Robust Light Field Depth Estimation using Occlusion-Noise Aware Data Costs

Williem, Student Member, IEEE, In Kyu Park, Senior Member, IEEE, and Kyoung Mu Lee, Member, IEEE,

Abstract—Depth estimation is essential in many light field applications. Numerous algorithms have been developed using a range of light field properties. However, conventional data costs fail when handling noisy scenes in which occlusion is present. To address this problem, we introduce a light field depth estimation method that is more robust against occlusion and less sensitive to noise. Two novel data costs are proposed, which are measured using the angular patch and refocus image, respectively. The constrained angular entropy cost (CAE) reduces the effects of the dominant occluder and noise in the angular patch, resulting in a low cost. The constrained adaptive defocus cost (CAD) provides a low cost in the occlusion region, while also maintaining robustness against noise. Integrating the two data costs is shown to significantly improve the occlusion and noise invariant capability. Cost volume filtering and graph cut optimization are applied to improve the accuracy of the depth map. Our experimental results confirm the robustness of the proposed method and demonstrate its ability to produce high-quality depth maps from a range of scenes. The proposed method outperforms other state-of-the-art light field depth estimation methods in both qualitative and quantitative evaluations.

Index Terms—light field, depth estimation, occlusion-aware, noise-aware, data cost, constrained angular entropy, constrained adaptive defocus

1 INTRODUCTION

The 4D light field camera is a promising potential technology in image acquisition owing to its ability to capture rich information. Unlike conventional technology, it does not capture the accumulated intensity of a pixel but rather captures the intensity of each direction of light. Commercial light field cameras, such as those by Lytro [1] and Raytrix [2], are attracting interest from both consumers and light field researchers because of their superiority to conventional light field camera arrays [3]. A light field image has wider applications than a conventional 2D image, including refocusing [4], saliency detection [5], matting [6], and editing [7]. Among the potential applications, light field depth estimation is the most active area of current research [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23].

To develop the light field depth estimation algorithms, researchers have used a range of properties of the light field, such as the epipolar plane image (EPI), angular patch, and refocus image. However, even state-of-the-art techniques have failed to successfully address occlusion, as it violates two key assumptions: photo consistency (correspondence cues) and focus area (defocus cues). Chen et al. [8] introduced a method that is robust to occlusion, but it is sensitive to noise. Wang et al. [19] proposed an occlusion-aware depth estimation method, but it is limited to a single occluder and is highly dependent on edge detection and optimization. It remains challenging to apply a depth estimation method to real data because of the presence of occlusion and noise. Most recent works have evaluated the results after applying the global optimization method. This means that the discrimination power of each data cost has not been properly evaluated, since the final results depend on the optimization method used.

In this paper, we introduce two novel data costs that are robust against both occlusion and noise. To achieve this, we utilize two different cues: correspondence and defocus. The preliminary data costs (angular entropy and adaptive defocus costs) have been presented in [21]. The refined data costs proposed in this paper are the constrained angular entropy cost (CAE) and the constrained adaptive defocus cost (CAD). The intuition for each data cost is that neighboring pixels should have a similar value as that of the center pixel. Instead of utilizing all pixels in the angular patch, CAE weights each pixel based on color similarity. Thus, occluder pixels in the angular patch make a smaller contribution in the entropy calculation. As the occluders tend to produce blurry artifacts in the refocus image, we divide the original refocus image patch into a set of subpatches and measure the conventional defocus cost of each subpatch; then, we add a color similarity constraint for each subpatch cost. CAD is then set as the minimum constrained cost of all the subpatches, allowing the area without blurry artifacts to be selected.

Next, we conduct extensive comparisons between the proposed and conventional data costs to compare their discriminative power. To ensure that the evaluation is fair, an identical method is used to optimize the state-of-the-art data costs. We also evaluate different optimization methods for each data cost. The evaluation is performed using a well-known light field dataset [24] and light field images captured by a Lytro Illum light field digital camera. For quantitative evaluation, the mean square error (MSE) and
bad pixel percentage (BP) are measured using all pixels within the image and pixels in the occlusion regions. Our experimental results demonstrate that the data costs of the proposed method are significantly better than those from conventional approaches, especially for the occlusion region and for noisy scenes.

The main contributions of this paper are as follows.
- Precise observation of the light field angular patch and the refocus image.
- Novel constrained angular entropy and constrained adaptive defocus costs for occlusion and noise invariant light field depth estimation.
- Intensive evaluation of existing cost functions for light field depth estimation.

The rest of this paper is organized as follows. In Section 2, we introduce the related literature. Section 3 describes the properties of the light field images, light field stereo matching, and proposed data costs. The experimental results are presented in Section 4. Section 5 discusses our conclusions.

## 2 Related Works

Light field depth estimation has been an active research topic over the last few years. While range of methods have been presented, the focus of this paper is on the data cost design of light field depth estimation. More specifically, we evaluate the robustness of each data cost against occlusion and noise. In contrast, previous studies have investigated a range of depth estimation approaches, including EPI-based, angular patch-based, and refocus image-based techniques.

The application of EPI analysis to the extraction of depth information was introduced by Bolles et al. [25], who used it to detect edges, peaks, and troughs and to extract features and their depth with a focus on sparse depth information in the image. Matoušek et al. [13] extracted EPI lines with similar intensities using dynamic programming. The line candidates were generated using the first and last rows of an EPI, and the variance value for each line candidate was computed as the data cost. In [26], Criminisi et al. proposed an iterative method for extracting the EPI tube, which is a set of EPI strips with the same depth. For each pixel, the EPI strip was then detected by measuring the vertical variance of the rectified EPI. Wanner and Goldluecke [20] measured the local line orientation in the EPI to estimate the depth information. They used the structure tensor method to calculate the orientation and assess its reliability and introduced the variational method for optimizing the depth information. However, their method is sensitive to noise and occlusion.

