IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 2, 2021, accepted July 17, 2021, date of publication July 20, 2021, date of current version July 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3098819

Occlusion Handling by Successively Excluding Foregrounds for Light Field Depth Estimation Based on Foreground-Background Separation

JAE YOUNG LEE[1], RAE-HONG PARK[2], (Life Senior Member, IEEE), AND JUNMO KIM[1], (Member, IEEE)

[1]School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea
[2]Department of Electronic Engineering, ICT Convergence Disaster/Safety Research Institute, Sogang University, Seoul 04107, South Korea

Corresponding author: Junmo Kim (junmo.kim@kaist.ac.kr)

ABSTRACT This paper proposes a depth from light field (DFLF) method specifically to deal with occlusion based on the foreground-background separation (FBS). The FBS-based methods infer the disparity maps by accumulating the binary maps which divide whether each pixel is the foreground or background. Although there have been widely studied to handle the occlusion problem with the cost-based method, there are not enough researches to handle the occlusion problem with the FBS-based methods yet. We found that errors around the occlusion boundary in the resulting disparity maps of the FBS-based methods arise from the fattened foreground by the light field reprameterization. To avoid fattened foregrounds, the inferred foreground maps in the front region with respect to the disparity axis could be utilized in the back region in the three-dimensional volume construction, which corresponds to the cost volume construction in the cost-based methods. With the front-to-back scanning manner of the FBS-based method, by successively excluding inferred foreground maps, errors around occlusion boundary could be effectively reduced in the resulting disparity maps. With synthetic and real LF images, the proposed method shows reasonable performance compared to the existing methods and better performance than existing FBS-based methods.

INDEX TERMS Occlusion, depth estimation, foreground-background separation, light field, cost volume.

I. INTRODUCTION

Depth information plays important roles in various computer vision and image processing applications such as saliency detection [1]–[4], pose estimation [5]–[7], segmentation [8], [9], and object detection [10], [11]. To obtain depth information, various three-dimensional (3-D) imaging techniques [12] had been developed including triangulation-based (such as light field (LF) camera [13] and stereo camera [14]), time-of-flight based imaging [12] (such as light detection and ranging [15]–[17]), and etc. Among them, a LF camera simultaneously captures four-dimensional (4-D) information including both two dimensional (2-D) spatial and 2-D angular information, where depth information is encoded.

In LF image processing, depth from LF (DFLF) is considered as a mid-level process whose performance directly influences high-level user interface or applications-related algorithms [18]. As a mid-level process, various methods have been studied to obtain a depth map with high accuracy from LF image [18]–[20]. However, there are still many practical problems such as occlusion boundary, camera noise, low texture, fine structure, and so on [19].

According to a taxonomy of the DFLF [20], the existing local DFLF methods could be categorized into 9 (3×3) groups with the combination of two criteria: three LF representations and three signal processing viewpoints. In view of the LF representation, the local DFLF methods could be categorized according to LF representation used for derivations of the depth information: the epipolar plane image (EPI)-based methods, the angular patch-based methods, and the focal stack-based methods. In the signal processing viewpoint, according to properties of intermediate images, there are three groups of the existing DFLF methods: the cost-based methods, the foreground-background separation (FBS)-based methods, and the depth model-based methods. These nine categories could be individually divided into two groups: conventional and learning-based approaches.

The associate editor coordinating the review of this manuscript and approving it for publication was Turgay Celik.

In the depth from stereo (DFS), which could be considered as a subset of LF in view of input images, learning-based approaches has been widely studied and shown the state-of-the-art performance. With the learning-based approaches, Žbontar and LeCun [21] proposed a method to learn stereo matching, which can be considered as a cost-based method. Concurrently, Dosovitskiy *et al*. [22] proposed an end-to-end method for disparity estimation, which can be considered as a depth model-based method. Recently, Zhang *et al*. [23] and Xu and Zhang [24] proposed methods for the DFS with the learning-based approaches in an end-to-end manner whose pipelines are backboned on those of the conventional DFS approaches. Also, the DFLF using learning-based approaches show the state-of-the-art performance [25], [26]. Shin *et al*. [25] proposed the fully-convolutional neural network in the end-to-end manner, which could be considered as the depth model-based method. Tsai *et al*. [26] embedded view selection network using the attention module in an end-to-end manner based on the conventional DFLF pipelines. Because the conventional approaches had been sufficiently studied, it was possible for the learning-based approaches to achieve the state-of-the-art performance. Although it is obvious that the research trend is the learning-based approach for high performance, the conventional approaches are also important because it is a baseline for network architectures of the learning-based approaches.

This paper proposes a method using the conventional approach to handle the occlusion boundary problem in the framework of foreground-background separation (FBS)-based methods in the DFLF [27], [28]. Although the practical problems have been widely dealt with frameworks of the cost-based and depth model-based methods, those are not sufficiently dealt with frameworks of the FBS-based methods. Using the LF reparameterization [29] for every possible disparity plane with fixed disparity step size, the FBS-based methods construct a 3-D volume where each 2-D slice is divided into two regions: foreground and background. As a local DFLF method, the 3-D volume constructed by the FBS-based methods corresponds to the cost-volume in the cost-based methods [20]. Around the occlusion boundary in resulting disparity map of the FBS-based methods, errors occurred because background regions are influenced by the foreground regions. To reduce these errors, using the front-to-back scanning manner of the FBS-based methods, the proposed method successively excludes the foreground information when computing slices in back sides of a given disparity range. Fig. 1 briefly shows a flow of the foreground information in the 3-D volume construction with the front-to-back scanning manner. A disparity skip counting parameter J determines which disparity plane in back side of the 3-D volume would use the foreground information in front side of the 3-D volume.

The framework of FBS-based methods was derived from the optical phenomenon, which is called flipping [30]. With respect to an in-focused plane, while the objects in the foreground are non-flipped, the objects in the background are flipped. Using the flipping phenomenon in LF images and the LF reparameterization [29], Lee and Park [27] proposed a framework for DFLF by accumulating binary maps, which divide the foreground and background with respect to the in-focused plane.

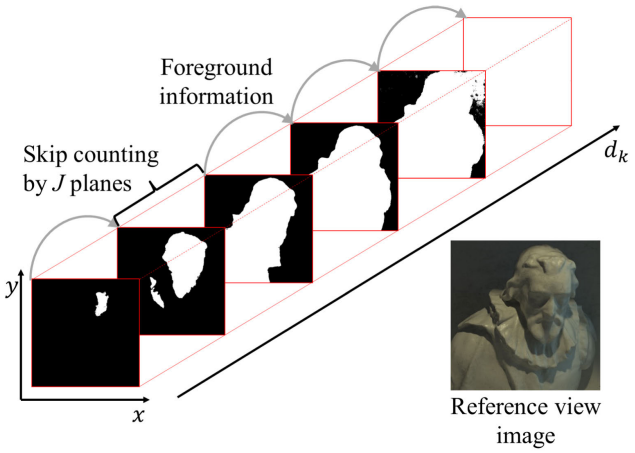

FIGURE 1. A flow of the foreground information in the 3-D volume construction with the front-to-back scanning manner in the proposed method. By using the skip counting parameter J, a foreground map of $k - J^{th}$ plane is utilized for computing foreground maps of k^{th} plane.

