Initial Depth Estimation using EPIs and Structure Tensor

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

***Abstract*— Light field cameras are enlarged in the field of computational photography because of the increased information they capture compared to traditional cameras. Moreover, the estimated depth is widely used for photographic editing. While, the intrinsic information of the light field has the advantage to allow disparity computation without the image calibration and rectification steps required in classic binocular stereo or multi-view algorithms. In this paper, we propose an initial depth estimation method using epipolar plane images (EPIs) and structure tensor, which appropriately applicable to the sub-aperture light field images. First, the sub-aperture images are displaced using raw light field image data. We analyze the sub-aperture images to EPIs and compute depth and corresponding information using the structure tensor on EPIs. Finally, the experiment results show that an initial disparity map of light field image is reconstructed.**

I. INTRODUCTION

The light field is a vector function that describes the amount of light propagating in every direction through every point in space [1]. The light field imaging has been recently introduced to the mass market by the handheld plenoptic camera, like Lytro [2], and Raytrix [3]. Techniques have been developed to extract the depth information and to reconstruct images with viewpoint changes at different focus planes using a single shot of the scene. Thus, light field acquisition and processing present high feasibility in the applications like digital refocusing, depth estimation, free-viewpoint image synthesis, 3D displays, and panorama image generation. Meanwhile, the problem of estimating a depth map from a light field camera is investigated. Different to conventional cameras, a light field camera captures not only a 2D image, but also the directions of the incoming light rays. The additional light directions allow the image to be re-focused and the depth map of a scene to be estimated. 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

A. Subimage and sub-aperture images

Concerning Lytro, the raw data is not provided with its software, therefore we have used the tool to convert the data with the Lytro file format to RAW data. Generally, a RAW image is converted (de-multiplexed) into a set of multi-viewpoint images called sub-aperture images. A sub-aperture image is created by gathering a pixel from each subimage, which is located in the same relative coordinate with respect to the subimage. The subimage is the image which formed under a microlens and on the image sensor.

Fig. 1 The set of sub-aperture images

The de-multiplexing process consists in reorganizing the pixels of the raw image in such a way that all pixels capturing the light rays with a certain angle of incidence are stored in the same image creating the so-called sub-aperture images. Each sub-aperture image is a projection of the scene under a different angle of view. The set of sub-aperture images creates a block matrix where the central of sub-aperture image store the pixels capturing perpendicular light rays to the sensor. In fact, the angular information of the light rays is given by the relative pixel positions in the subimages with respect to the subimage centers. The subimages and sub-aperture images is shown as Fig.2.

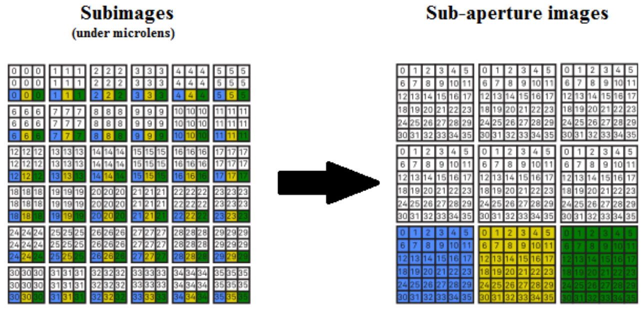


Fig. 2 Subimages and sub-aperture images

We use Lytro light field camera [2] as a device to generate the image datasets. We split the raw light field file to different angle of views by the Lytro Power Tools [4]. It provides the images which display same location view with the various angles of views as shown in Fig.1.

B. Two-plane parameterization

In this paper, we adopt the two-plane parameterization [5]. In this parameterization, oriented rays are represented by their intersections with two paralleled planes (Π: camera plane, Ω: image plane), as shown in Fig. 3. The intersection point on the Π plane is represented as (s, t), and the point on the Ω plane is represented as (x, y). Therefore a 4-D light field can be denoted as L(x, y, s, t).

