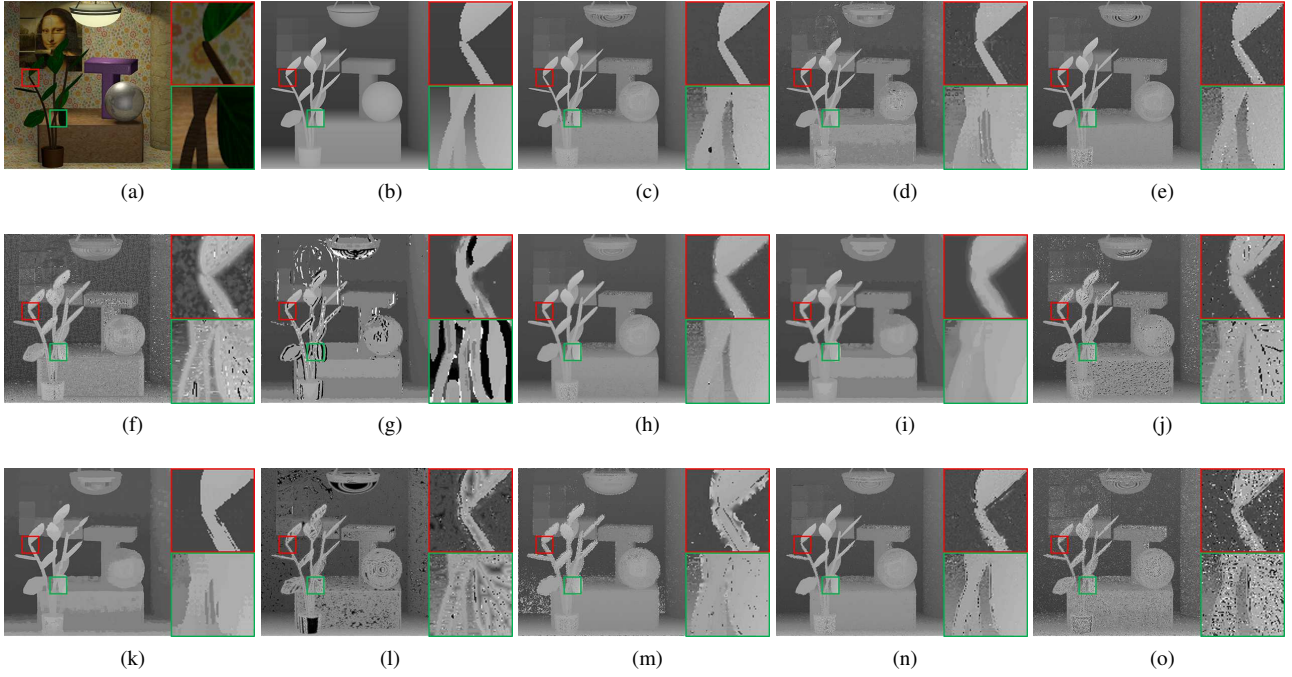
Figure 7: Comparison of the disparity maps generated from the local data cost; (a) Center pinhole image; (b) Ground truth; (c) Proposed angular entropy cost; (d) Proposed adaptive defocus cost; (e) Chen's bilateral cost [4]; (f) Tao's correspondence cost [19]; (g) Tao's defocus cost [19]; (h) Tao's correspondence cost [20]; (i) Tao's defocus cost [20]; (j) Jeon's data cost [11]; (k) Kang's shiftable window [12]; (l) Vaish's binned entropy cost [21]; (m) Lin's defocus cost [15]; (n) Wang's correspondence cost [22]; (o) Wang's defocus cost [22].

Table 1: Computational time for the cost volume generation (in seconds).

Data Type	Synthetic Data	Original Lytro	Lytro Illum
	$(9 \times 9 \times 768 \times 768)$	$(9 \times 9 \times 379 \times 379)$	$(11 \times 11 \times 625 \times 434)$
Chen *et al.* [4]	892	260	612
Jeon *et al.* [11]	2,590	689	1,406
Kang *et al.* [12]	102	31	89
Lin *et al.* [15]	167	40	93
Tao *et al.* [19]	878	110	246
Tao *et al.* [20]	252	68	149
Vaish *et al.* [21]	598	173	386
Wang *et al.* [22]	256	76	183
Proposed method	528	115	207

Table 2: The mean squared error of various dataset.