While it was possible to derive a hand-crafted feature for the FBS because the LF has rich information, it is still challenging to derive a hand-crafted feature for the FBS in the DFS. Badki *et al*. [31] recently proposed the FBS for the DFS using the learning-based approach. If the FBS framework for the DFS with the conventional approach had been developed, the FBS for the DFS could have been developed earlier. In addition, the occlusion problem in DFS [32], monocular depth estimation [33], and semantic segmentation [34] are still studied within both conventional and learning-based approaches. In these methods, performance near occlusion is not perfect yet. Especially in autonomous driving applications, to prepare for safety and unexpected circumstances, we believe that any solution should properly handle the occlusion problem and be explainable for their performance near the occlusion boundary. In these contexts, a well defined conventional approach might be not only needed for better performance within learning-based framework but also helpful for explainability with deep learning techniques.

The contribution of this paper is as follows. First, an occlusion handling method in the framework of the FBS-based method is proposed without additional scene information, which is different from the cost-based method. Second, with a given disparity plane, a method converting a given single 2-D image (e.g., binary mask) to 4-D LF image is presented. We believe that the second contribution can be utilized when detected features at a certain disparity plane are transferred to the other disparity plane.

The rest of this paper is organized as follows. In Section II, the framework of FBS-based methods and occlusion handling approaches in the existing DFLF methods are briefly

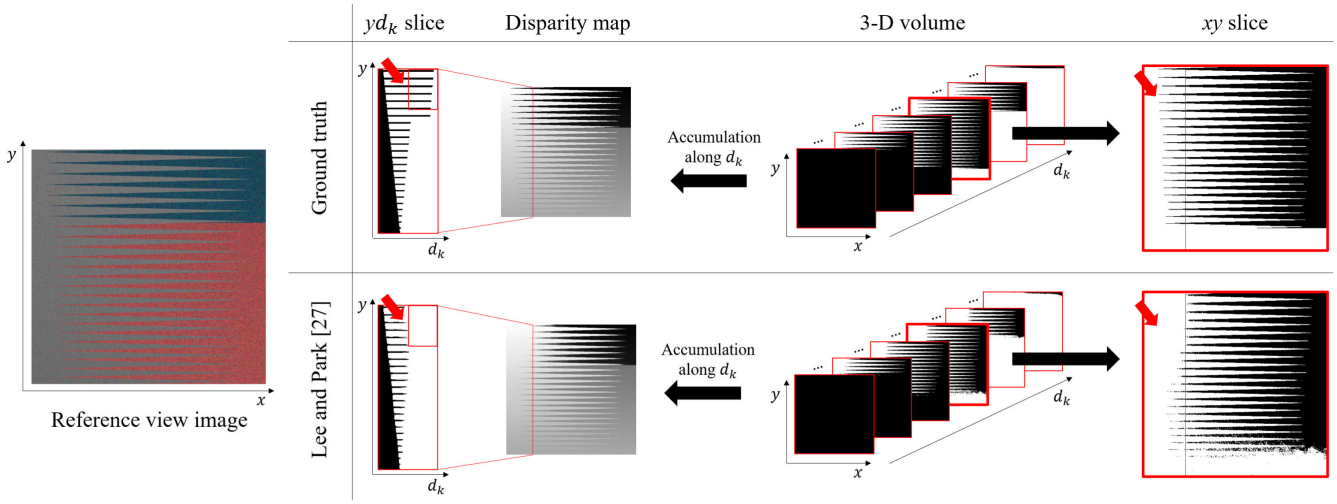

FIGURE 2. Foreground fattening around occlusion boundary in the FBS [27]. In yd_k slices, 3-D volume, and xy slice, the white and black pixels represent the foreground and background, respectively. In both yd_k and xy slice, while GTs show sharp depth discontinuities, the resulting foreground maps by Lee and Park [27] show that the depth discontinuities are ambiguous due to the foreground fattening.

introduced. In Section III, the proposed method is presented in detail. In Section IV, the experimental results with synthetic and real images are illustrated and discussed. Finally, in Section V, the conclusion is given.

II. RELATED WORKS

A. LF REPARAMETERIZATION

With an input light field $L_0(x, y, u, v)$ and arbitrary disparity plane d, the reparameterized LF L_d could be expressed as [29]

$$L_d(x, y, u, v) = L_0(x + ud, y + vd, u, v), \tag{1}$$

where (x, y) and (u, v) signifies the spatial and angular coordinates, respectively. The light field reparameterization is a part of digital refocusing algorithm [13]. By averaging the reparameterized LF, refocused image could be obtained. Although the LF reparameterization was derived by moving virtual focal plane in view of lenslet images, it could be also interpreted in the subaperture images and EPIs. In view of the subaperture images (or multi-view images), LF reparameterization could be considered to warp multi-view images with a certain disparity value at once. In view of EPIs, it could be considered as EPI shearing, which slopes in EPIs correspond to certain disparity values.

B. FBS-BASED METHODS

With given disparity range $[d_{\min}, d_{\max}]$ and disparity step size d_{step}, a set of quantized disparity planes $\Pi = \{d_1, d_2, \ldots, d_k, \ldots, d_K\}$ could be pre-defined. The reparameterized LF with a quantized disparity plane $L_{d_k}(x, y, u, v)$ can be obtained by reparameterizing captured LF image $L_0(x, y, u, v)$ to every possible disparity plane d_k. With $L_{d_k}(x, y, u, v)$, the FBS-based methods construct a 3-D volume $C(x, y, d_k)$, which corresponds to the cost volume in view of the cost-based methods. The 3-D volume of the FBS-based methods is composed of multiple slices, and each slice is divided into two regions: foreground and background. Each slice of the initial 3-D volume $C_{d_k}(x, y)$ is computed by the disparity sign voting which is derived from the optical phenomenon "flipping" [30]. Then, using the guided image filtering [35], $C(x, y, d_k)$ is refined to edge-preserving smoothed 3-D volume $\tilde{C}(x, y, d_k)$, which corresponds to the cost volume filtering in the cost-based methods [36]. In the FBS-methods, the ideal one-dimensional (1-D) profile, which corresponds to the cost profile in the cost-based methods, along disparity plane is a sign function. At every spatial location (x, y), zero crossing points are considered as the candidates of optimal disparity. With candidates, errors between ideal sign functions centered at the zero-crossing points and an edge preserving smoothed 1-D profile $\tilde{C}_{x,y}(d_k)$ are computed by the sign function approximation. Among the candidates, an optimal 1-D profile $\hat{C}^{x,y}(d_k)$ with a minimal error is chosen [28]. In the rest of this paper, the notations of the 3-D volumes for 1-D, 2-D, and 3-D viewpoints are used as follows: initial 3-D volume $C^{x,y}(d_k) = C_{d_k}(x, y) = C(x, y, d_k)$, edge-preserving smoothed 3-D volume $\tilde{C}^{x,y}(d_k) = \tilde{C}(x, y, d_k) = \tilde{C}(x, y, d_k)$, and sign function approximated 3-D volume $\hat{C}^{x,y}(d_k) = \hat{C}_{d_k}(x, y) = \hat{C}(x, y, d_k)$.

Fig. 2 shows the foreground fattening around the occlusion boundary in the FBS [27]. The first and second rows represent the ground truth (GT) and Lee and Park's method [27], respectively. With respect to the 3-D volume, the right side shows a xy slice. On the left side of the 3-D volume, yd_k slice and the resulting disparity map by accumulating 3-D volume along d_k are shown. Compared to the GT, the resulting disparity map of Lee and Park [27] is fattened around the occlusion boundary, where red arrows in yd_k and xy slices indicate. The fattened foreground is presented because the foreground are spread into the angular patch of background by the LF reparameterization. Thus, by successively excluding the

foreground map obtained at the front plane along d_k when computing the foreground map at the back plane along d_k, the fattened foreground could be effectively reduced.

