Efficient Depth Estimation for Light Field Images

Wisarut Chantara, Ji-Hun Mun, and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Republic of Korea
E-mail: {wisarut, jhm, hoyo}@gist.ac.kr

***Abstract*— Light field images consist of a set of sub-aperture images with slight shifts in their positions. They can be used to obtain the depth map without camera calibration and image rectification. However, since the baseline between sub-aperture images is very narrow, it limits the depth estimation range and accuracy. In this paper, we propose a depth estimation method to solve such problems. After we analyze the sub-aperture images in the form of epipolar plane images (EPIs), we modify the edge confidence measure to test which part of EPIs is good for depth estimation. Then, we apply the modified structure tensor (MST) to compute depth and corresponding information. Finally, we reconstruct a depth map from light field images.**

I. INTRODUCTION

Stereo matching is a very popular image processing algorithm for depth estimation from stereoscopic images, and several variants are available. However, it requires preprocessing operations, such as camera calibration, image rectification, and color correction. Since stereo matching is a pattern matching technique, it requires a heavy computation to find the best matching position. Furthermore, it is quite difficult to obtain the correct correspondence between the left and right images in occlusion regions. In this paper, thus, we try to use light field images to estimate the depth map in different ways.

In recent years, the light field camera, also called as the plenoptic camera, has been used for several applications, such as depth map estimation, image compression, and 3D scene reconstruction. It is constructed with an internal microlens array to capture the light field information and generate a set of multiple view images, which is called as the sub-aperture images. Each sub-aperture image is created by gathering pixel values from all the micro images that are located in the same coordinate of the 3D scene. Each micro image is the image that was formed on the photosensor by the microlens array.

However, since each sub-aperture image from the light field camera has a very narrow baseline, the conventional stereo matching method does not provide an accurate depth map. In this paper, we propose a new method for depth map estimation from the light field images.

Until now, several prior papers have addressed the bottleneck of the light field camera, namely the resolution problem. Besides of super-resolution, depth estimation has also been investigated as an application of light field images. Yu et al. proposed an analysis of color demosaicing in the plenoptic camera, which performed the demosaicing on the raw image [1]. Their method transformed the radiance to the desired focal plane, and then applied the frequency domain plenoptic resampling. A color-filtered image of full resolution was created by performing a 2D integral projection from the reparameterized radiance. Then, it conducted demosaicing to obtain the color result.

Yu et al. introduced a line assisted light field triangulation and stereo matching [2]. Their algorithm applied a line-assisted graph cut method where the line segments with known disparities were used as hard constraints in the graph cut algorithm. It analyzed the 3D geometry of lines in the light field image and computed the disparity map through line matching between the sub-aperture images.

Bishop and Favaro presented a formal methodology for the restoration of high-resolution images from light field data [3]. Their method used an image formation model. The model incorporated Lambertian and texture preserving priors to reconstruct both scene depth and its super-resolved texture in a Bayesian framework. It eliminated aliasing by fusing multi-view information.

Lastly, Tao et al. proposed an interesting approach that combines defocus and correspondence to estimate the scene depth [4]. After processing the initial estimation, they applied a multi-label optimization to refine the estimated disparity map.

In this paper, we describe a depth estimation method that integrates the modified edge confidence measure and the modified structure tensor to reconstruct a depth map from light field images.

II. PROPOSED METHOD

A. Light field EPIs generation and depth information

We decide to use Lytro camera [5] as a device to generate the image datasets. A scene captured with a different angle of viewpoint is called sub-aperture images in the light field camera system, as shown in Fig. 1.

For representing the light field data, we exploit the epipolar plane images (EPIs), as shown in Fig. 2. We get 2D (x, s) and 2D (y, t) slices of a light filed by fixing y-t and x-s respectively. It is called the horizontal EPI ($EPI(x, s)$) and the vertical EPI ($EPI(y, t)$).

EPIs is introduced by Bolles et al. in [6]. It is also shown in their work that any point in the space can be projected to a line on the EPIs. Additionally, the slope of the line is related to the depth of the corresponding point in the space [6], as shown in Fig. 3. Therefore, the depth information can be obtained by estimating the slope of lines in EPIs [7].