A fine-to-coarse framework was introduced by Kim et al. [27] for the reconstruction of depth images from a high spatio-angular resolution light field. They used modified Parzen window estimation with an Epanechnikov kernel to measure the distance between pixels and the mean value for each EPI line candidate. Tošić and Berkner [17] developed a light field scale-depth space and detected the local extrema in that space. These local extrema represent the depth and width information of an object in the EPI. However, the EPI data costs used in [13], [17], [20], [25], [26], [27] (i.e., variance-based cost, structure tensor, etc.) do not take into account occlusion and noise, and therefore cannot be applied to real light field data captured by commercial light field cameras, such as those by Lytro [1]. To deal with occlusion and noise when using EPI information, Zhang et al. [22] recently introduced a novel spinning parallelogram operator by measuring the weighted histogram distance between two regions next to the EPI line. However, their method requires further optimization to produce accurate results. They also applied different optimization methods and parameters to synthetic data and real data. In this paper, we took Zhang’s method [22] to represent the state-of-the-art EPI-based approaches.

In contrast with EPI-based methods, angular patch-based and refocus image-based methods are often integrated [9], [14], [15]. For example, Tao et al. [14] combined correspondence and defocus cues to obtain accurate depth, using 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. This approach was extended by Tao et al. [15] by adding a shading constraint as the regularization term and by modifying the original correspondence and defocus measure. In place of a variance-based correspondence data cost, they used the standard multiview stereo data costs, calculated from the sum of absolute differences. In addition, they derived the defocus data cost as the average intensity difference between patches in the refocus and center pinhole images. Lin et al. [9] analyzed the color symmetry in the light field focal stack, and introduced novel infocus and consistency measures that were integrated with traditional multiview data costs. However, no detailed comparisons were made for each data cost independently, without applying global optimization. Jeon et al. [12] proposed a method for handling narrow baseline multiview images based on the phase shift theorem, and using the sum of absolute and gradient differences as the data costs. Although these methods can provide accurate depth information, they fail in the presence of occlusion.

Using the angular patch, Vaish et al. [18] proposed applying the binned entropy data cost to reconstruct an occluded surface. They measured the entropy value of a binned 3D color histogram. This may lead to quantization errors in the entropy measurement and incorrect depth estimation, especially on smooth surfaces with small color changes. To resolve the occlusion problem, Chen et al. [8] adopted the bilateral consistency metric on the angular patch as the data cost, and showed that the data cost was robust against handling occlusion but sensitive to noise. Wang et al. [19] assumed that the edge orientation in angular and spatial patches was invariant. They separated the angular patch into two regions based on the edge orientation, and used conventional correspondence and defocus data costs in each region to identify the minimum cost. An occlusion-aware regularization term was also introduced. However, their method is limited to a single large occluder in an angular patch, and the performance is affected by how accurately the angular patch is divided.

Several studies have used multiview stereo matching to address the occlusion problem. Kolmogorov and Zabih [28] utilized the visibility constraint to model the occlusion which was then optimized by a graph cut method. Instead
of adding a new term, Wei and Quan [29] handled the occlusion cost within the smoothness term. Bleyer et al. [30] proposed the application of a soft segmentation method to the occlusion model in [29]. These methods use the visibility of a pixel in the corresponding images to derive the occlusion cost, but it remains challenging to apply such methods when a large number of views are present, such as in a light field. Kang et al. [31] used a shiftable window to refine the data cost in occluded pixels and showed that the method could be applied to the conventional defocus cost [15]. However, there was ambiguity between the occluder and occluded pixels.

Heber et al. [10] generated virtual multiview images using active wavefront sampling (AWS), and solved the depth estimation problem using a general variational model with the sum of absolute differences as the data cost. Yu et al. [23] developed 3D line matching between subaperture images and used it to estimate the depth information. Heber and Pock [11] extended their work in [10] and improved its stability and accuracy by introducing a modified global matching cost using a low-rank model. Tao et al. [16] introduced an iterative depth estimation and specular separation framework using the depth estimation method in [14] to simultaneously compute the free specular image and depth information.

The goal of the paper is to introduce novel occlusion and noise-aware data costs and to evaluate the data costs of different approaches to light field depth estimation. In contrast with the conventional approaches, the proposed data costs do not require additional information, for example, from edge detection. Instead, to reduce the effect of the occluders, we use a constraint in the angular patch for the correspondence costs and a refocus image patch for the defocus costs.

3 Light Field Depth Estimation for a Noisy Scene with Occlusion

3.1 Light Field Image

In general, three properties from the light field image are used to measure the data cost: EPI, the angular patch, and the refocus image. In this paper, we utilize $L(x, y, u, v)$ light field parameterization as illustrated in Fig. 1. We use the angular patch and refocus image to estimate the data cost for each depth label candidate. To generate the angular patch, each pixel in the light field $L(x, y, u, v)$ is remapped to a sheared light field image $L_\alpha(x, y, u, v)$ based on the depth label candidate $\alpha$ as follows.