C. OCCLUSION HANDLING IN DFLF

Although there are various DFLF methods, an overview in this section is focused on the DFLF methods handling the occlusion. The angular patch & cost-based methods [37]–[41] used additional scene information such as color and/or edge to make the cost volume itself robust to the occlusion boundary. Chen *et al*. [37] proposed bilateral consistency metric based on the color information in the angular patch. Williem *et al*. [38] used the range kernel of bilateral filter in the angular patch. Wang *et al*. [39] used edge information to find occlusion boundary and divided the angular patch into two regions based on a single occlusion model. Zhu *et al*. [40] generalized a single occlusion model [39] to a multiple occlusion model using K-means clustering, with edge inforamtion as in [39]. Ai *et al*. [41] proposed a more suitable cost volume construction method based on the multiple occlusion model [40]. The focal stack & cost-based method [42] constructed a single cost volume from multiple partial focal stacks to obtain un-occluded information. Strecke *et al*. [42] constructed 4 directed partial focal stacks to obtain un-occluded focal stack. Similarly, as one of the EPI & cost-based methods, [43] constructed a single cost volume from those of multi-orientation EPIs to obtain un-occluded information. Sheng *et al*. [43] constructed a single cost volume from 4 cost volumes, which are constructed by different directional EPIs, to obtain un-occluded information. As one of the desired approaches in the EPI & cost-based methods, Schilling *et al*. [44] proposed an occlusion handling method, which directly integrated into a depth model in the optimization without additional scene information. As in Schilling *et al*.'s method [44], the proposed method handles the occlusion boundary without additional scene information in the framework of FBS. To make the cost volume itself robust to the occlusion boundary, the proposed method uses the information in the cost volume itself.

III. PROPOSED OCCLUSION HANDLING DFLF METHOD

Fig. 3 shows the block diagram of the proposed method. Based on the framework of the existing FBS-based methods [27], [28], in occlusion handling FBS method, the disparity sign voting is modified and the 4-D mask generation is added to handle occlusion. The foregrounds from d_{k-J} are considered as occlusions at d_k disparity plane, where J denotes the skip counting parameter. The 4-D foreground mask $F^{d_k}_{d_{k-J}}(x, y, u, v)$, which is initialized with zero, signifies the reparameterized 4-D foreground mask into d_k using the foreground map at d_{k-J}. With $F^{d_k}_{d_{k-J}}(x, y, u, v)$, to exclude the foreground information, the disparity sign voting is modified comparing to the existing FBS-based method [27].

Algorithm 1 shows a pseudo code for the occlusion handling FBS algorithm. With a given disparity range $[d_{min}, d_{max}]$ and disparity step size d_{step}, a set of

Algorithm 1 Occlusion Handling FBS Algorithm

Require: $L_0(x, y, u, v)$, $[d_{min}, d_{max}]$, d_{step}, J, Π, Φ
$L_0(x, y, u, v)$: input LF image
$[d_{min}, d_{max}]$: disparity range
d_{step}: disparity step
J: skip counting parameter
$\Pi = \{d_1, d_2, \ldots, d_k, \ldots, d_K\}$
: a set of quantized disparity planes
$j \in \Phi = \{1, 2, \ldots, J\}$
: an initial set for occlusion handling FBS

for each $j \in \Phi$ **do**
  $k \leftarrow j$
  initialize $F^{d_k}_{d_{k-J}}(x, y, u, v)$ with 0
  while $k \leq K$ **do**
    $L_{d_k}(x, y, u, v) = LFreparam(L_0(x, y, u, v), d_k)$
    $C_{d_k}(x, y) = DSV(L_{d_k}(x, y, u, v), F^{d_k}_{d_{k-J}}(x, y, u, v))$
    $\tilde{C}_{d_k}(x, y) = GF(L_0(x, y, u_{ref}, v_{ref}), C_{d_k}(x, y))$
    $F^{d_{k+J}}_{d_k}(x, y, u, v) = 4DMaskGen(\tilde{C}_{d_k}(x, y))$
    $k \leftarrow k + J$
  end while
end for
return $\tilde{C}(x, y, d_k)$ for all d_k

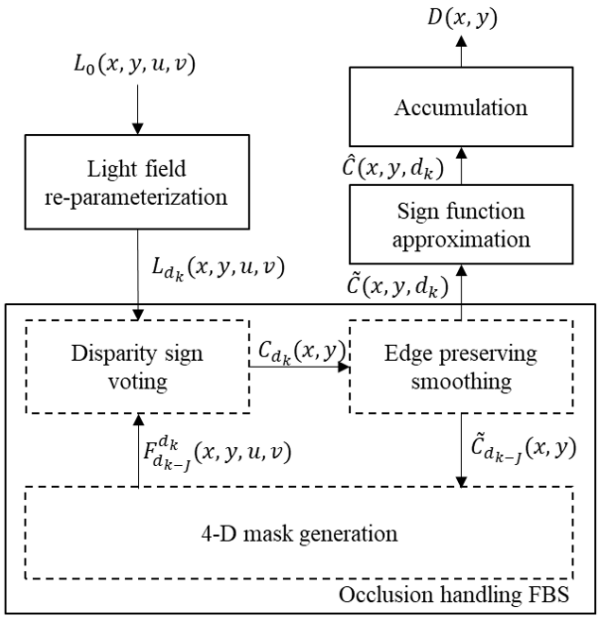

FIGURE 3. Block diagram of the proposed occlusion handling DFLF method. Based on FBS [27], the 4-D mask generation block is added to handle the occlusion boundary.

quantized disparity planes for disparity estimation $\Pi = \{d_1, d_2, \ldots, d_k, \ldots, d_K\}$ is pre-computed. d_1 and d_K are equal to d_{min} and d_{max}, respectively. d_k is computed as $d_k = d_{min} + d_{step} \times (k - 1)$. With skip counting parameter J, an initial set for occlusion handling FBS $\Phi = \{1, 2, \ldots, J\}$ is pre-defined. $LFReparam(\cdot, \cdot)$, $DSV(\cdot, \cdot)$, $GF(\cdot, \cdot)$, and $4DMaskGen(\cdot)$ are abbreviations of the functions for LF reparameterization, disparity sign voting, guided image filter,

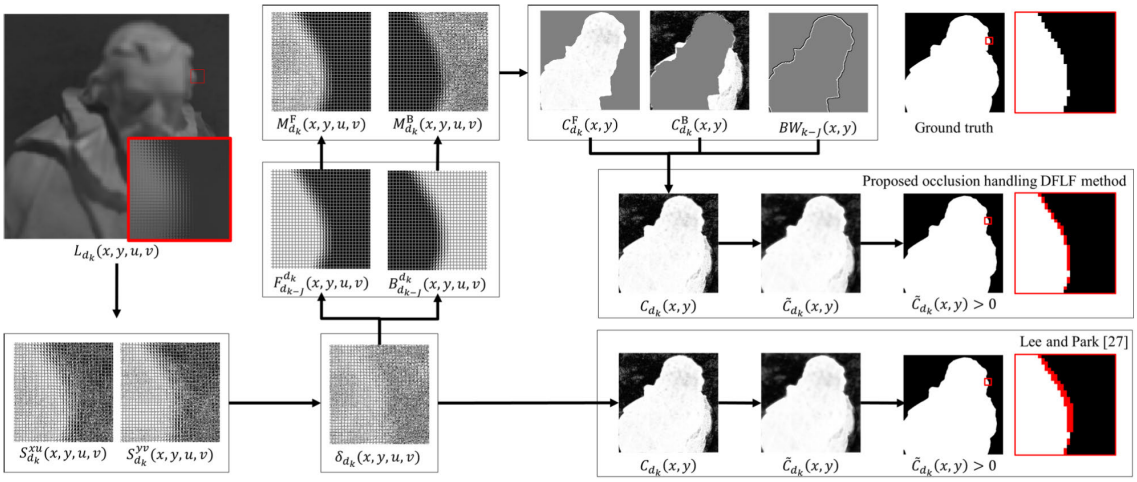

FIGURE 4. Intermediate results of the disparity sign voting and edge-preserving smoothing by the proposed occlusion handling DFLF method and Lee and Park [27]. For visualization, the red box in the lenslet image is enlarged. Whereas Lee and Park [27] directly inferred $C_{d_k}(x,y)$ using $S^{xu}_{d_k}$, $S^{yv}_{d_k}$, and δ_{d_k}, the proposed method additionally use $F^{d_k}_{d_{k-J}}$ whose foreground information is obtained from d_{k-J} plane. Compared to Lee and Park's method [27], errors around occlusion boundary are reduced (See three right-most sub-figures). Red pixels signify errors compared to GT.