Data Type	*Buddha*	Noisy *Buddha*	*StillLife*	Noisy *StillLife*
Chen's bilateral cost [4]	0.0129	1.0654	0.3918	4.9483
Jeon's data cost [11]	0.0730	0.2575	0.0759	0.2264
Kang's shiftable window [12]	0.0139	0.1944	0.0787	0.1818
Lin's defocus cost [15]	0.2750	0.3374	2.4554	2.4624
Tao's correspondence cost [19]	0.1198	0.1657	0.2314	0.2490
Tao's defocus cost [19]	0.2267	0.2220	0.5728	0.5394
Tao's correspondence cost [20]	0.0186	0.2908	0.0473	0.2422
Tao's defocus cost [20]	0.0189	**0.0713**	0.0538	0.0711
Vaish's binned entropy cost [21]	0.1646	0.1500	0.0377	0.0980
Wang's correspondence cost [22]	0.0304	0.8980	0.0624	5.3913
Wang's defocus cost [22]	0.2085	0.7173	0.9148	2.3622
Proposed correspondence cost	0.0058	0.1174	**0.0176**	0.0608
Proposed defocus cost	0.0266	0.2500	0.0634	0.1961
Proposed integrated cost	**0.0047**	0.0989	0.0206	**0.0532**
Chen's cost + optimization [4]	0.0041	0.1646	**0.0126**	0.1386
Jeon's cost + optimization [11]	0.0149	0.0250	0.0283	0.0432
Tao's cost + optimization [19]	0.0270	0.0300	0.0572	0.0593
Tao's cost + optimization [20]	0.0154	0.0282	0.0491	0.0587
Proposed cost + optimization	**0.0036**	**0.0170**	0.0142	**0.0251**

while others are general or only occlusion aware. Furthermore, the proposed method, Tao *et al.* [19], Tao *et al.* [20], and Wang *et al.* [22] have two data costs in each method.

4.1. Synthetic Clean and Noisy Data

Figure 7 shows the non-optimized depth comparison for the synthetic dataset from Wanner *et al.* [24]. Since the proposed data costs consider the occlusion, it is shown that they outperform the conventional data costs, yielding less fattening effect in the occluded region (*i.e.* leaves or branches). Similar to the proposed method, Wang *et al.* [22] also model the occlusion in their data cost, but their method highly de-

pends on the edge detection result.

Furthermore, we also evaluate the optimized results of the selected methods ([4],[11],[19],[20]) using clean and noisy light field data, as shown in Figure 8. The noisy image is generated by adding additive Gaussian noise with standard deviation $\sigma = 10$. For fair comparison, we per-
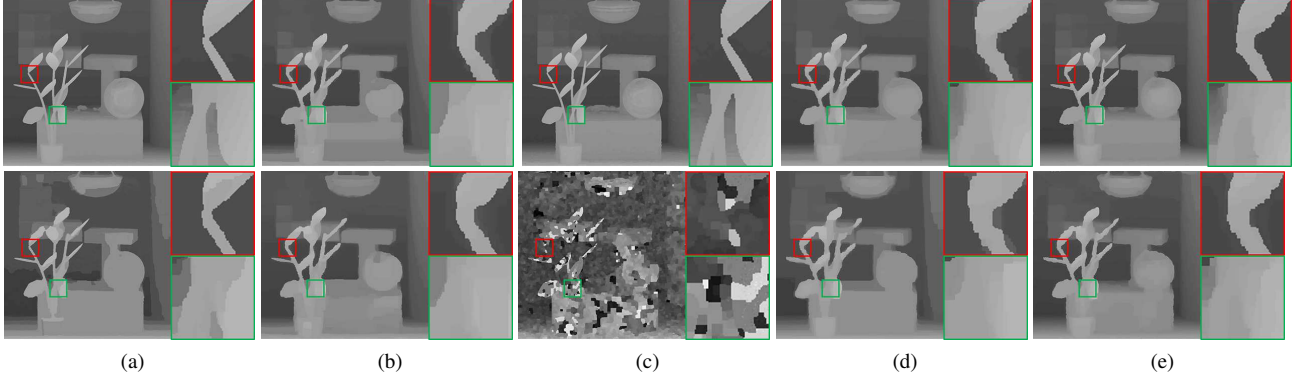
Figure 8: Comparison of the optimized disparity maps (synthetic data); (a) Proposed method; (b) Tao's method [19]; (c) Chen's method [4]; (d) Tao's method [20]; (e) Jeon's method [11]; (First row) Clear light field image; (Second row) Noisy light field image ($\sigma = 10$).