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

(1)

$$\nabla_x(u, \alpha) = (u - u_c)(\alpha - \alpha_c)k$$

(2)

$$\nabla_y(v, \alpha) = (v - v_c)(\alpha - \alpha_c)k$$

(3)

where the center pinhole image position is denoted as $(u_c, v_c)$. $\nabla_x$ and $\nabla_y$ are the shift values in the $x$ and $y$ directions with the unit disparity label $k$, respectively. $\alpha_c$ represents the depth label $\alpha$ with zero disparity. The shift value increases as the distance between the light field subaperture image and the center pinhole image increases. Without loss of generality, in this paper, the depth/disparity map does not denote the real depth/disparity value but rather represents the depth label map $\alpha$. An angular patch can be generated by extracting the pixels in the angular images from the sheared light field, as follows.

$$A_\alpha^p(u, v) = L_\alpha(x, y, u, v)$$

(4)

where $A_\alpha^p$ is the angular patch on pixel $p = (x, y)$ with depth label $\alpha$. The refocus image $R_\alpha$ is generated by averaging the angular patches for all pixels, defined as follows.

$$R_\alpha(p) = \frac{1}{|A|} \sum_{u, v} A_\alpha^p(u, v)$$

(5)

where $|A|$ is the number of pixels in the angular patch.

3.2 Light Field Stereo Matching

In this paper, light field depth estimation is modeled on the MAP-MRF framework [32] as follows.

$$E = \sum_p E_{unary}(p, \alpha(p)) + \lambda \sum_p \sum_{q \in N(p)} E_{binary}(p, q, \alpha(p), \alpha(q))$$

(6)

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

We propose two novel data costs for correspondence and defocus cues. For the correspondence data cost $C(p, \alpha(p))$, we measure the pixel color randomness in the angular patch by calculating the constrained angular entropy metric. Then, we calculate the constrained adaptive defocus response $D(p, \alpha(p))$ to obtain robust performance in the presence of occlusion. Each data cost is normalized and integrated to
Fig. 2: Angular patch analysis. (a) The center pinhole image with a spatial patch; (b) Angular patch and its histogram \((\alpha = 8)\); (c) Angular patch and its histogram \((\alpha = 28)\); (d) Angular patch and its histogram \((\alpha = 48)\); (First column) Non-occluded pixel; (Second column) Multi-occluded pixel. Ground truth \(\alpha\) is 8. The contrast of each patch is enhanced for better visualization.

the final data cost. The final data and smoothness costs are defined as follows.

\[
E_{\text{unary}}(p, \alpha(p)) = \beta C(p, \alpha(p)) + (1 - \beta) D(p, \alpha(p))
\]

\[
E_{\text{binary}}(p,q,\alpha(p),\alpha(q)) = \nabla I(p,q) \min(|\alpha(p) - \alpha(q)|, \tau)
\]

where \(\nabla I(p,q)\) is the intensity difference between pixel \(p\) and \(q\), \(\tau\) is the threshold value, and \(\beta\) is the weighting factor for the correspondence data cost. Every slice in the final data

Fig. 3: Data cost curve analysis for angular patch in (a) Fig. 2 (first column); (b) Fig. 2 (second column); (c) Fig. 4. Red line is the ground truth depth label.

cost volume is filtered using the edge-preserving filter [33], [34]. We then perform a graph cut to optimize the energy function [32]. The details of each data cost are discussed in the following subsections.

3.3 Constrained Angular Entropy Cost (CAE)

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 broken for the pixels in the angular patch, but the majority of the pixels are still photo consistent. We therefore design a novel, occlusion-aware correspondence data cost to capture this property from the intensity probability of the dominant pixels.

The first column in Fig. 2 shows the angular patches of a pixel and its intensity histograms for several depth candidates. In the absence of occlusion, the angular patch with the correct depth value \((\alpha = 8)\) has a uniform color and the intensity histogram has sharper and higher peaks, as shown in Fig. 2 (b). Based on this observation, we measure the entropy in the angular patch, which is called the angular entropy cost (AE), and use this to evaluate the randomness in the photo consistency. Since a light field comprises many more views than a conventional multiview stereo setup, the angular patch contains sufficient pixels to allow the entropy to be reliably computed. The angular entropy cost \(H\) is formulated as follows.

\[
H(p, \alpha) = -\sum_i h(i) \log(h(i))
\]  

where \(h(i)\) is the probability of intensity \(i\) in the angular patch \(A_p^\alpha\). Unlike [18], the entropy cost is computed for each color channel independently without binning the histogram in our approach. AE is also robust against 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 cost gives a low response. The second column in Fig. 2 shows the angular patches when occluders are present. Note that the proposed data cost yields the minimum response although there are multi-occluders in the angular patch. The data cost curve of each angular patch is shown in Fig. 3 (a) and (b). The preliminary results of the angular entropy cost are available in [21].