and 4-D mask generation, respectively. If the occlusion handling FBS is completed, the edge-preserving smoothed 3-D volume $\tilde{C}(x,y,d_k)$ is returned.

A. DISPARITY SIGN VOTING

The disparity signs $S^{xu}_{d_k}$ and $S^{yv}_{d_k}$ from xu and yv planes are obtained using the LF gradients $\partial L_{d_k}/\partial x$, $\partial L_{d_k}/\partial y$, $\partial L_{d_k}/\partial u$, and $\partial L_{d_k}/\partial v$ [30]

$$S^{xu}_{d_k}(x,y,u,v) = \mathrm{sgn}\left(\left(\frac{\partial L_{d_k}}{\partial x}\right)\left(\frac{\partial L_{d_k}}{\partial u}\right)\right), \tag{2}$$

$$S^{yv}_{d_k}(x,y,u,v) = \mathrm{sgn}\left(\left(\frac{\partial L_{d_k}}{\partial y}\right)\left(\frac{\partial L_{d_k}}{\partial v}\right)\right), \tag{3}$$

respectively, where $\mathrm{sgn}(\cdot)$ signifies the sign function. The indicator functions of foreground $M^F_{d_k}(x,y,u,v)$ and background $M^B_{d_k}(x,y,u,v)$ are computed by the 4-D foreground mask $F^{d_k}_{d_{k-J}}(x,y,u,v)$ and the 4-D background mask $B^{d_k}_{d_{k-J}}(x,y,u,v) = 1 - F^{d_k}_{d_{k-J}}(x,y,u,v)$, respectively, with the indicator function $\delta_{d_k}(x,y,u,v)$ based on the relaxed Lambertian assumption [30]. The disparity signs are separately voted for foreground $C^F_{d_k}(x,y)$ and background $C^B_{d_k}(x,y)$. With $M^F_{d_k}$ and $M^B_{d_k}$, as an intermediate step, the foreground and background of the disparity sign voting are represented as

$$C^F_{d_k}(x,y) = \frac{\sum_{u,v}(M^F_{d_k}(x,y,u,v)\sum_{\omega}S^{\omega}_{d_k})}{2\sum_{u,v}M^F_{d_k}(x,y,u,v)}, \tag{4}$$

$$C^B_{d_k}(x,y) = \frac{\sum_{u,v}(M^B_{d_k}(x,y,u,v)\sum_{\omega}S^{\omega}_{d_k})}{2\sum_{u,v}M^B_{d_k}(x,y,u,v)}, \tag{5}$$

respectively, where $\omega \in \{xu, yv\}$. Because two directional EPIs (xu and yv planes) are combined, the numerator in (3) and (4) are devided by 2. A slice of the initial 3-D volume $C_{d_k}(x,y)$ could be obtained by

$$C_{d_k}(x,y) = \begin{cases} C^F_{d_k}(x,y), & if\ \ BW_{d_{k-J}}(x,y) = 1, \\ C^B_{d_k}(x,y), & otherwise, \end{cases} \tag{6}$$

where the decision (binary) map $BW_{d_{k-J}}(x,y) = F^{d_k}_{d_{k-J}}(x,y,u_{ref},v_{ref})$ and (u_{ref},v_{ref}) signifies the coordinates of the reference view image. Since the reference view image is always the same whether the LF image is reparameterized or not, $BW_{d_{k-J}}(x,y)$ could be extracted from $F^{d_k}_{d_{k-J}}(x,y,u,v)$. That is, $F^{d_k}_{d_{k-J}}(x,y,u_{ref},v_{ref})$ and $F^{d_{k-J}}_{d_{k-J}}(x,y,u_{ref},v_{ref})$ are equal to each other because the reference (or center) view image is not warped by the LF reparameterization. $C^F_{d_k}(x,y)$, $C^B_{d_k}(x,y)$, and $C_{d_k}(x,y)$ could have the value between -1 and 1.

Fig. 4 shows the intermediate results of the disparity sign voting and edge-preserving smoothing by the proposed method and Lee and Park [27]. Note that the disparity sign voting is the same as used in [27] and [28]. While Lee and Park's method [27] directly separates the foreground and background using $S^{xu}_{d_k}$ and $S^{yv}_{d_k}$, the proposed method separated the foreground and background using $S^{xu}_{d_k}$, $S^{yv}_{d_k}$, and $F^{d_k}_{d_{k-J}}$. $B^{d_k}_{d_{k-J}}$, $M^F_{d_k}$, $M^B_{d_k}$, and $BW_{d_{k-J}}(x,y)$ could be obtained from $F^{d_k}_{d_{k-J}}$. In the mid-top region of Fig. 4, for $C^F_{d_k}(x,y)$ and $C^B_{d_k}(x,y)$, the gray pixels represent not-interested region. The other pixels represent the value of the disparity sign voting based on gray scale color map with range $[-1,1]$. That is, the pixels closer to 1 and -1 are represented as white and black, respectively. Although the decision map $BW_{k-J}(x,y)$ is a binary map, for visualization around the occlusion boundary, it is represented as tri-map, where the white, black, and gray pixels signify the foreground, background, and not-interested region. To compare the proposed method and Lee and Park's method [27], the red boxes in