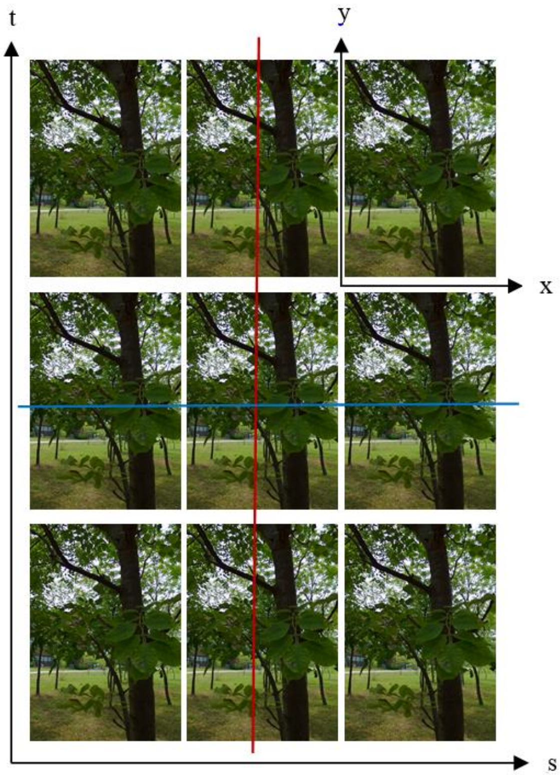

Fig. 1 Sub-aperture images

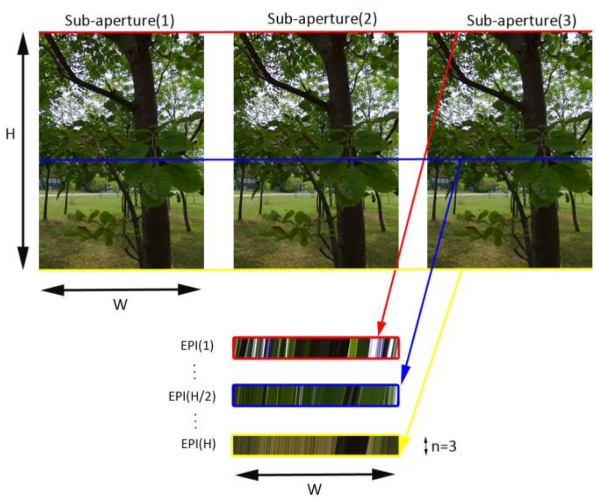

Fig. 2 An illustration of EPIs

Fig. 3 Depth estimation from EPIs

B. Modified edge confidence

We modify edge confidence [8] that is intended to be a fast test for which parts of the EPIs depth estimation seem promising, the measure is defined by:

$$C_{E(x,s)} = \sum_{x' \in N_h(x,y,s,t)} ||L(x,y,s,t) - L(x',y,s,t)|| + \sum_{y' \in N_v(x,y,s,t)} ||L(x,y,s,t) - L(x,y',s,t)|| \tag{1}$$

where $C_{E(x,\ s)}$ is the confidence at position *(x, s)* from the *EPI(x, s)*, $N_h(x, y, s, t)$ is 1D horizontal window centered at position *(x, y)* from viewpoint *(s, t)* and $N_v(x, y, s, t)$ is 1D vertical window centered at position *(x, y)* from viewpoint *(s, t)*. *L(x, y, s, t)* represents intensity value in the image plane position *(x, y)* through viewpoint *(s, t)*. We choose the window size to be 9. $C_{E(x,\ s)}$ is the threshold (with a value of 0.5), resulting in a binary confidence mask $M_{E(x,\ s)}$. We apply an opening morphological operator to the mask for removing faked isolated regions. It can be applied to the *EPI(y, t)* to get $C_{E(y,\ t)}$ and $M_{E(y,\ t)}$ information as well.

$$M_{E(x,s)} = \begin{cases} 1, & C_{E(x,s)} > threshold \\ 0, & otherwise \end{cases} \tag{2}$$

C. Modified structure tensor (MST)

The structure tensor [9] produces orientation estimations and the certainty of each estimation by using the gradient information of an image. The structure tensor is expressed by:

$$J = \begin{bmatrix} G_\sigma * (E_x E_x) & G_\sigma * (E_x E_y) \\ G_\sigma * (E_x E_y) & G_\sigma * (E_y E_y) \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix} \tag{3}$$

where G_σ: Gaussian smoothing kernel with standard deviation σ. E_x, E_y: the horizontal and vertical component of the gradient vector at each pixel in EPIs E respectively. ($E_x = \partial E/\partial x$, $E_y = \partial E/\partial y$)