Table 3: The mean squared error of results from *Mona* dataset with different noise levels.

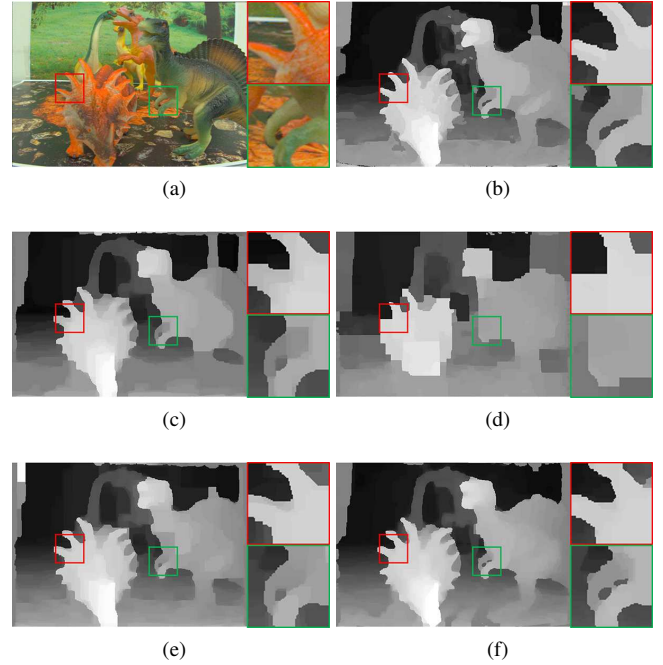| Data Type	$\sigma = 0$	$\sigma = 5$	$\sigma = 10$	$\sigma = 15$
Chen's bilateral cost [4]	0.0233	0.6834	0.8754	0.9373
Jeon's data cost [11]	0.0801	0.1588	0.2665	0.3791
Kang's shiftable window cost [12]	0.0259	0.0799	0.1772	0.2610
Lin's defocus cost [15]	0.0626	0.0851	0.1162	0.1453
Tao's correspondence cost [19]	0.0601	0.0772	0.1191	0.1755
Tao's defocus cost [19]	0.1350	0.1246	0.1252	0.1395
Tao's correspondence cost [20]	0.0183	0.1300	0.2752	0.4025
Tao's defocus cost [20]	0.0244	0.0409	**0.0806**	**0.0350**
Vaish's binned entropy cost [21]	0.0928	0.1170	0.1480	0.2115
Wang's correspondence cost [22]	0.0207	0.3072	0.5914	0.8130
Wang's defocus cost [22]	0.1301	0.4357	0.5809	0.6809
Proposed correspondence cost	0.0118	0.0381	0.1152	0.2043
Proposed defocus cost	0.0217	0.0954	0.1962	0.2788
Proposed integrated cost	**0.0081**	**0.0337**	0.1011	0.1792
Chen's cost + optimization [4]	0.0052	0.0308	0.1466	0.8547
Jeon's cost + optimization [11]	0.0122	0.0179	0.0224	0.0274
Tao's cost + optimization [19]	0.0238	0.0251	0.0227	0.0226
Tao's cost + optimization [20]	0.0203	0.0245	0.0301	0.0350
Proposed cost + optimization	**0.0045**	**0.0072**	**0.0125**	**0.0185**

Figure 9: Comparison of the optimized disparity maps (real data); (a) Center pinhole image; (b) Proposed method; (c) Tao's method [19]; (d) Chen's method [4]; (e) Tao's method [20]; (f) Jeon's method [11].

form the same optimization technique for all methods. As the initial data costs fail to produce the minimum cost on the occluded region, the conventional methods produce the artifacts around the object boundary even after global optimization. Chen *et al.* [4] achieves comparable performance on the clean data. However, it produces significant error on the noisy data. It is shown that the proposed method achieves stable performance in both environments.

Next, mean squared error is measured to evaluate the computed depth accuracy. Table 2 and Table 3 show the comparison of the mean squared error for various data with multiple noise levels. On the clean data, the proposed method and Chen's method [4] achieve comparable performance. However, Chen's method fails on the noisy data

while the proposed method obtains the minimum error for most cases. The proposed method obtains the best performance among all the conventional data costs.