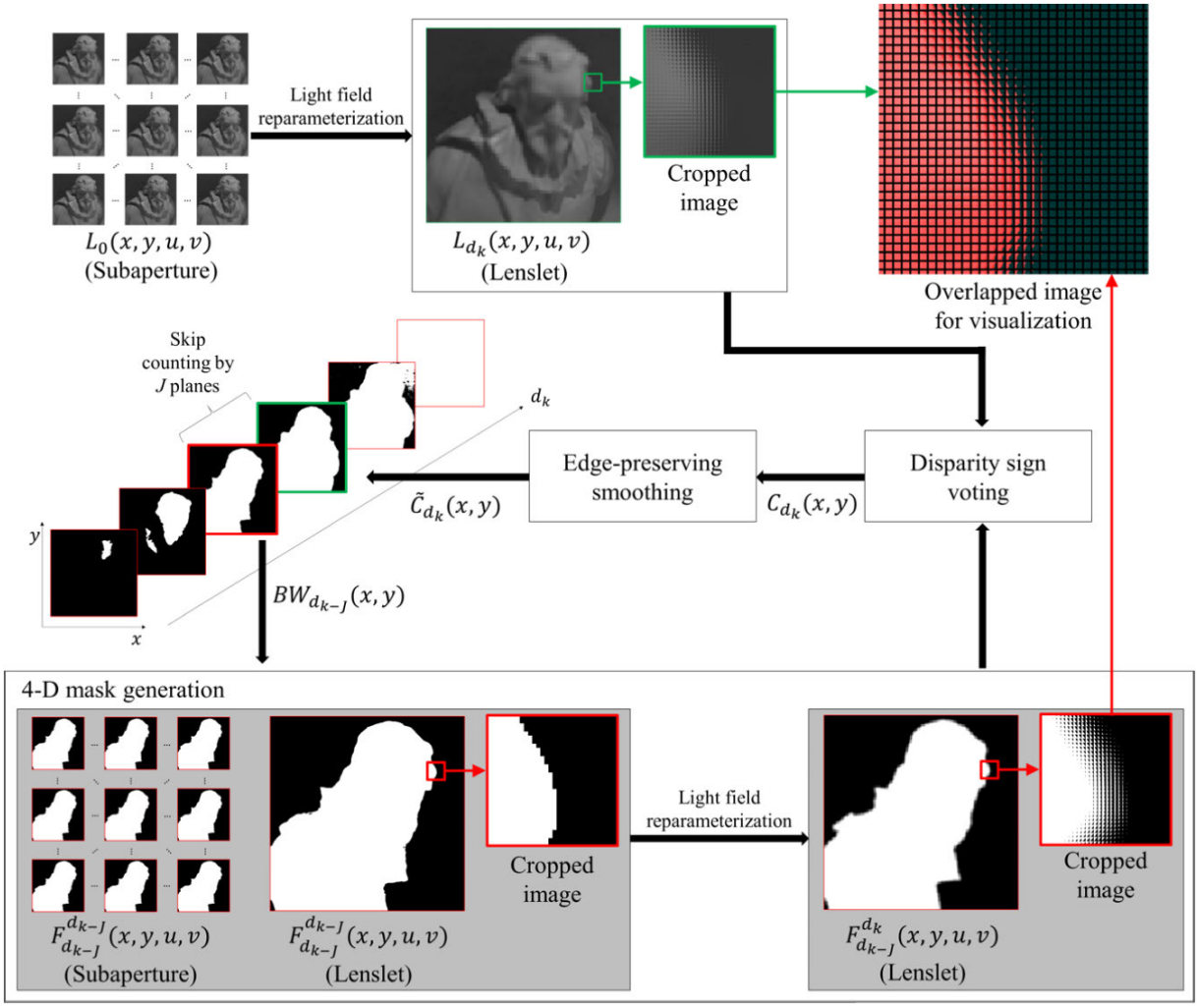

FIGURE 5. Intermediate results of 4-D mask generation by the proposed occlusion handling DFLF method and comparison to input LF image. In the top-right, a reparameterized LF image and foreground map are overlapped for visualization. The overlapped image shows that the foreground map generated by the proposed 4-D mask generation are well-aligned with the reparameterized LF image.

$\tilde{C}_{d_k}(x, y)$ thresholding by 0 are enlarged on the right side of each figure. The red pixels signify the errors compared to the GT. Compared to Lee and Park's method, the errors of the proposed method are reduced.

B. 4-D MASK GENERATION

With a given edge-preserving smoothed slice $\tilde{C}_{d_{k-J}}(x, y)$ of the 3-D volume, the 2-D foreground mask $BW_{d_{k-J}}(x, y)$ can be obtained by

$$BW_{d_{k-J}}(x, y) = \begin{cases} 1, & if \ \ \tilde{C}_{d_{k-J}}(x, y) > 0, \\ 0, & otherwise. \end{cases} \tag{7}$$

Algorithm 2 shows a pseudo code of the 4-D mask generation. To generate the 4-D foreground mask that is aligned with reparameterized LF image $L_{d_k}(x, y, u, v)$, the 4-D foreground mask is initialized with the 2-D foreground mask $BW_{d_{k-J}}(x, y)$ for all views. Since the 2-D mask $BW_{d_{k-J}}(x, y)$ is generated with the reparameterized LF image $L_{d_{k-J}}(x, y, u, v)$, all object on the disparity plane d_{k-J} falls into the same angular patch of $L_{d_{k-J}}(x, y, u, v)$ in view of lenslet image. Therefore, the initialized 4-D foreground mask $F_{d_{k-J}}^{d_{k-J}}(x, y, u, v)$ with $BW_{d_{k-J}}(x, y)$ is aligned with $L_{d_{k-J}}(x, y, u, v)$. To reparameterize the initialized 4-D foreground mask $F_{d_{k-J}}^{d_{k-J}}(x, y, u, v)$ to the disparity plane d_k, the amount of $d^{step} \times J$ needs to be reparameterized in view of the disparity plane d_{k-J}. By reparameterizing $F_{d_{k-J}}^{d_{k-J}}(x, y, u, v)$ to the amount of $d^{step} \times J$, $F_{d_{k-J}}^{d_k}(x, y, u, v)$ can be obtained. $F_{d_{k-J}}^{d_k}(x, y, u, v)$ is returned and used in the next step $k + J$.

Algorithm 2 4-D Mask Generation (*4DMaskGen*)

Require: $BW_{d_{k-J}}(x, y)$
1: **for** each (u, v) **do**
2: $\quad F_{d_{k-J}}^{d_{k-J}}(x, y, u, v) = BW_{d_{k-J}}(x, y)$
3: **end for**
4: $F_{d_{k-J}}^{d_k}(x, y, u, v) = LFreparam(F_{d_{k-J}}^{d_{k-J}}(x, y, u, v), d_k)$
5: **return** $F_{d_{k-J}}^{d_k}(x, y, u, v)$

Fig. 5 shows the intermediate results of 4-D mask generation by the proposed occlusion handling DFLF method and

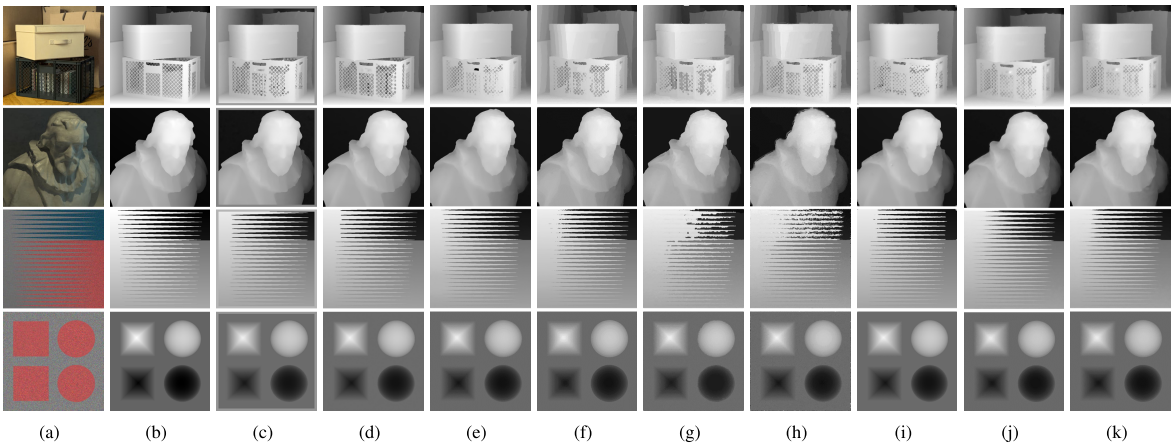

FIGURE 6. Comparison of resulting disparity maps of the existing methods and the proposed method on the synthetic dataset [48] (Boxes, Cotton, Backgammon, and Pyramids from the top to bottom rows). (a) Reference view images. (b) GT disparity maps. (c) EPINET [25]. (d) LFAttNet [26]. (e) SPO-MO [43]. (f) CAE [38]. (g) LF_OCC [39]. (h) OMG [40]. (i) OFSY [42]. (j) FBS-SFA [28]. (k) Ours.