We improve this kind of computing method by modifying the computing formulations of the structure tensor. Since structure tensor J is symmetric, positive semidefinite, it has two orthogonal eigenvectors (V) and the correspondence of eigenvalues (v) as follows:

$$V = \begin{pmatrix} J_{yy} - J_{xx} + \sqrt{(J_{yy} - J_{xx})^2 + 4J_{xy}^2} \\ -2J_{xy} \end{pmatrix} \tag{4}$$

$$V^{\perp} = \begin{pmatrix} 2J_{xy} \\ J_{yy} - J_{xx} + \sqrt{(J_{yy} - J_{xx})^2 + 4J_{xy}^2} \end{pmatrix}$$

$$v = \frac{1}{2}\left(J_{yy} + J_{xx} - \sqrt{(J_{yy} - J_{xx})^2 + 4J_{xy}^2}\right)$$
$$v^{\perp} = \frac{1}{2}\left(J_{yy} + J_{xx} + \sqrt{\left(J_{yy} - J_{xx}\right)^2 + 4J_{xy}^2}\right) \quad (5)$$

The relationship of the eigenvectors with EPIs is shown in Fig. 4. We can compute the slope of EPIs line by using the intersection angle of two eigenvectors with the x-position axis. Moreover, local structures can be determined as edges ($v^{\perp} \gg v \approx 0$) based on the two eigenvalues.

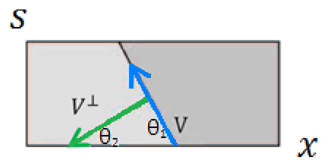

Fig. 4 The relationship of the eigenvectors with the EPIs

We can compute the depth information and the relationship of eigenvectors with the EPIs by:

$$d = -f\frac{\Delta s}{\Delta x} = -f \cdot \tan(\theta_1) \quad (6)$$

where f is a constant value referring to the distance between parallel planes. The reliability measure is applied from the correspondence of the structure tensor J [9]. The reliability measure is defined by:

$$r = \frac{(J_{yy} - J_{xx})^2 + 4(J_{xy})^2}{(J_{xx} + J_{yy})^2} \quad (7)$$

D. Depth estimation reconstruction

After calculating the depth information d ($d_{(x,s)}$, $d_{(y,t)}$), reliability measure r ($r_{(x,s)}$, $r_{(y,t)}$), and confidence mask M_E ($M_{E(x,s)}$, $M_{E(y,t)}$) from EPIs ($EPI(x, s)$, $EPI(y, t)$), we construct the final depth d_{final} by superimposing all depth information d based on the corresponding reliability measure and the confidence mask. We select the depth information that corresponds to higher reliability measure with $M_E = 1$; if $M_E = 0$, we directly choose the depth information from higher reliability measure. So the high credibility depth map is reconstructed. The merging process can be described by:

$$d_{final} = \begin{cases} d_{(x,s)}, & M_{E(x,s)} = 1 \text{ and } r_{(x,s)} > r_{(y,t)} \\ d_{(y,t)}, & M_{E(y,t)} = 1 \text{ and } r_{(y,t)} > r_{(x,s)} \\ d_{M_E=0} = \begin{cases} d_{(x,s)}, & r_{(x,s)} > r_{(y,t)} \\ d_{(y,t)}, & r_{(y,t)} > r_{(x,s)} \end{cases}, & M_E = 0 \end{cases} \quad (8)$$

III. Experiment Results

In this paper, the proposed method conducts the experiment using test images which are captured by Lytro light field camera [5]. Because Lytro light field camera does not have ground truth. Then, we compare the results visually.

When we compare the proposed results to Lytro software [5], the proposed method can preserve the details of depth information better than other methods, as shown in Fig. 5.

IV. Conclusions

In this paper, we proposed the efficient depth estimation method for light field images. The proposed method generates the depth map using only the sub-aperture images and estimates depth map on EPIs, The method obtains depth information from the modified structure tensor and reliability measure. The modified edge confidence is applied for testing which parts of the EPIs depth estimation seem promising. It is used to prevent the computation of depth estimate at ambiguous EPI-pixels for a final depth map. As a result, our method generated improved the depth map compared with other methods.

Acknowledgment

This work was supported by the 'Civil-Military Technology Cooperation Program' grant funded by the Korea government.