However, AE is less reliable when the occluder or noise becomes more dominant than the non-occluded or clean pixels in the angular patch, as shown in Fig. 4. We therefore refine the data cost and propose the constrained angular
Fig. 4: Constrained histogram analysis. (a) The center pinhole image with a spatial patch; (b) Angular patch ($\alpha = 5$); (c) Angular patch ($\alpha = 35$); (d, e) The ordinary histogram of (b, c); (f, g) The constrained histogram of (b, c). Ground truth $\alpha$ is 5. The contrast of each patch is enhanced for better visualization.

entropy cost (CAE) to reduce the effect of the occluder and noise. Instead of applying a uniform weight to each intensity in the histogram, we use a weight function to build the adaptive histogram $g_i$, as follows.

$$w(i) = \exp\left(-\frac{|i - A_0(u_c, v_c)|^2}{2\sigma^2}\right)$$  \hspace{1cm} (10)

$$g(i) = w(i) \cdot h(i)$$  \hspace{1cm} (11)

where $|i - A_0(u_c, v_c)|^2$ is the difference between intensity $i$ and the intensity of the center pixel in angular patch $A_0$. We do not consider the geometric proximity as the bilateral filter does because there is no spatial relationship between each pixel in the angular patch. Note that each pixel in the angular patch with the correct depth should have the same intensity value. As the adaptive histogram gives a weight of almost zero to an intensity that is far from the center pixel intensity, we can assume that the adaptive histogram is the constrained histogram. The constrained angular entropy $\tilde{H}$ for each color channel is then measured using the constrained histogram, as follows.

$$\tilde{H}(\mathbf{p}, \alpha) = -\sum_i \frac{g(i)}{|g|} \log(g(i))$$  \hspace{1cm} (12)

where $|g|$ is the sum of the constrained histogram $g_i$, $\frac{g(i)}{|g|}$, the normalization value of the constrained histogram $g_i$, denotes

the weight for each logarithmic value of the constrained histogram. To integrate the costs from the three channels, we apply average pooling, formulated as follows.

$$C(\mathbf{p}, \alpha) = \frac{\tilde{H}_R(\mathbf{p}, \alpha) + \tilde{H}_G(\mathbf{p}, \alpha) + \tilde{H}_B(\mathbf{p}, \alpha)}{3}$$  \hspace{1cm} (13)

where $\{R, G, B\}$ denotes the color channels. Unlike [21] which utilizes both average and max pooling together, we only utilize average pooling because it performs better in the general case. While max pooling performs well on an object with a dominant color (i.e., a red object with high intensity in the red channel and approximately zero intensity in the green and blue channels), it fails on other objects or surfaces because it only selects an entropy value from one channel. Conversely, average pooling considers the entropy values in all channels which make it more robust to various kinds of objects and noise.

Fig. 4 shows the original and constrained histograms of two angular patches with different depth labels. It is shown that the dominant occluder pixels produce inconsistent histograms on the ground truth depth label, which affects the cost computation. The previous angular entropy cost fails to achieve the minima on the ground truth depth label as shown in Fig. 3 (c). However, the constrained histogram reduces the effect of the occluder pixels, giving the novel constrained angular entropy superior discrimination power.
Fig. 6: Disparity maps of (a) Adaptive defocus cost; (b) Constrained adaptive defocus cost.

Fig. 3 shows that CAE achieved the minima in ground truth depth in all cases.

3.4 Constrained Adaptive Defocus Cost (CAD)

Conventional defocus costs for light field depth estimation are robust against noisy scenes but fail in occluded regions [14], [15]. We observe that the blurry artifact from the occluder in the refocus image produces ambiguity in the conventional data costs. Fig. 5 (a) and (c)–(f) show the spatial patches in the center pinhole image and refocus images, respectively. The conventional defocus data cost fails to produce optimal response on the patches. We compute the difference maps between the patches in the center image and refocus images to clarify the observations, as shown in Fig. 5 (g)–(i). The large patch in the non-ground truth label (α = 67) has a smaller difference than the ground truth (α = 27). To address this problem, we propose the adaptive defocus cost (AD) that is robust against both noise and occlusion.

Based on the difference map observations, AD should be able to find the minimum response among the subregions. Instead of measuring the response across the whole region (15 × 15) that is affected by the blurry artifact, we look for a subregion without blur, i.e., one that is not affected by the occluder, by dividing the original patch (15 × 15) into 9 subpatches (5 × 5). We then measure the defocus response $D_{c}^{res}(p, \alpha)$ of each subpatch $N_c(p)$ independently as follows.

$$D_{c}^{res}(p, \alpha) = \frac{1}{|N_c(p)|} \sum_{q \in N_c(p)} | R_\alpha(q) - P(p) |$$

(14)

where $c$ is the index of the subpatch and $P$ is the center pinhole image. The initial defocus cost (ID) is computed as the minimum patch response at the subpatch $c^*$ (i.e., $c^* = \min_c D_{c}^{res}(p, \alpha)$) [31].

However, the initial cost still produces ambiguity between the occluder and occluded regions as shown in Fig. 5 (b). To discriminate between the two cases, we introduce an additional color similarity constraint $D_{col}^{c}$, representing the difference between the mean color of the minimum subpatch and the center pixel color $P(p)$. This is formulated as follows.

$$D_{c}^{col}(p, \alpha) = \frac{1}{|N_c(p)|} \sum_{q \in N_c(p)} R_\alpha(q) - P(p)$$

(15)

where $\gamma$ is the influence parameter of the constraint. Fig. 5 (b) compares the data cost curves of the proposed AD and Tao’s defocus cost (CD) [15]. It is shown that the proposed data cost is able to find the correct disparity in the occluded region. The preliminary results of the adaptive defocus data cost are available in [21].