comparison to the lenslet image of the input LF $L_0(x, y, u, v)$. As input images of the disparity sign voting, the reparameterized LF $L_{d_k}(x, y, u, v)$ and 4-D foreground mask $F^{d_k}_{d_{k-J}}(x, y, u, v)$ are used. An input LF image $L_0(x, y, u, v)$ is shown in form of the subaperture image in the top-left side. Using $L_0(x, y, u, v)$, $L_{d_k}(x, y, u, v)$ is computed as shown in the top-middle. At the bottom of the Fig. 5, $BW_{d_{k-J}}$, which is computed at $k - J$ plane, is used as an input of the 4-D mask generation. At every viewpoint (u, v), $F^{d_{k-J}}_{d_{k-J}}(x, y, u, v)$ is initialized with $BW_{d_{k-J}}$. In view of the lenslet image, each angular patch at (x, y) has the same value as shown in the cropped image. The 4-D foreground mask $F^{d_k}_{d_{k-J}}(x, y, u, v)$, which is used for the disparity sign voting at d_k plane, is generated by reparameterizing $F^{d_{k-J}}_{d_{k-J}}(x, y, u, v)$ from d_{k-J} to d_k plane. For visualization, in the top-right of Fig. 5, the cropped images of $L_{d_k}(x, y, u, v)$ and $F^{d_k}_{d_{k-J}}(x, y, u, v)$ are overlapped to qualitatively evaluate 4-D mask generation. With a three channeled color image representation, the cropped image of $F^{d_k}_{d_{k-J}}(x, y, u, v)$ is inserted into the red channel and that of $L_{d_k}(x, y, u, v)$ is inserted into the green and blue channels. Although $F^{d_k}_{d_{k-J}}(x, y, u, v)$ slightly over-covers $L_{d_k}(x, y, u, v)$, the overlapped image shows a reasonable result. As shown in Fig. 5, if the true disparity values at certain pixel (x, y) are known, it is possible to obtain the well-fitted foreground mask comparing to the reparameterized LF. However, the disparity values are unknown in the estimation step and the optimal J would be different for each scene and even for each region in the same scene. Thus, the skip counting parameter J is introduced for handling occlusion. Although the optimal J is not applied to every pixel, occlusion could be effectively handled by just introducing the skip counting parameter in the proposed method.

The proposed 4-D mask generation method can be utilized to not only the proposed DFLF method but also the other applications where masking operation is needed. For example, in LF matting [45]–[47], to further analyze light field image according to the disparity axis, there might be some

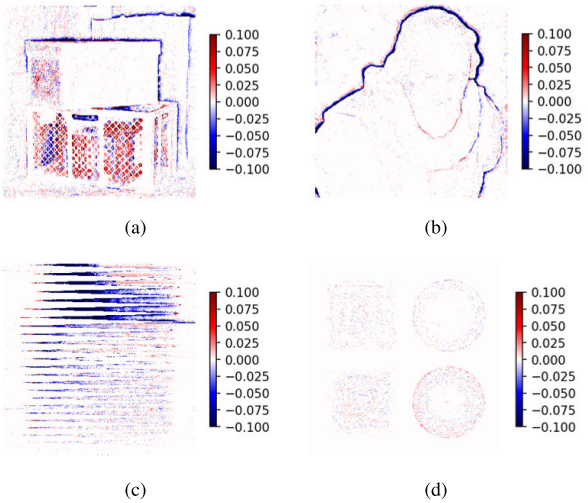

FIGURE 7. Comparison to FBS-SFA [28]. The blue pixels signify the region where the proposed method is better and red pixels signify the region where FBS-SFA is better. (a) Boxes. (b) Cotton. (c) Backgammon. (d) Pyramids.

cases that an obtained 4-D LF mask at certain disparity plane needs to be transferred to the other disparity plane. Although the existing light field matting methods show solutions for a single layer, independently of the framework of the proposed method, the 4-D mask generation might be usefully utilized in a scenario where an obtained mask needs to be transferred over disparity axis.

IV. EXPERIMENTAL RESULTS

With the synthetic image dataset [48], the proposed and existing methods are quantitatively and qualitatively evaluated. With the real image dataset [49], the proposed method and existing methods are qualitatively compared. In the synthetic dataset, each scene has $x \times y = 512 \times 512$ spatial resolution and $u \times v = 9 \times 9$ angular resolution. In the real dataset, each scene has $x \times y = 541 \times 375$ spatial resolution and $u \times v = 14 \times 14$ angular resolution. For experiments with the

TABLE 1. Performance evaluation on synthetic image dataset [48].

	Learning-based approaches		Conventional approaches						
Metrics	EPINET	LFAttNet	SPO-MO	CAE	LF_OCC	OMG	OFSY	FBS-SFA	Ours
BadPix(0.07)	**4.646**	**3.756**	**5.708**	8.211	17.579	15.403	12.036	12.039	10.739
MSE	**2.418**	**1.904**	3.518	3.730	6.690	6.586	7.030	3.591	**3.231**
Discontinuity	**18.855**	**14.938**	22.848	**21.750**	38.263	33.346	33.71	36.493	31.954
Runtime (second)	**10.659**	**5.862**	4304.333	832.081	10508.469	1163.938	**200.282**	259.2	560.896

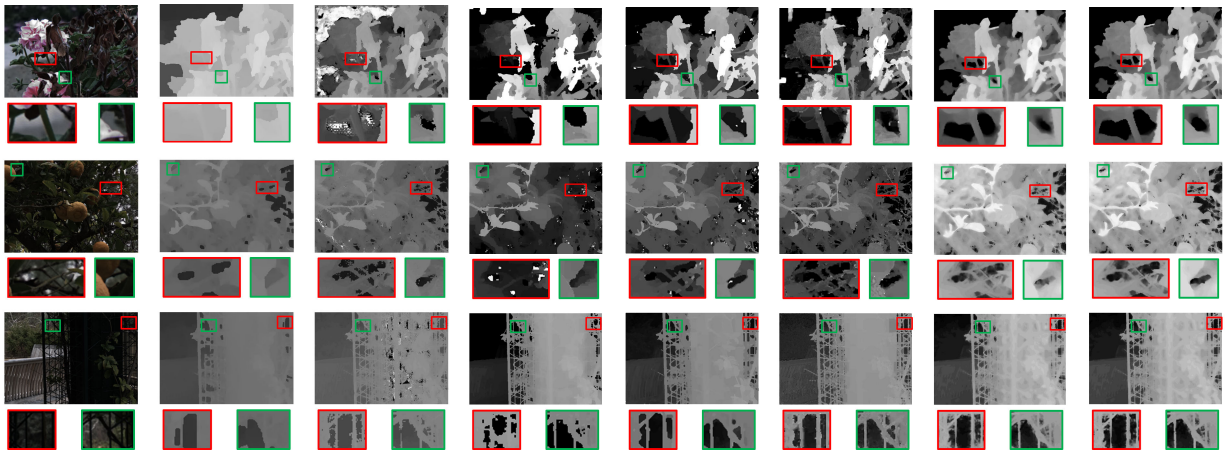

FIGURE 8. Comparison of resulting disparity maps of the existing methods and the proposed method in the real dataset [49]. (a) Reference view images. (b) Lytro [50]. (c) SPO [51]. (d) CAE [38], (e) LF_OCC [39], (f) OMG [40], (g) FBS-SFA [28], (h) Ours.

real dataset, angular resolution is cropped to $u \times v = 9 \times 9$ with respect to the center pixel. The proposed method is implemented in MATLAB on a PC with a 4.1 GHz CPU and 16 GB of memory. For fair comparison of the time complexity, although the proposed method could be implemented with parallel processing tool, the proposed method is implemented without parallel processing tool. Averaged runtimes of the proposed method for a single slice with the synthetic dataset and real dataset are 3.347 and 2.607 seconds, respectively. With the same environment, averaged run-times of FBS-SFA is measured as 2.110 and 1.592 seconds, respectively. The skip counting parameter J and d^{step} for the experiments on the synthetic images are set to 7 and 0.02, respectively. Those on the real images are set to 5 and 0.05, respectively. For runtimes, if the number of quantized disparity planes is 100 levels for an LF image, the proposed method approximately takes $3.347 \times 100 = 334.7$ seconds for computing a single disparity map from a given LF image.