References

[1] Z. Yu, J. Yu, A. Lumsdaine, and T. Georgiev, "An analysis of color demosaicing in plenoptic cameras," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 901-908, 2012.
[2] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu, "Line assisted light field triangulation and stereo matching," in IEEE Conference on Computer Vision (ICCV), pp. 2792-2799, 2013.
[3] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *TPAMI*, vol. 34(5), pp. 972-986, 2012.
[4] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in IEEE Conference on Computer Vision (ICCV), pp. 673-680, 2013.
[5] Lytro, "The Lytro camera," http://www.lytro.com.
[6] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1(1), pp. 7-55, 1987.
[7] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4d light fields," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 41-48, 2012.
[8] P. H. Lin, J. S. Yeh, F.C. Wu, and Y. Y. Chuang, "Depth estimation for lytro images by adaptive window matching on EPI," *J. Imaging*, pp. 1-15, 2017.
[9] J. Bigun and G.H. Granlund, "Optimal orientation detection of linear symmetry," *in First International Conference on Computer Vision (ICCV)*, pp. 433-438, 1987.

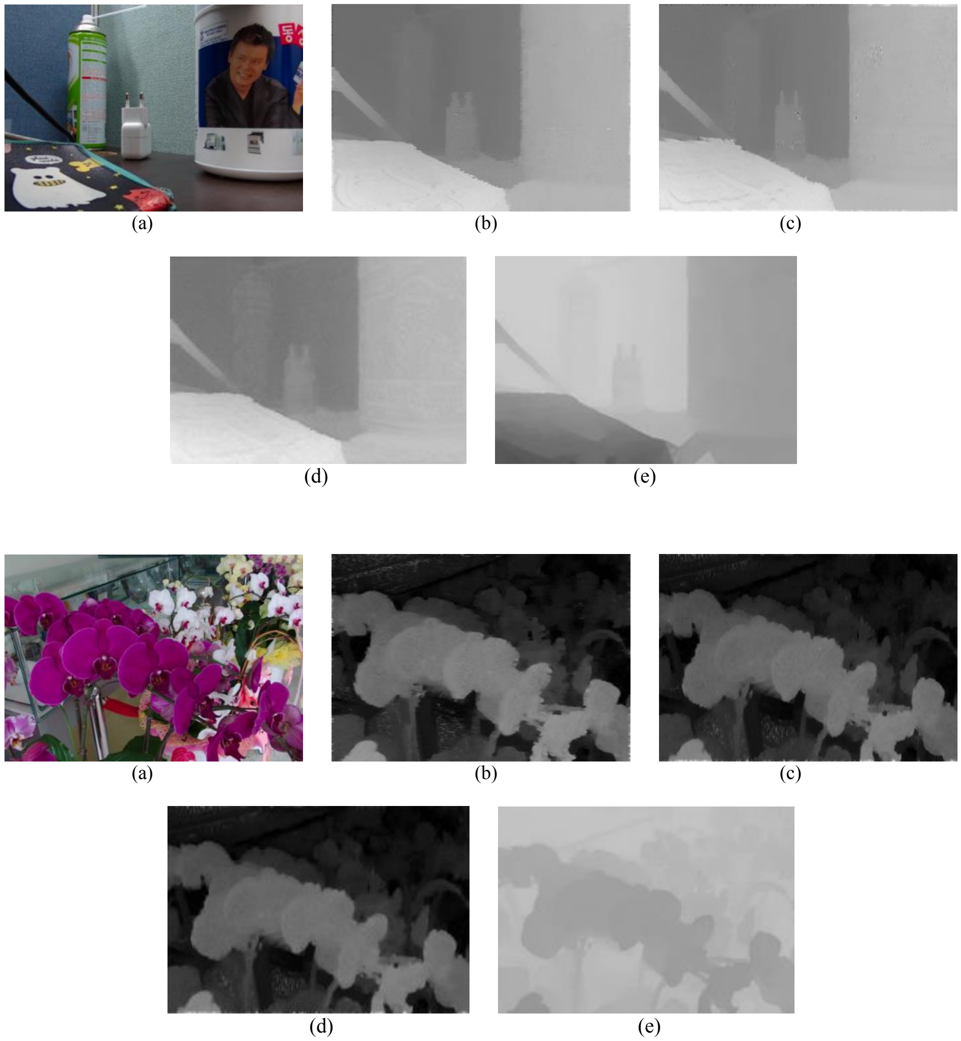

Fig. 5 Results of depth map: (a) center view sub-aperture image, (b) depth map in horizontal direction, (c) depth map in vertical direction, (d) proposed method, (e) Lytro software [5]