However, the adaptive defocus cost may result in a noisy depth label, due to the color similarity constraint. It can be seen from Fig. 6 (a) that the result of the adaptive defocus cost is still noisy. To refine the defocus data cost, we develop a constrained adaptive defocus cost (CAD). Instead of adding the color similarity constraint after finding the minimum subpatch cost, we refine and add the constraint to the defocus response before finding the minimum cost. The refined color similarity constraint is defined as follows.

$$D_{c}^{col}(p, \alpha) = \min | R_\alpha(q) - P(p) |$$

(17)

where $q \in N_c(p)$. The constraint denotes the minimum difference between the pixels in a subpatch $N_c(p)$ and the center pixel of the patch $P(p)$. Instead of dividing the large patch into 9 subpatches, we use all the possible subpatches inside the large patch. For each subpatch, we compute the defocus response and the constraint and then identify the minimum response. The final data cost is therefore defined as follows.

$$D(p, \alpha) = \min(D_{c}^{res}(p, \alpha) + \gamma D_{c}^{col}(p, \alpha))$$

(18)

The performance of the data cost depends on successfully identifying a clean subpatch, and therefore on the size of the main patch and the subpatch, as the larger the main patch, the greater the possibility of finding a clean subpatch. However, this introduces a high level of complexity. We select the optimum size empirically so that the main patch and subpatch are of the proper size. Fig. 5 (b) shows that the constrained adaptive defocus cost achieves the minimum cost in the ground truth depth label and has no cost ambiguity. Furthermore, the constrained adaptive defocus cost produces less noisy results than the adaptive defocus cost as shown in Fig. 6.

3.5 Data Cost Integration

Both proposed data costs are then combined to take advantage of their respective strengths, as CAE is robust against


Table 1: The MSE across all light field datasets and noise levels.

<table>
<thead>
<tr>
<th>dataset</th>
<th>Local</th>
<th>+GC</th>
<th>+EPF</th>
<th>+EPF-GC</th>
<th>Local</th>
<th>+GC</th>
<th>+EPF</th>
<th>+EPF-GC</th>
<th>Local</th>
<th>+GC</th>
<th>+EPF</th>
<th>+EPF-GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>0.1037</td>
<td>0.0365</td>
<td>0.0123</td>
<td>0.0113</td>
<td>0.1283</td>
<td>0.1125</td>
<td>0.0795</td>
<td>0.0763</td>
<td>0.0643</td>
<td>0.0202</td>
<td>0.0072</td>
<td>0.0067</td>
</tr>
<tr>
<td>V</td>
<td>0.0625</td>
<td>0.1055</td>
<td>0.0117</td>
<td>0.0827</td>
<td>0.2818</td>
<td>0.2796</td>
<td>0.2745</td>
<td>0.2628</td>
<td>0.0498</td>
<td>0.0984</td>
<td>0.0617</td>
<td>0.0610</td>
</tr>
<tr>
<td>BCM</td>
<td>0.1557</td>
<td>0.0872</td>
<td>0.0294</td>
<td>0.0271</td>
<td>0.1850</td>
<td>0.1021</td>
<td>0.0740</td>
<td>0.0684</td>
<td>0.1182</td>
<td>0.0690</td>
<td>0.0088</td>
<td>0.0073</td>
</tr>
<tr>
<td>SSD_B</td>
<td>0.0644</td>
<td>0.0213</td>
<td>0.0131</td>
<td>0.0288</td>
<td>0.1518</td>
<td>0.1481</td>
<td>0.1669</td>
<td>0.1616</td>
<td>0.0542</td>
<td>0.0162</td>
<td>0.0138</td>
<td>0.0117</td>
</tr>
<tr>
<td>SSD_P</td>
<td>0.0641</td>
<td>0.0305</td>
<td>0.0325</td>
<td>0.0289</td>
<td>0.1474</td>
<td>0.1341</td>
<td>0.1975</td>
<td>0.1520</td>
<td>0.0467</td>
<td>0.0131</td>
<td>0.0125</td>
<td>0.0106</td>
</tr>
<tr>
<td>PRED</td>
<td>0.1608</td>
<td>0.2082</td>
<td>0.0408</td>
<td>0.0394</td>
<td>0.2107</td>
<td>0.2096</td>
<td>0.1862</td>
<td>0.1820</td>
<td>0.1465</td>
<td>0.0705</td>
<td>0.0190</td>
<td>0.0187</td>
</tr>
<tr>
<td>OPFD</td>
<td>0.4033</td>
<td>0.0416</td>
<td>0.1096</td>
<td>0.0698</td>
<td>0.3758</td>
<td>0.0852</td>
<td>0.2548</td>
<td>0.2408</td>
<td>0.1420</td>
<td>0.0450</td>
<td>0.1199</td>
<td>0.1231</td>
</tr>
<tr>
<td>FSS</td>
<td>0.5957</td>
<td>0.5710</td>
<td>0.6504</td>
<td>0.6523</td>
<td>0.6481</td>
<td>0.5980</td>
<td>0.5543</td>
<td>0.5392</td>
<td>0.3583</td>
<td>0.3741</td>
<td>0.3462</td>
<td>0.3690</td>
</tr>
<tr>
<td>AD</td>
<td>0.0524</td>
<td>0.0360</td>
<td>0.0348</td>
<td>0.0280</td>
<td>0.1511</td>
<td>0.1286</td>
<td>0.1018</td>
<td>0.0910</td>
<td>0.0440</td>
<td>0.0161</td>
<td>0.0206</td>
<td>0.0138</td>
</tr>
<tr>
<td>CAD</td>
<td>0.0315</td>
<td>0.0299</td>
<td>0.0267</td>
<td>0.0237</td>
<td>0.1212</td>
<td>0.1163</td>
<td>0.1139</td>
<td>0.1071</td>
<td>0.0186</td>
<td>0.0100</td>
<td>0.0122</td>
<td>0.0106</td>
</tr>
<tr>
<td>V-LG</td>
<td>0.0905</td>
<td>0.0513</td>
<td>0.0453</td>
<td>0.0441</td>
<td>0.2285</td>
<td>0.2216</td>
<td>0.2017</td>
<td>0.1907</td>
<td>0.0458</td>
<td>0.0410</td>
<td>0.0159</td>
<td>0.0160</td>
</tr>
<tr>
<td>SSD_B-PRD</td>
<td>0.0315</td>
<td>0.0268</td>
<td>0.0344</td>
<td>0.0323</td>
<td>0.1753</td>
<td>0.1794</td>
<td>0.1750</td>
<td>0.1710</td>
<td>0.0301</td>
<td>0.0199</td>
<td>0.0144</td>
<td>0.0130</td>
</tr>
<tr>
<td>SSD_P-GRAD</td>
<td>0.1097</td>
<td>0.0447</td>
<td>0.0291</td>
<td>0.0272</td>
<td>0.1759</td>
<td>0.1546</td>
<td>0.1487</td>
<td>0.1385</td>
<td>0.0703</td>
<td>0.0235</td>
<td>0.0108</td>
<td>0.0086</td>
</tr>
<tr>
<td>OVF-OPFD</td>
<td>0.0539</td>
<td>0.0147</td>
<td>0.0668</td>
<td>0.0480</td>
<td>0.1006</td>
<td>0.0962</td>
<td>0.2040</td>
<td>0.1790</td>
<td>0.0143</td>
<td>0.0122</td>
<td>0.0401</td>
<td>0.0264</td>
</tr>
<tr>
<td>AE-AD</td>
<td>0.0151</td>
<td>0.0106</td>
<td>0.0083</td>
<td>0.0078</td>
<td>0.0798</td>
<td>0.0667</td>
<td>0.0588</td>
<td>0.0563</td>
<td>0.0143</td>
<td>0.0075</td>
<td>0.0066</td>
<td>0.0044</td>
</tr>
<tr>
<td>AE-CAD</td>
<td>0.0128</td>
<td>0.0097</td>
<td>0.0079</td>
<td>0.0075</td>
<td>0.0587</td>
<td>0.0510</td>
<td>0.0562</td>
<td>0.0537</td>
<td>0.0145</td>
<td>0.0060</td>
<td>0.0066</td>
<td>0.0063</td>
</tr>
</tbody>
</table>