4.2. Noisy Real Data

Light field image captured by commercial light field camera contains noise due to its small sensor size. We evaluate the proposed method with several scenes captured inside a room with limited lighting, which degrades the signal-to-noise ratio. The typical light level in the room
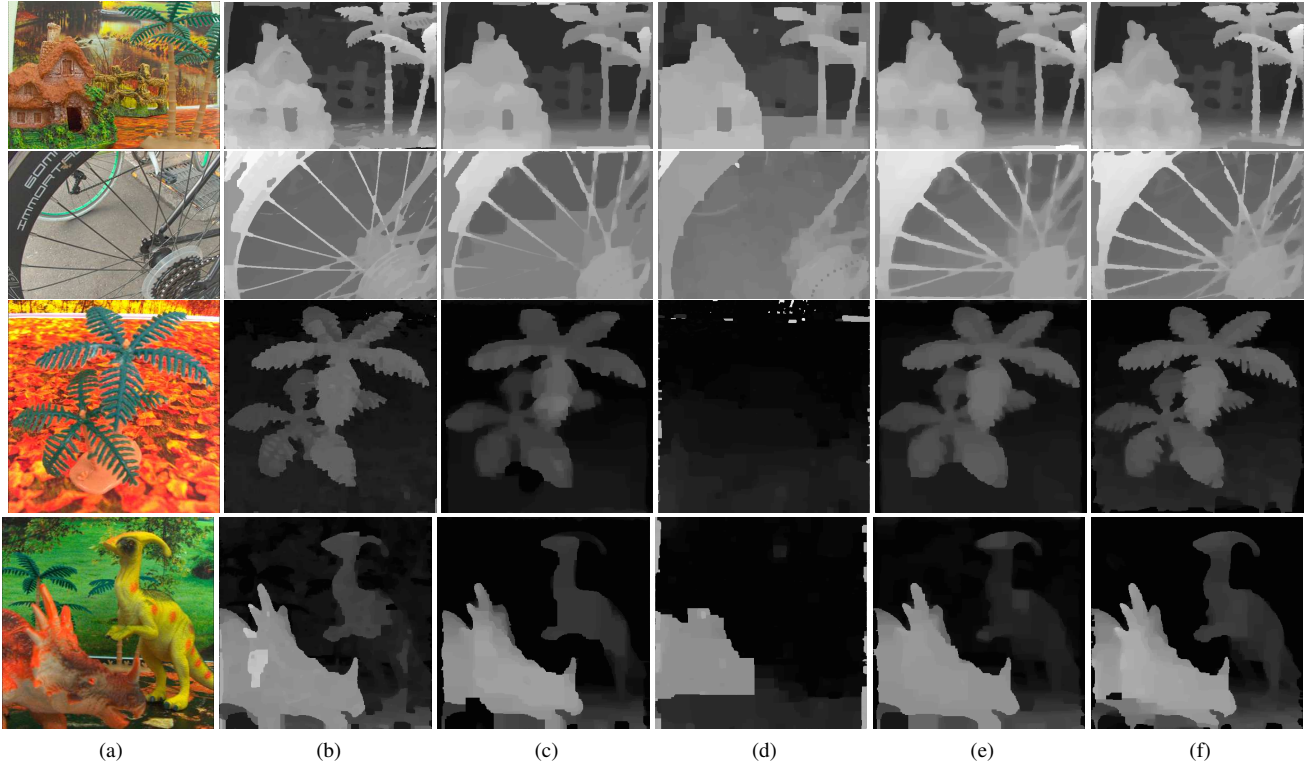
Figure 10: Additional comparison of the optimized disparity maps (real data);(a) Center pinhole image; (b) Proposed method; (c) Tao's method [19]; (d) Chen's method [4]; (e) Tao's method [20]; (f) Jeon's method [11]. The first and second rows are captured by Lytro Illum camera while the others are captured by original Lytro camera.

(≈ 400 lux) is much lower than the light level of outdoor under daylight (≈ 10000 lux). Figure 9 and Figure 10 show the disparity maps generated from the real light field images. Conventional approaches [4, 11, 19, 20] exhibit blurry or fattening effect around the object boundaries, as shown in Figure 9. On the other hand, the proposed method shows sharp and unfattened boundary. Note that only the proposed method can estimate the correct shape of the house entrance and the spokes of the wheel as shown in Figure 10.