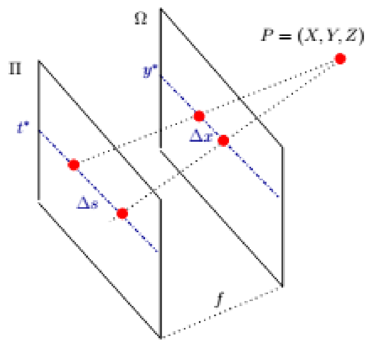


Fig. 3 Two-plane parameterization

C. Epipolar plane images (EPIs)

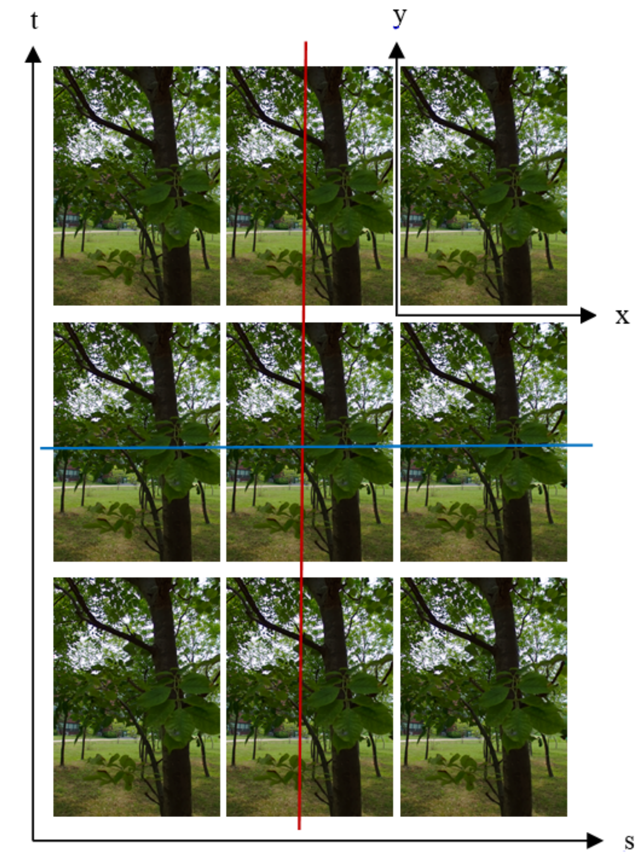
Under two-plane parameterization, we can get a 2D slice of the light field. By fixing y and t, we can obtain a 2D (x, s) slice as shown in Fig. 4(b). The slice can be seen as one horizontal scan-line of the scene observed from observers with different horizontal positions. Similarly, 2D (y, t) slices can be obtained if x and s are fixed as shown in Fig. 4(c). These 2D slices are called epipolar-plane image.

Epipolar-plane image 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], which is written as:

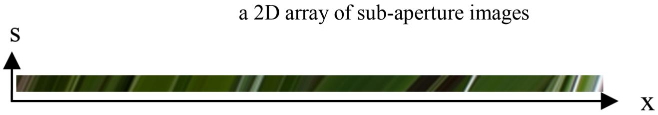
$$\Delta x = -\frac{f}{Z}\Delta s \;\Rightarrow\; Z = -f\frac{\Delta s}{\Delta x} \qquad (1)$$

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 line in EPIs which indicates the real metric depth value Z

Therefore, the depth value can be obtained by estimating the slope of lines in EPIs [7].



(a) A 4D light field can be interpreted as a 2D array of sub-aperture images



(b) The EPIs is related to the horizontal line direction (blue line)

(c) The EPIs is related to the vertical line direction (red line)

Fig. 4 An illustration of EPIs

D. *Modified structure tensor*

Local structure tensor [8, 9] have been used in image processing to solve problems such as motion detection. This method 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:

$$J = \begin{bmatrix} G_\sigma * (S_x S_x) & G_\sigma * (S_x S_y) \\ G_\sigma * (S_x S_y) & G_\sigma * (S_y S_y) \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix} \tag{2}$$

where *Gσ*: Gaussian smoothing kernel with standard deviation σ. *Sx, Sy*: Horizontal and vertical component of the gradient vector at each pixel in EPIs *S* respectively.