In the website of 4-D LF benchmark [52], the proposed method is abbreviated as FBSOCC-SFA. The resulting disparity maps of the other scenes not presented in this paper can be found at the website [52] in detail.

A. SYNTHETIC IMAGES

Fig. 6 and Table 1 show the qualitative and quantitative evaluations on the synthetic dataset, respectively. As shown in Fig. 6, for synthetic dataset, the resulting disparity maps of FBSOCC-SFA is compared with a fully-convolutional neural network using epipolar geometry (EPINET) [25], attention-based view selection network (LFAttNet) [26], the spinning parallelogram operator using multi-orientation EPIs (SPO-MO) [43], constrained angular entropy cost (CAE) [38], occlusion-aware LF depth estimation (LF_OCC) [39], occlusion model guided antiocclusion (OMG) [40], occlusion-aware focal stack symmetry (OFSY) [42], and FBS-based method with the sign function approximation (FBS-SFA) [28]. Although the results of various methods are available on 4-D LF benchmark [52], the two methods among the end-to-end learning-based methods (EPINET and LFAttNet) and local methods handling occlusion among the conventional approaches, which are grouped into the EPI & cost-based (SPO-MO), angular patch & cost-based (CAE, LF_OCC, and OMG), focal stack & cost-based (OFSY), and angular patch & FBS-based (FBS-SFA) methods, are chosen for the comparison. Qualitatively, the proposed method shows reasonable performance compared to the existing methods.

In Table 1, the proposed method is compared to the existing methods in terms of bad pixel ratio (BadPix (0.07)), mean squared error (MSE), discontinuity, and averaged run-time for images. BadPix (0.07) measures the percentage of error pixels that are greater than 0.07 compared to the GT disparity map. Discontinuity measures errors around occlusion boundary with given occlusion boundary mask. The top three performances are bold-faced for each metric. Although the proposed method shows somewhat bad performance in terms of BadPix(0.07) and discontinuity, it shows a reasonable per-

formance in terms of MSE. Compared to the FBS-SFA [28], the proposed method shows better performance for three metrics. Averaged run-times of each algorithm for images are shown in the last row in Table 1.

Fig. 7 shows comparisons between FBS-SFA [28] and the proposed method. Fig. 7 is obtained by |FBSOCC-SFA − GT| − |FBS-SFA − GT|. Blue pixels signify that the FBSOCC-SFA is better whereas red pixels signify that the FBS-SFA is better. Figs. 7(a)-(c) show that the blue pixels are concentrated around relatively large occlusion boundary. If occlusion boundary is relatively small or occlusion boundary is not presented in a scene such as Fig. 7(d), there are cases that the red pixels are concentrated. However, overall quantitative performance of FBSOCC-SFA is improved as shown in Table 1. If the disparity values for each pixel are known, it is possible to reparameterize the foreground mask to true disparity plane. Then, errors like the red pixels could be also reduced. However, in this paper, our goal is to estimate the disparity map, which is unknown value. Thus, by simply introducing the skip counting parameter J, the occlusion boundary is effectively handled in the proposed method.

B. REAL IMAGES

In Fig. 8, for three scenes from real dataset [49], the resulting disparity maps of the proposed method is compared with Lytro [50], the spinning parallolgram operator (SPO) [51], CAE [38], LF_OCC [39], and OMG [40] whose codes are publicly available. FBS-SFA [28] is also compared with the proposed method for evaluating the improvement of performance. For each scene, red and green boxes are enlarged on the bottom side of each scene. Although the GT disparity maps are not provided, the proposed method shows the reasonable performance. Especially around occlusion boundary (the enlarged boxes of each scene), compared to the FBS-SFA [28], the resulting disparity maps of the proposed method have sharper boundaries than those of FBS-SFA. Note that some resulting disparity maps of the existing methods have less contrast due to spiky artifacts.

V. CONCLUSION

In this paper, an occlusion handling method for DFLF based on FBS is proposed. Based on the observation (foreground fattening), with the front-to-back scanning manner, the proposed method successively excludes the foreground information obtained from $(k - J)^{th}$ step when inferring foreground map of k^{th} step. For both synthetic and real datasets, the proposed method shows reasonable performance compared to the existing methods and better performance than the existing FBS-based methods. Recently, although the methods based on deep learning have shown better performance than those based on conventional approaches, we believe that the success of the methods based on deep learning is supported by sufficient researches of the conventional approaches. Based on the proposed method, future work will focus on combining the framework of the proposed method with deep learning techniques.

REFERENCES

[1] C. Chen, J. Wei, C. Peng, W. Zhang, and H. Qin, "Improved saliency detection in RGB-D images using two-phase depth estimation and selective deep fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4296–4307, 2020.
[2] X. Wang, S. Li, C. Chen, A. Hao, and H. Qin, "Depth quality-aware selective saliency fusion for RGB-D image salient object detection," *Neurocomputing*, vol. 432, pp. 44–56, Apr. 2021.
[3] X. Wang, S. Li, C. Chen, Y. Fang, A. Hao, and H. Qin, "Data-level recombination and lightweight fusion scheme for RGB-D salient object detection," *IEEE Trans. Image Process.*, vol. 30, pp. 458–471, 2021.
[4] C. Chen, J. Wei, C. Peng, and H. Qin, "Depth-quality-aware salient object detection," *IEEE Trans. Image Process.*, vol. 30, pp. 2350–2363, 2021.
[5] G. Borghi, M. Fabbri, R. Vezzani, S. Calderara, and R. Cucchiara, "Face-from-depth for head pose estimation on depth images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 596–609, Mar. 2020.
[6] J.-H. Song and S.-J. Kang, "Fast 3D hand pose estimation for real-time system," in *Proc. Int. SoC Design Conf. (ISOCC)*, Oct. 2020, pp. 121–122.
[7] L. Chen, S.-Y. Lin, Y. Xie, Y.-Y. Lin, W. Fan, and X. Xie, "DGGAN: Depth-image guided generative adversarial networks for disentangling RGB and depth images in 3D hand pose estimation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 400–408.
[8] S. Zhu, G. Brazil, and X. Liu, "The edge of depth: Explicit constraints between segmentation and depth," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13113–13122.
[9] Z. Gu, L. Niu, H. Zhao, and L. Zhang, "Hard pixel mining for depth privileged semantic segmentation," *IEEE Trans. Multimedia*, early access, Nov. 2, 2020, doi: 10.1109/TMM.2020.3035231.
[10] C. Li, J. Ku, and S. L. Waslander, "Confidence guided stereo 3D object detection with split depth estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 5776–5783.
[11] M. Ding, Y. Huo, H. Yi, Z. Wang, J. Shi, Z. Lu, and P. Luo, "Learning depth-guided convolutions for monocular 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11669–11678.
[12] A. Bhandari and R. Raskar, "Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-D imaging," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 45–58, Sep. 2016.
[13] R. Ng, "Digital light field photography," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2006.
[14] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
[15] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.
[16] S. Lee, J. Lee, B. Kim, E. Yi, and J. Kim, "Patch-wise attention network for monocular depth estimation," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 3, pp. 1873–1881.
[17] D. Kim, S. Lee, J. Lee, and J. Kim, "Leveraging contextual information for monocular depth estimation," *IEEE Access*, vol. 8, pp. 147808–147817, 2020.
[18] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.
[19] O. Johannsen *et al.*, "A taxonomy and evaluation of dense light field depth estimation algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1795–1812.
[20] J. Y. Lee and R.-H. Park, "Complex-valued disparity: Unified depth model of depth from stereo, depth from focus, and depth from defocus based on the light field gradient," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 830–841, Mar. 2021.
[21] J. Zbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1592–1599.
[22] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2758–2766.
[23] F. Zhang, V. Prisacariu, R. Yang, and P. H. S. Torr, "GA-Net: Guided aggregation net for end-to-end stereo matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 185–194.
[24] H. Xu and J. Zhang, "AANet: Adaptive aggregation network for efficient stereo matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1956–1965.