The occlusion region and less sensitive to noise, whereas CAD is robust against noise and less sensitive to occlusion. CAD depends on an angular patch of only a single pixel, and is therefore not robust when the noise level is high. In contrast, CAD utilizes the spatial patch information in the refocus image, and is therefore more stable at different noise levels. Fig. 7 shows the MSE curve of the proposed data costs for noisy light field images with variances from 0 to 0.3 and an interval of 0.02. CAE gives a smaller error when the noise is weak, whereas CAD is superior to CAE at high noise levels. Similar results are obtained in both non-optimized and optimized evaluations. It is therefore demonstrated that combining the two data costs yields an improved data cost that is robust against both occlusion and noise.

4 Experimental Results

The proposed algorithm is implemented on an Intel i7 4770 @ 3.4 GHz with 16 GB RAM. The performance of the proposed data costs is compared with recently reported light field depth estimation data costs. Where possible, we use the code shared by Jeon et al. [12], Tao et al. [14], and Wang et al. [19] and implement the code of other methods that are note available. We compared 17 individual data costs (11 correspondence costs and 6 defocus costs) and 6 integrated data costs. The correspondence data costs are the binned entropy cost (BE) [18], variance based cost (V) [14], bilateral consistency metric (BCM) [8], SSD with bilinear interpolation (SSD_B) [15], SSD with phase based interpolation (SSD_P) [12], sum of gradient differences with phase based interpolation (GRAD) [12], occlusion-aware variance based cost (OV) [19], spinning paralellogram operator cost (SPO) [22], modified Parzen window estimation (PEF) [27], angular entropy cost (AE) [21], and constrained angular entropy cost (CAE). Sequentially, the defocus data costs are the Laplacian operator cost (LO) [14], pinhole and re-images differences (PRD) [15], occlusion-aware PRD (OPRD) [19], focal stack symmetry cost (FSS) [9], adaptive defocus cost (AD) [21], and constrained adaptive defocus cost (CAD). The integrated data cost is the summation of two data costs as presented in previous works: V-LO [14], SSD_B-PRD [15], SSD_P-GRAD [12], OV-OPFD [19], AE-AD [21], and CAE-CAD (Proposed).

As conventional depth estimation methods apply different methods of optimization, it is challenging to make a fair comparison between them. We first compare the depth estimation results without global optimization to identify the discrimination power of each data cost. The globally optimized depth is compared, using a range of challenging scenes. We use three optimization methods: graph cut (GC) [32], edge-preserving filter (EPF) [34], and edge-preserving filter [34] + graph cut [32] (EPF-GC). In the quantitative evaluation, we use the MSE and BP methods, derived as follows.

\[
MSE = \frac{1}{N} \sum_p |GT(p) - \alpha^*(p)|^2
\]

\[
BP = \frac{1}{N} \sum_p |GT(p) - \alpha^*(p)| > \delta
\]
### TABLE 2: The BP (%) across all light field datasets.

