Occlusion Aware Reduced Angular Candidates based Light Field Depth Estimation from an Epipolar Plane Image

Ji-Hun Mun and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, South Korea

Abstract

Light field images are represented by capturing a densely projected rays from object to camera sensor. In this paper we propose a novel method for depth estimation from light field epipolar plane image (EPI). Contrary to the conventionally used depth estimation method such as stereo matching, the depth map is generated without high computational complexity via EPI. In order to obtain an accurate depth value from EPI, optimal angular value has to be founded for each pixels. If we consider all the angular candidate value for optimal angle value selection, that cause high computational complexity. Instead we consider all candidate value, reduce the angular candidate by analyzing the EPI patterns. In addition, to improve the quality of estimated depth map from EPI, occlusion area is handled before computing a matching cost. As a pre-processing, average and variance value are computed within specific window size to detect and replace the occlusion area. To validate the efficiency of our algorithm, we experiment with computer graphics and also dense light field data set. The experiment results show that our algorithm achieve better performance than conventionally used depth estimation methods.

1. Introduction

Light field camera include a function that preserves the light rays which flowing in any direction through an image plane. Unlike the general 2D image, the light field image additionally includes angular coordinate data. In order to represent the light field light ray which projected from object, we need to defined two planes as indicated in figure 1. The first plane is image plane$[\Omega; x, y]$ that contain ray data from main lens. The second plane is camera plane$[\Pi; s, t]$ which represent the set of image plane data. Since those two plane has 4 coordinate information, we define that light field is composed with 4D space.

The light field images are usually produced by gantry n x n multiple camera set [1] and rendered by 3D computer graphics. All the light field data generation procedure has to include amount of different view-point images must be obtained. However, the gantry light field camera system has complex hardware structure and also need a comprehensive calibration procedure for image ratification and error correction.

Recently, instead of the gantry based light field camera system many commercial plenoptic light field cameras are invented such as Lytro[2] and Raytrix[3]. The plenoptic light field camera can capture scenes with slightly different view point. Since microlens array in plenopic light field camera, various view point images are available. Currently, plenoptic light field camera is used for 3D reconstruction, virtual view synthesis, matting [4] and depth estimation [5-8].

Estimating a depth map from the light field image has been a rough problem for several years. For example, focal stack, defocused image confusion and epipolar plane image based depth estimation methods have been developed. Especially, EPI based depth estimation method via optimal angular value selection method [9] has low computational complexity than other depth estimation method, since that does not use global optimization technique.

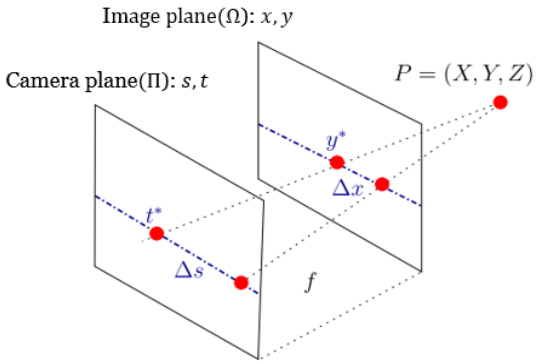


Figure 1. Light field camera plane structure

In this paper, we introduce novel depth estimation method from light field EPI data. The proposed method composed with two main contributions. The first function is decreasing a computational complexity by reducing angular candidates for cost computation. Generally, to find an optimal angle value from angular candidates (0°~180°), the costs are computed in that angular range. In order to reduce the number of angular candidates, we analyze the pattern of each EPI data. Since each EPI data has similar gradient pattern, the number of angular candidate for cost computation is efficiently reduced.

An occlusion problem always happens in depth estimation not only stereo image but also in light field images. Microlens array which used in light field camera split the light rays. The splitted ray provide different view point image and cause an occlusion problem. To take occlusion problem into account during depth estimation procedure, we compute a variance value within the specific window patch. Generally, the occlusion area occurs near the object boundary region. Therefore, we measure the variance value around the object boundary region before computing a cost value for optimal angle selection.