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:

$$V = \begin{pmatrix} J_{yy} - J_{xx} + \sqrt{(J_{yy} - J_{xx})^2 + 4J_{xy}^2} \\ -2J_{xy} \end{pmatrix} \tag{3}$$
$$V^{\perp} = \begin{pmatrix} 2J_{xy} \\ J_{yy} - J_{xx} + \sqrt{(J_{yy} - J_{xx})^2 + 4J_{xy}^2} \end{pmatrix}$$

The corresponding eigenvalues for each eigenvector are as follows:

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

The relationship of the eigenvectors with the EPIs is shown in Fig. 5. We can compute the slope of the EPIs line by using the intersection angle of two eigenvectors with x-position axis. Moreover, local structures can be determined as edges ($v^{\perp} \gg v \approx 0$) based on the two eigenvalues.

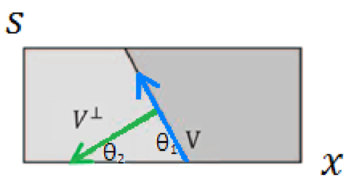


Fig. 5 The relationship of the eigenvectors with the EPIs

E. *Depth and reliability information on EPIs*

We can compute the depth information via equation (1) and the relationship of eigenvectors with the EPIs as:

$$d = -f\frac{\Delta s}{\Delta x} = \tan(\theta_1) \tag{5}$$

We use the correspondence of the structure tensor *J* as the reliability information [8].

$$r = \frac{(J_{yy} - J_{xx})^2 + 4(J_{xy})^2}{(J_{xx} + J_{yy})^2} \tag{6}$$

F. *Initial depth estimation*

Using the depth information *d* and reliability information *r* for all the EPIs in horizontal and vertical direction. We superimpose all depth *d* based on the corresponding reliability information by select the depth information that corresponds to higher reliability information 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

In this section, the proposed method is implemented by using test images which capture by Lytro light field camera [2] and the Stanford light field archive [10]. Because the Lytro light field camera and the Stanford light field archive do 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. 6.

Finally, even if our method considers only the sub-aperture images, it provides the initial depth estimation and we have noticed errors on some regions of the image. Adding further the all-in-focused, focal stack and profiles of the light field images datasets could be effective to accurate disparity estimation. This is one of our perspectives for future work.

IV. Conclusions

In this paper, we proposed an initial depth estimation method using EPIs and modified structure tensor. The proposed method generated the initial depth map via only the sub-aperture images. We estimate depth using dominant on EPIs, which are obtained using the modified structure tensor. As a result, our method generated improved the initial disparity map compared with other methods. Our future plan will combine the all-in-focused, focal stack and profiles of the light field image information to enhance and make more accurate on the final depth map.

Acknowledgment

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

REFERENCES

[1] G. Lippmann, La photographie integrale, Academic des Sciences, vol. 146, pp. 446-451, 1908.

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

[3] Raytrix, "The Raytrix camera," http://www.raytrix.de.

[4] Lytro Power Tools, http://forums.lytro.com.

[5] M. Levoy and P. Hanrahan. "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 31–42. 1996.