<table>
<thead>
<tr>
<th>LF Dataset</th>
<th>(\delta = 2)</th>
<th>(\delta = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{Local} + \text{GC})</td>
<td>(\text{EPF})</td>
<td>(\text{EPF-GC})</td>
</tr>
<tr>
<td>BE</td>
<td>38.24</td>
<td>15.47</td>
</tr>
<tr>
<td>V</td>
<td>62.03</td>
<td>44.48</td>
</tr>
<tr>
<td>BCM</td>
<td>35.24</td>
<td>30.76</td>
</tr>
<tr>
<td>SSD_B</td>
<td>20.04</td>
<td>8.09</td>
</tr>
<tr>
<td>SSD_P</td>
<td>23.13</td>
<td>12.71</td>
</tr>
<tr>
<td>GRAD</td>
<td>37.59</td>
<td>23.77</td>
</tr>
<tr>
<td>OV</td>
<td>23.32</td>
<td>9.98</td>
</tr>
<tr>
<td>CAE</td>
<td>13.86</td>
<td>10.06</td>
</tr>
<tr>
<td>LO</td>
<td>29.17</td>
<td>26.47</td>
</tr>
<tr>
<td>PRD</td>
<td>27.29</td>
<td>26.36</td>
</tr>
<tr>
<td>OPRD</td>
<td>52.49</td>
<td>16.92</td>
</tr>
<tr>
<td>FSS</td>
<td>35.63</td>
<td>25.63</td>
</tr>
<tr>
<td>AD</td>
<td>38.98</td>
<td>34.13</td>
</tr>
<tr>
<td>CAD</td>
<td>26.90</td>
<td>26.36</td>
</tr>
</tbody>
</table>

![Fig. 8: The MSE comparison for various light field datasets.](image)

(a) Non-optimized results of the image (Local); (b) Non-optimized results of the occlusion regions (Local); (c) Optimized results of the image (Local+EPF-GC); (d) Optimized results of the occlusion regions (Local+EPF-GC). For better visualization, we use the logarithm value of \(MSE \times 1000\).
TABLE 3: The BP (%) across all noise levels (Variance = 0.02, 0.04, 0.06, 0.08, 0.1).

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Squared Error</th>
<th>Noise Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Local</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>Local</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>Local</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>Local</td>
<td>0.06</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Fig. 9: The MSE comparison for various noise levels using each local data cost. (a) Correspondence costs; (b) Defocus costs; (c) Best of correspondence and defocus costs; (From left to right) Non-optimized results of the image (Local); Non-optimized results of the occlusion regions (Local); Optimized results of the image (Local+EPF-GC); Optimized results of the occlusion regions (Local+EPF-GC).
Fig. 10: Comparison of the non-optimized disparity maps (Mona dataset) using the individual data costs; (a) BE; (b) V; (c) BCM; (d) SSD_B; (e) SSD_P; (f) GRAD; (g) OV; (h) SPÖ; (i) AE; (j) CAE; (k) LO; (l) PRD; (m) OPRD; (n) FSS; (o) AD; (p) CAD.

The depth search range is 1, 2, 3, ..., 74, 75 for all datasets. Using the Lytro Illum light field, the computational times for the CAE and CAD data costs are 359.74 and 33.35 seconds in the Matlab environment, respectively. Note that we did not perform any code optimization in these experiments. The computational time can be increased up by implementing the data costs computation on C and GPU environments. Our code implementation is available on the project website (http://image.inha.ac.kr/lfdepth/).

4.1 Synthetic Light Fields

First, we evaluate the performance of each data cost under different optimization methods for all synthetic datasets [24]. Table 1 and Table 2 show the average values of MSE and BP across all datasets. For the BP calculation, we compare the results when using two different values. To evaluate the data cost for each light field image, Fig. 8 presents the bar charts of MSE for non-optimized (Local) and optimized (EPF-GC) data costs. The proposed data costs (CAE, CAD, CAE-CAD) achieve the best performance overall, especially in the occlusion region. While conventional data costs depend on the optimization method used, the proposed data costs produce the smallest error without applying optimization.

We then generated the synthetic noisy light field images of the Mona dataset, using a Gaussian noise with variances from 0 to 0.1 and an interval of 0.02. The average values of MSE and BP are shown in Table 1 and Table 3, respectively, and the performance of the non-optimized and optimized data costs are plotted in Fig. 9. It can be seen that the proposed data costs are also robust against noise. The constrained angular entropy and constrained adaptive defocus costs maintain the small MSE value.

For qualitative evaluation, we show the disparity maps of each non-optimized (Local) and optimized (EPF-GC) cost in Fig. 10 and Fig. 11, respectively. The disparity maps of the optimized integrated data costs for the clean and noisy Mona datasets are shown in Fig. 12. The proposed data costs prove to be more robust against occlusion and noise than conventional data costs both with and without optimization.

4.2 Real Light Fields

Fig. 13 shows the center pinhole images of light fields captured by the Lytro Illum. To compare the performance
of each proposed data cost, the local disparity maps are shown in Fig. 13 (b) and (c). In both local data costs, the edges of thin objects are well preserved, such as the leaves in the second column, the racket in the third column, and the spokes of the wheel in the fifth column. These results demonstrate that the constrained angular entropy and constrained adaptive defocus costs are robust against occlusion in noisy scenes. Fig. 14 compares the optimized disparity maps for each integrated data cost, and shows that the proposed method preserves the edges of thin objects better than other methods. While conventional approaches depend on optimization methods to deal with occlusion and noise, our method uses occlusion and noise-aware data costs. Furthermore, the proposed data costs do not require the detection of edges in the center pinhole image as done in [19].

