Initial Depth Estimation using Adaptive Window Size with Light Field Image

Wisarut Chantara and Yo-Sung Ho
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology
123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Republic of Korea
{wisarut, hoyo}@gist.ac.kr

***Abstract*—Light field image generates a set of focusing images at different depth levels, which suggests that one can determine object, foreground, or background regions. Moreover, the image can provide another set of multiple view images with slight shifts (sub-aperture images) which allows disparity computation without the image calibration and rectification steps. Nevertheless, the baseline between sub-aperture images is narrow. It limits the depth estimation range and accuracy. To solve this problem, we present an initial depth estimation method using adaptive window matching on epipolar plane images (EPIs) and structure tensor, which appropriately applied to the sub-aperture light field images. We analyze the sub-aperture images to EPIs and compute depth and corresponding information using the structure tensor on EPIs. An adaptive window size is used to improve the robustness of corresponding information. Finally, an initial depth map of light field image is reconstructed.**

Keywords*— *Light field image; sub-aperture images; epipolar plane images; structure tensor; adaptive window size

I. Introduction

Reconstructing depth map from a stereo pair of images, also known as stereo matching, is one of the key problems in computer vision. A large number of algorithms have been proposed to solve this problem. While the studies on depth estimation from light fields have just started, and most of them are focused on certain plenoptic camera [1]. A light field camera also called a plenoptic camera has been recently shown to be very effective in applications such as digital refocusing, depth estimation and 3D reconstruction. It is constructed with an internal microlens array to capture light field information that one can generate a set of multiple view images call sub-aperture images after the acquisition. A sub-aperture image is created by gathering a pixel from each micro-image, which located in the same relative coordinate with respect to the micro-image. The micro-image is the image which formed under microlens and on the photosensor. Because the baseline between sub-aperture images from a light field camera is narrow, how to estimate disparity map of the light field has become an attractive problem strongly affecting its application.

II. Proposed Method

Due to the narrow baseline of each sub-aperture image, a conventionally used stereo matching method does not generate an accurate depth map. We adopt an epipolar plane image [2] based depth map generation method. The flow of the proposed method in this paper is summarized as shown in Fig.2.

A. Two-plane parameterization

We adopt the two-plane parameterization [3] of the 4D light field and denote a light field as *L(x, y, s, t)* as shown in Fig. 1., where *(s, t)* can be seen as the index of different views and *(x, y)* are spatial coordinates with each view (see Fig.2. sub-aperture images).

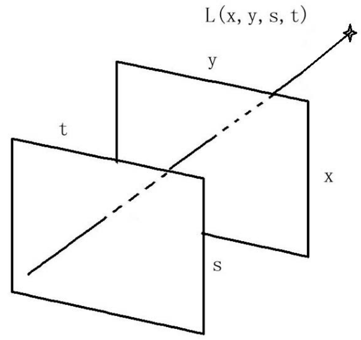

Fig. 1. Two-plane parameterization

B. Epipolar plane images (EPIs)

By fixing y and t, we can obtain a 2D *(x, s)* slice of a light field, as shown in Fig.2. (blue arrow). Similarly, 2D *(y, t)* slices can be obtained if x and s are fixed. These 2D slices are called epipolar plane images (EPIs). We call EPIs with fixed y and t the horizontal EPI (*EPI(x, s)*), and the one with fixed s and x the vertical EPI (*EPI(y, t)*).

Epipolar-plane image is introduced by Bolles et al. in [4]. 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 [4], which is written as:

$$\Delta x = -\frac{f}{Z}\Delta s \Rightarrow Z = -f\frac{\Delta s}{\Delta x} \quad (1)$$

978-1-5386-2615-3/18/$31.00 ©2018 IEEE

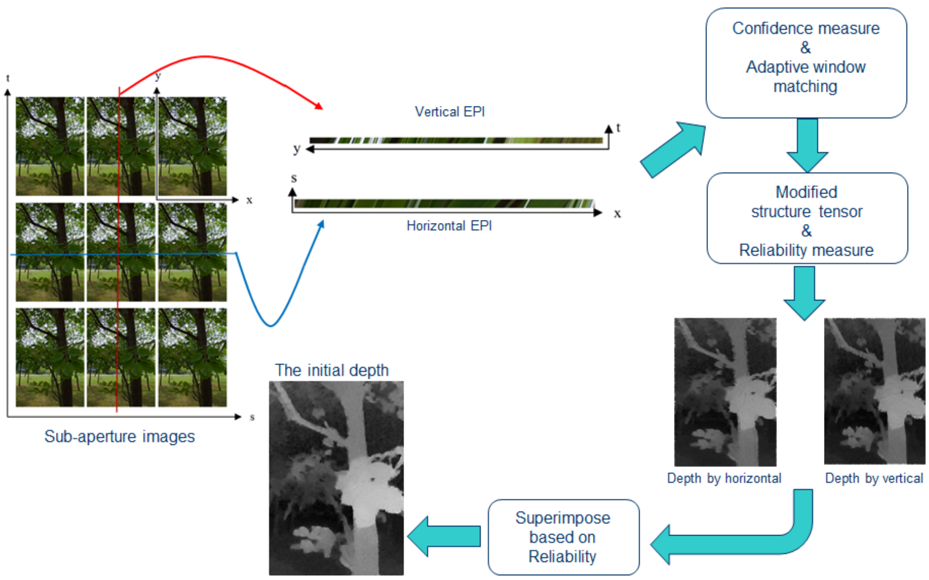

Fig. 2. The flowchart of proposed method

where Δs: distance between the sub-aperture image, Δx: distance between scene points, f: distance between parallel planes (focal length), $\Delta s/\Delta x$: The slope of a line in EPIs which indicates the real metric depth value Z