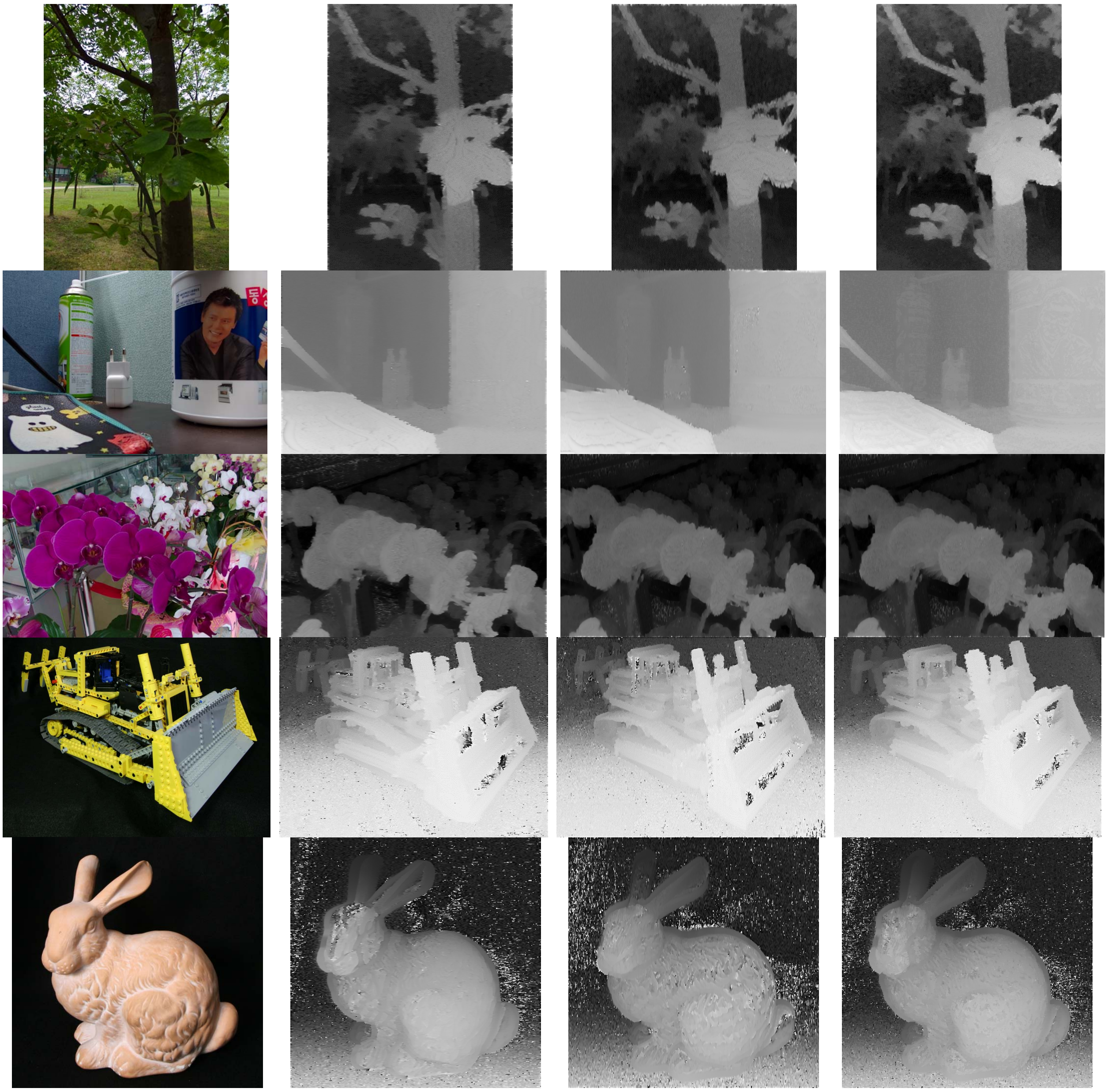
[6] R. C. Bolles, H. H. Baker, and D. H. Marimont. "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1(1), pp.7–55, 1987.

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

[9] A. Baghaie and Z. Yu, "Structure tensor based image interpolation method," *AEU - International Journal of Electronics and Communications*, vol.69(2), pp. 515-522, 2015.

[10] "Stand (new) light field archive", http://lightfield.stanford.edu/lfs.html.



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

Fig. 6 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