4.3. Limitation and Future Work

The angular entropy metric becomes less reliable when the noise/occluder is more dominant than the clean/non-occluded pixels in the angular patch. Thus, it performs better when there are lots of subaperture images. As an example, it performs better on the Lytro Illum images than the images captured by original Lytro camera, as shown in Figure 10. This problem might be addressed by using spatial neighborhood information to increase the probability of the dominant pixels or capturing more angular images. In addition, it is also useful to find the reliability measure for the entropy data cost. To extend the current work, we also plan to perform exhaustive comparison and informative benchmarking on the state-of-the-art data costs for light field depth estimation.

5. Conclusion

In this paper, we proposed an occlusion and noise aware light field depth estimation framework. Two novel data costs were proposed to obtain robust performance in the occluded region. Angular entropy metric was introduced to measure the pixel color randomness in the angular patch. In addition, adaptive defocus response was determined to gain robust performance against occlusion. Both data costs were integrated in the MRF framework and further optimized using graph cut. Experimental results showed that the proposed method significantly outperformed the conventional approaches in both occluded and noisy scenes.

Acknowledgement

References

[1] M. Bleyer, C. Rother, and P. Kohli. Surface stereo with soft segmentation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 1570–1577, 2010. 2

[2] Y. Bok, H.-G. Jeon, and I. S. Kweon. Geometric calibration of micro-lens-based light-field cameras using line features. In *Proc. of European Conference on Computer Vision*, pages 47–61, 2014. 1

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov. 2001. 3

[4] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu. Light field stereo matching using bilateral statistics of surface cameras. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 1518–1525, 2014. 1, 2, 4, 6, 7, 8

[5] D. Cho, S. Kim, and Y.-W. Tai. Consistent matting for light field images. In *Proc. of European Conference on Computer Vision*, pages 90–104, 2014. 1

[6] 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. of IEEE International Conference on Computer Vision*, pages 3280–3287, 2013. 1

[7] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 1027–1034, 2013. 1, 5

[8] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, June 2013. 3

[9] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(2):504–511, Feb. 2013. 3

[10] A. Jarabo, B. Masia, A. Bousseau, F. Pellacini, and D. Gutierrez. How do people edit light fields? *ACM Trans. on Graphics*, 33(4), July 2014. 1

[11] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon. Accurate depth map estimation from a lenslet light field camera. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 1547–1555, 2015. 1, 2, 5, 6, 7, 8

[12] S. B. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2001. 2, 6, 7

[13] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Proc. of European Conference on Computer Vision*, pages 82–96, 2002. 2

[14] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 2806–2813, 2014. 1

[15] H. Lin, C. Chen, S. B. Kang, and J. Yu. Depth recovery from light field using focal stack symmetry. In *Proc. of IEEE International Conference on Computer Vision*, 2015. 1, 2, 6, 7

[16] Lytro. The Lytro camera, 2014. 1, 5

[17] R. Ng. Fourier slice photography. *ACM Trans. on Graphics*, 24(3):735–744, July 2005. 1

[18] Raytrix. 3D light field camera technology, 2013. 1

[19] M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *Proc. of IEEE International Conference on Computer Vision*, pages 673–680, 2013. 1, 2, 4, 5, 6, 7, 8

[20] M. Tao, P. P. Srinivasan, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi. Depth from shading, defocus, and correspondence using light-field angular coherence. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 1940–1948, 2015. 1, 2, 4, 5, 6, 7, 8

[21] 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. of IEEE Computer Vision and Pattern Recognition*, 2006. 2, 6, 7

[22] T. C. Wang, A. A. Efros, and R. Ramamoorthi. Occlusion-aware depth estimation using light-field cameras. In *Proc. of IEEE International Conference on Computer Vision*, 2015. 1, 2, 5, 6, 7

[23] S. Wanner and B. Goldluecke. Globally consistent depth labelling of 4D lightfields. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 41–48, 2012. 1, 2

[24] S. Wanner, S. Meister, and B. Goldluecke. Datasets and benchmarks for densely sampled 4D light fields. In *Proc. of Vision, Modeling & Visualization*, pages 225–226, 2013. 5, 6

[25] Y. Wei and L. Quan. Asymmetrical occlusion handling using graph cut for multi-view stereo. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 902–909, 2005. 2

[26] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Artunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. on Graphics*, 24(3):765–779, July 2005. 1