C. Confidence measure and adaptive window size

We modify edge confidence [5, 6] for a fast test for which parts of the EPIs a depth estimation seems promising, the measure of it is:

$$C_{E(x,s)} = \sum_{x' \in N_h(x,y,s,t)} ||L(x,y,s,t) - L(x',y,s,t)|| + \sum_{y' \in N_v(x,y,s,t)} ||L(x,y,s,t) - L(x,y',s,t)|| \quad (2)$$

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 view point (s, t). We choose the window size to be 9. It can be applied to the $EPI(y, t)$ as well. The window size is determined by its confidence. For a pixel with higher confidence, a smaller window size is needed for estimating depth. In other words, the pixel has lower confidence, the bigger window is required.

$$W_{size} = \begin{cases} 3 & if\ C_E \geq 0.8 \\ 7 & if\ 0.8 > C_E \geq 0.5 \\ 9 & if\ 0.5 > C_E \geq 0.2 \\ 11 & if\ 0.2 > C_E \end{cases} \quad (3)$$

D. Modified structure tensor

The structure tensor [7] uses the gradient information of an image in order to determine the orientation information of the edges and corner. The structure tensor is defined as:

$$J = \begin{bmatrix} G_\sigma * (E_x E_x) & G_\sigma * (E_x E_y) \\ G_\sigma * (E_x E_y) & G_\sigma * (E_y E_y) \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix} \quad (4)$$

where G_σ: Gaussian smoothing kernel with standard deviation σ . E_x, E_y: Horizontal and vertical component of the gradient vector at each pixel in EPIs E respectively. The window size of the kernel will be followed by the confidence measure from the previous section.

We improved this kind of computing method by modifying the computing formulations of the structure tensor. Since structure tensor J is symmetric and positive semi-definite, it has two orthogonal eigenvectors as follows:

$$V = \begin{pmatrix} J_{yy} - J_{xx} + \sqrt{(J_{yy} - J_{xx})^2 + 4J_{xy}^2} \\ -2J_{xy} \end{pmatrix} \quad (5)$$

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

The relationship of the eigenvectors with the EPIs is shown in Fig. 3. We can compute the slope of the EPIs line by using the intersection angle of two eigenvectors with x-position axis.

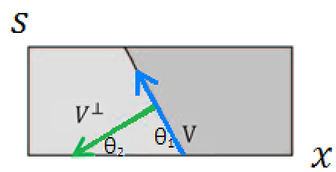

Fig. 3. The EPIs relationship

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

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

We use the correspondence of the structure tensor J as the reliability measure [7].

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

E. Initial depth estimation

Using the depth information d and reliability measure r for all the EPIs (*EPI(x, s), EPI(y, t)*), we superimpose all depth d based on the corresponding reliability measure by selecting the depth information that corresponds to higher reliability measure and reconstructs the initial depth map based on these depth d. So the new depth map has a higher credibility than the original one.

III. EXPERIMENT RESULTS

The proposed method is implemented by using test images which capture by Lytro light field camera [8]. Because the Lytro light field camera does not have ground truth information, so we compare the results visually.

When we compare the proposed results to the single direction depth map image (horizontal or vertical direction), the proposed method can preserve details of the initial depth information better than other methods as shown in Fig. 4.

IV. CONCLUSION

In this paper, we proposed an initial depth estimation method using adaptive window size with light field image. The proposed method generated the initial depth map using only the sub-aperture images and estimates depth map on EPIs, which are obtained depth information from the modified structure tensor and reliability measure. An adaptive window size is used to improve the robustness and avoids incorrect estimation in the EPIs. As a result, our method generated improved the initial disparity map compared with other methods. Our future work will enhance and make a refinement on the final depth map.

ACKNOWLEDGMENT

This work was supported by 'The Cross-Ministry Giga KOREA Project' grant funded by the Korea government (MSIT) (GK17C0100, Development of Interactive and Realistic Massive Giga- Content Technology).

REFERENCES

[1] T. Bishop and P. Favaro, “Full-resolution depth map estimation from an aliased plenoptic light field,” Computer Vision–ACCV, pp. 186–200, 2011.

[2] 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.

[3] M. Levoy and P. Hanrahan, “Light field rendering,” Proceeddddddings of the 23rd annual conference on Computer graphics and interactive techniques. ACM, pp. 31-42, 1996.

[4] 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.

[5] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene reconstruction from high spatio-Angular resolution light fields," ACM Trans. Graph, vol.32, no. 4, pp. 73:1-73:12, July 2013.

[6] 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, May 2017.

[7] J. Bigun and G.H. Granlund, “Optimal orientation detection of linear symmetry,” First International Conference on Computer Vision (ICCV), pp. 433-438, 1987.

[8] Lytro, "The Lytro camera," http://www.lytro.com.

(a) (b) (c) (d)

Fig. 4. Results of initial disparity estimation: (a) center view sub-aperture image, (b) initial depth estimate in horizontal directions, (c) initial depth estimate in vertical directions, (d) proposed method