[25] C. Shin, H.-G. Jeon, Y. Yoon, I. S. Kweon, and S. J. Kim, "EPINET: A fully-convolutional neural network using epipolar geometry for depth from light field images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4748–4757.
[26] Y.-J. Tsai, Y.-L. Liu, M. Ouhyoung, and Y.-Y. Chuang, "Attention-based view selection networks for light-field disparity estimation," in *Proc. 34th Conf. Artif. Intell. (AAAI)*, 2020, pp. 12095–12103.
[27] J. Y. Lee and R.-H. Park, "Depth estimation from light field by accumulating binary maps based on foreground–background separation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 955–964, Oct. 2017.
[28] J. Y. Lee and R.-H. Park, "Reduction of aliasing artifacts by sign function approximation in light field depth estimation based on foreground–background separation," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1750–1754, Nov. 2018.
[29] M. Levoy, "Light fields and computational imaging," *Computer*, vol. 39, no. 8, pp. 46–55, Aug. 2006.
[30] J. Y. Lee and R.-H. Park, "Separation of foreground and background from light field using gradient information," *OSA Appl. Opt.*, vol. 56, no. 4, pp. 1069–1078, 2017.
[31] A. Badki, A. Troccoli, K. Kim, J. Kautz, P. Sen, and O. Gallo, "Bi3D: Stereo depth estimation via binary classifications," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1597–1605.
[32] X. Ye, J. Li, H. Wang, H. Huang, and X. Zhang, "Efficient stereo matching leveraging deep local and context information," *IEEE Access*, vol. 5, pp. 18745–18755, 2017.
[33] M. Ramamonjisoa, Y. Du, and V. Lepetit, "Predicting sharp and accurate occlusion boundaries in monocular depth estimation using displacement fields," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14636–14645.
[34] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Computer Vision—ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 173–190.
[35] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[36] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 504–511, Feb. 2013.
[37] 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.*, Jun. 2014, pp. 1518–1525.
[38] I. K. Park and K. M. Lee, "Robust light field depth estimation using occlusion-noise aware data costs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2484–2497, Oct. 2018.
[39] T.-C. Wang, A. 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.
[40] H. Zhu, Q. Wang, and J. Yu, "Occlusion-model guided antiocclusion depth estimation in light field," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 965–978, Oct. 2017.
[41] W. Ai, S. Xiang, and L. Yu, "Robust depth estimation for multi-occlusion in light-field images," *Opt. Exp.*, vol. 27, no. 17, pp. 24793–24807, Aug. 2019.
[42] 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)*, Jul. 2017, pp. 2529–2537.
[43] H. Sheng, P. Zhao, S. Zhang, J. Zhang, and D. Yang, "Occlusion-aware depth estimation for light field using multi-orientation EPIs," *Pattern Recognit.*, vol. 74, pp. 587–599, Feb. 2018.
[44] H. Schilling, M. Diebold, C. Rother, and B. Jahne, "Trust your model: Light field depth estimation with inline occlusion handling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4530–4538.
[45] D. Cho, S. Kim, and Y.-W. Tai, "Consistent matting for light field images," in *Computer Vision—ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 90–104.
[46] J. Fiss, B. Curless, and R. Szeliski, "Light field layer matting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 623–631.
[47] D. Cho, S. Kim, Y.-W. Tai, and I. S. Kweon, "Automatic trimap generation and consistent matting for light-field images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1504–1517, Aug. 2017.
[48] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Computer Vision—ACCV 2016*, S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds. Cham, Switzerland: Springer, 2017, pp. 19–34.
[49] A. S. Raj, M. Lowney, and R. Shah. (Nov. 2020). *Light-Field Database Creation and Depth Estimation*. [Online]. Available: http://lightfields.stanford.edu/
[50] Lytro. (Nov. 2020). *The Lytro Camera*. [Online]. Available: http://lytro.com
[51] 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.
[52] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke. (Nov. 2020). *4-D Light Field Benchmark*. [Online]. Available: http://lightfield-analysis.net

JAE YOUNG LEE received the B.S. and M.S. degrees from the Department of Electronic Engineering, Sogang University, Seoul, South Korea, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with KAIST. His current research interests include image processing, computer vision, computational photography, machine learning, and deep learning.

RAE-HONG PARK (Life Senior Member, IEEE) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, South Korea, in 1976 and 1979, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1981 and 1984, respectively.

In 1984, he joined the Faculty of the Department of Electronic Engineering, Sogang University, Seoul. In 1990, he spent his sabbatical year as a Visiting Associate Professor with the Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park. In 2001 and 2004, he spent sabbatical semesters with the Digital Media Research and Development Center (DTV image/video enhancement), Samsung Electronics Company, Ltd., Suwon, South Korea. In 2012, he spent his sabbatical year with the Digital Imaging Business (Research and Development Team) and Visual Display Business (Research and Development Office), Samsung Electronics Company, Ltd. He is currently an Emeritus Professor with the Department of Electronic Engineering and a Distinguished Professor with the ICT Convergence Disaster/Safety Research Institute, Sogang University. His current research interests include computer vision, pattern recognition, and video communication. He was a recipient of the 1987 Academic Award presented by the Korea Institute of Telematics and Electronics (KITE), the 1990 Postdoctoral Fellowship presented by the Korea Science and Engineering Foundation (KOSEF), the 1997 First Sogang Academic Award, the 1999 Professor Achievement Excellence Award presented by Sogang University, and the 2000 Haedong Paper Award presented by the Institute of Electronics Engineers of Korea (IEEK). He is also a co-recipient of the Best Student Paper Award of the IEEE International Symposium Multimedia (ISM 2006) and the IEEE International Symposium Consumer Electronics (ISCE 2011). From 1995 to 1996, he served as an Editor for *Journal of Electronics Engineering* (KITE).

JUNMO KIM (Member, IEEE) received the B.S. degree from Seoul National University, Seoul, South Korea, in 1998, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 2000 and 2005, respectively. From 2005 to 2009, he was with the Samsung Advanced Institute of Technology (SAIT), South Korea, as a Research Staff Member. He joined the Faculty of KAIST, in 2009, where he is currently an Associate Professor in electrical engineering. His research interests include image processing, computer vision, statistical signal processing, machine learning, and information theory.