In order to generate a final depth map from light field EPI, weight median filter is applied to initially estimated depth map. As a weighting factor, bilateral filter is adopted. The post processing improve the accuracy of depth map than initially estimated depth map by reducing the small error region. Our algorithm firstly reduced the computation complexity by decreasing the angular candidates, and also improve the depth map accuracy throughout the

occlusion aware post processing. Since light field depth estimation applicable to many image processing technique, we perform experiment with CG and dense light field dataset.

2. Light Field EPI Analysis

The light field camera system composed with two plane; image plane Ω and camera plane Π as indicated in figure 1. In order to simplify the procedure of depth estimation from light field image, we represent the light field system using two plane. We assume that the camera plane Π is located in $z = 1$ and image plane Ω has distance with that plane at f. Any pixel coordinate in light field camera system can be expressed in 4D space such as $[x, y, s, t]$. Thus, the 4D light field coordinate is correspond to (1).

$$L: \Omega \times \Pi \rightarrow \mathbb{R}, \qquad L(x, y, s, t) \tag{1}$$

To efficiently describe a light field EPI data generation procedure, we consider a fixed image plane coordinate y^* and camera plane t^*. The camera plane describes a set of image plane which caused by different camera view point. From each horizontal or vertical camera plane data set, we fix a specific image coordinate in horizontal and vertical line. The EPI data is generated by stacking a fixed horizontal and vertical pixel value in regular order as represented in figure 2. The horizontally constant red line which correspond to each sub-aperture image indicates a sample horizontal pixel line for EPI generation.

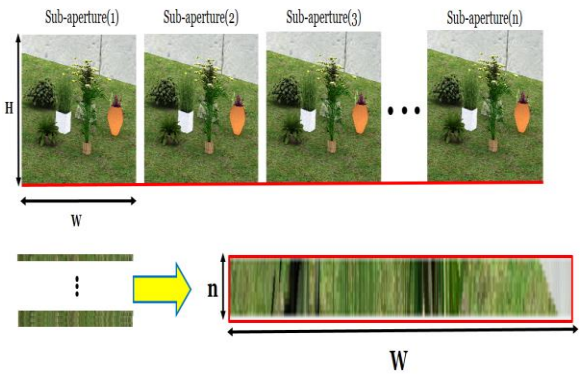


Figure 2. Light field EPI data generation

Each image plane data is same with the sub-aperture image as indicated in figure2. Depending on the number of sub-aperture images, height of EPI data is affected such as n and the width of sub-aperture is same with the width of EPI data W. In addition, total number of EPI data has same number with sub-aperture image height H. From the EPI generation procedure, we notice that the EPI data is simply constructed even though a dataset include complex structures or homogeneous pattern.

Concerning the light field structure in figure1, a triangular relationship need to be considered for depth value generation from EPI data. As displayed in figure 3, we observe the light field camera system in 2D plane. If the triangular relationship equation is organized in terms of Δx with varying parameter Δs, then it can be represented in (2).

$$\Delta x = -f\frac{\Delta s}{Z} \tag{2}$$

In equation (2), f is the focal length of light field camera which capturing scenes. Δs indicates a camera view point variation parameter and Δx is pixel point distance between different image plane.

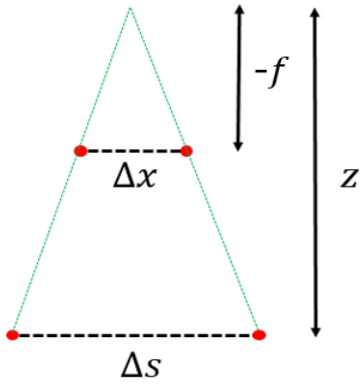


Figure 3. Light field EPI data generation

To obtain a real depth value, the equation (2) has to be reorganize in terms of Z as indicated in (3). The ratio value $\frac{\Delta s}{\Delta x}$ indicates a slope of line in EPI image as represented in figure 4. Multiplying focal length with variation of image plane coordinate and camera plane coordinate provide a real depth value. In this paper we basically use EPI slope computation technique for depth estimation.