In addition, we test our method on the real light field dataset used in [27]. Note that the dataset is denser compared to the Lytro Illum light field. The light field image is captured by a commercial DSLR camera on a motorized linear stage so that the image is less noisy and higher resolution than the image captured by the Lytro Illum. Due to memory limitation, we resize the original image with a factor $\frac{1}{4}$. Figure 15 shows the qualitative comparison between the CAE and PWE data costs. It proves that the proposed method performs better on this dataset.

4.3 Limitations and Future Work

The proposed data costs are shown to perform well on occlusion and noise, but the constrained adaptive defocus cost fails when no non-blurred subpatch is available. As with the other methods, the proposed data costs also perform poorly on textureless regions. Finally, the quality of the final result still depends on the optimization method selected. We believe that further improvement is possible by applying a better optimization method, although this is not the main focus of the present paper. In this research, we integrate the data costs using a weighted summation, but it is important that a confidence metric be found that sets a weight for each data cost, rather than using the uniform weight. To the best of our knowledge, no studies have been conducted on deriving a reliability value for light field data costs.
Fig. 12: Comparison of the optimized disparity maps (Mona dataset) using the integrated data costs. (a) CAE-CAD (Proposed); (b) V-LO [14]; (c) SSD_B-PRD [15]; (d) SSD_P-GRAD [12]; (e) OV-OPRD [19]; (f) AE-AD [21]. (Left) Clean light fields; (Right) Noisy light fields with variance 0.10.

Fig. 13: Results for the real light fields captured by Lytro Illum. (a) Center pinhole images; (b) Disparity maps of CAE (Local); (c) Disparity maps of CAD (Local).

5 CONCLUSIONS

In this paper, we proposed a framework for occlusion and noise-aware light field depth estimation. Two observations on the angular patch and refocus image were found when occlusion exists. Two novel data costs were proposed to allow robust performance in noisy occluded regions. The constrained angular entropy metric was introduced to measure the randomness of pixel color in the angular patch while reducing the effect of the occluder and noise. A constrained adaptive defocus response was determined to increase, which provided robust performance against occlusion while maintaining noise robustness. Both data costs were integrated into the MRF framework and further optimized using edge preserving filtering and a graph cut method. In our experimental tests, the proposed method significantly outperformed conventional approaches in both occluded and noisy scenes. Finally, we conducted an exhaustive comparison and benchmarking of state-of-the-art data cost methods.
Fig. 14: Comparison of the optimized disparity maps of light fields in Fig. 13. (a) CAE-CAD (Proposed); (b) V-LO [14]; (c) SSD_B-PRD [15]; (d) SSD_P-GRAD [12]; (e) OV-OPRD [19]; (f) AE-AD [21].

Fig. 15: Comparison of the disparity maps of very dense light field dataset in [27]. (a) PWE [27]; (b) CAE (Proposed); (c) PWE+EPF-GC [27]; (d) CAE+EPF-GC (Proposed).

ACKNOWLEDGMENTS
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2016R1A2B4014731).

REFERENCES


William received his B.S. degree in computer science from Bina Nusantara University, Indonesia, in 2011, and Ph.D. degree in information and communication engineering from Inha University, Korea, in August 2017. His research interests include computer vision, computational photography, and image processing, including 3D reconstruction and GPGPU.

Kyoung Mu Lee Kyoung Mu Lee received the B.S. and M.S. degrees in Control and Instrumentation Eng. from Seoul National University (SNU), Seoul, Korea in 1984 and 1986, respectively, and Ph. D. degree in Electrical Engineering from the University of Southern California in 1993. He is currently with the Dept. of ECE at Seoul National University as a professor. Prof. Lee has received several awards, in particular, the Most Influential Paper over the Decade Award by the IAPR Machine Vision Application in 2009, the ACCV Honorable Mention Award in 2007, the Okawa Foundation Research Grant Award in 2006, the Distinguished Professor Award from the college of Engineering of SNU in 2009, and both the Outstanding Research Award and the Shinyang Engineering Academy Award from the College of Engineering of SNU in 2010. He is currently serving as an AEIC (Associate Editor in Chief) of the IEEE TPAMI, Area Editor of the CVIU, and has served as an Associate Editor of the IEEE TPAMI, the Machine Vision Application Journal and the IPSJ Transactions on Computer Vision and Applications, and the IEEE Signal Processing Letter. He also has served (or will serve) as a General Chair of ICCV2019, ACM MM2018, Program Chair of ACCV2012, a Track Chair of ICPR2012, Area Chair of CVPR2012, CVPR2013, CVPR2015, ECV2013, ECV2014, ECV2016 and a Workshop Chair of ICCV2013. He was a Distinguished Lecturer of the Asia-Pacific Signal and Information Processing Association (APSIPA) for 2012-2013. More information can be found on his homepage http://cv.snu.ac.kr/kmlee.

In Kyu Park received his B.S. and M.S. degree and Ph. D. from Seoul National University, Korea, in 1995, 1997, and 2001, respectively, all in electrical engineering and computer science. From September 2001 to March 2004, he was a member of the technical staff at the Samsung Advanced Institute of Technology, Giheung, Korea. Since March 2004, he has been with the School of Information and Communication Engineering, Inha University, Incheon, Korea, where he is a full professor. From January 2007 to February 2008, he was an exchange scholar at Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. From September 2014 to August 2015, he was a visiting associate professor at MIT Media Lab, Cambridge, MA, USA. His research interests include the joint area of computer vision and graphics, including 3D shape reconstruction from multiple views, image-based rendering, computational photography, and GPGPU for image processing and computer vision. He is a senior member of the IEEE and a member of the ACM.