$$Z = -f\frac{\Delta s}{\Delta x} \tag{3}$$

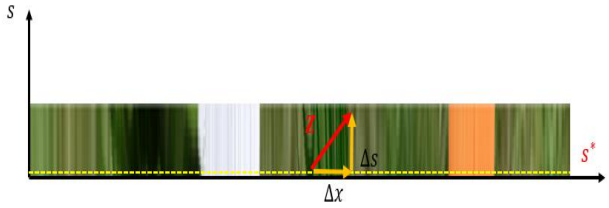


Figure 4. Depth generation from EPI data

3. Depth Estimation from light field EPI

Depth map generation from EPI data via slope computation method need to consider all angular candidates. Since it causes a high computational complexity while generating a depth map, we propose angular candidate reduction method. In addition, occlusion aware pre-processing method is proposed. Since the conventional EPI depth estimation method and stereo matching method does not handle the occlusion area, depth map includes error near the occlusion area. To prevent occlusion artifact in final depth map, we replace the occlusion area before cost computation for optimal angle selection.

3.1 Occlusion elimination

Not only for the stereo matching procedure in light field image in depth estimation procedure, but also slope computation from EPI data affected by occlusion. Since occlusion interrupt to photo consistency while cost computation, it derives an inaccurate depth map.

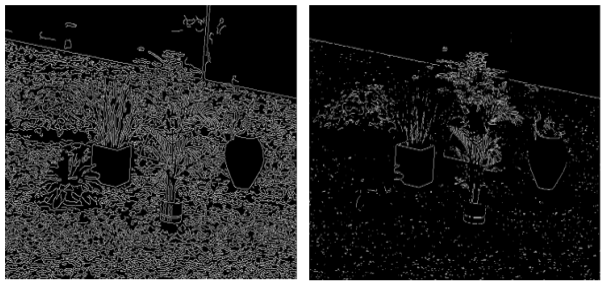
Generally, occlusion area appears near the object boundary region, since a view point change mainly affect to occluded region between objects area. In order to eliminate an occlusion area from input images, we extract the edge region via Canny edge detection on the left most image plane data. Since our algorithm generates a EPI data from left most image plane, edge regions are detected from starting viewpoint image.

From the extracted edge region an edge directional information is computed. The extracted edge region via Canny edge detection method has too many edge information as indicated in figure 5(a). Our interesting area is not a background or foreground region but the object boundary region. In order to find out the meaningful edge information edge surrounding condition is computed by considering a different intensity value near the extracted edge region. As defined in (4), the edge surrounding condition is easily computed by computing an average of pixel intensity and comparing an intensity value in RGB channel. In (4) k is the additional search range near the current pixel, i indicates pixels which included in window patch and n is the number of input image. If the intensity value difference is larger than Avg value (5) of window patch, then we define that region is our consideration for occlusion elimination.

$$E_{con} = \begin{cases} 1, if \sum_{k\in[-2,+2]} \sum_{i\in W} (I_n(i,i+k) - I_{n+k}(i,i+k) > Avg \\ 0, \quad else \end{cases} \tag{4}$$

$$Avg = \frac{1}{E} \sum_{k\in[-2,+2]} \sum_{i\in W} I_n(i,i+k) \tag{5}$$

The filtered extracted edge region is represented in figure 5 (b). Compared to the Canny edge detection result and filtered edge region, we notice that filtered edge region provides more meaningful edge information than pre-filtered edge region.



(a) Canny edge detection result (b) Filtered edge region

Figure 5. Edge extraction and filtering result

Based on the filtered edge detection result, pixel intensity is computed on each pixel as defined in (6). Since the filtered edge image still contains small part of edge information around the foreground region, we need to efficiently remove that edge information. If the difference of pixel intensity on filtered edge regions has higher value than pre-defined threshold value, we defined that region is real occlusion area. In this paper for the experiment we set the threshold value 0.5.

$$Occ = \begin{cases} 1, & if\ |L_{n_{RGB}}(x',y') - L_{n+1_{RGB}}(x',y')| \geq threshold \\ 0, & else \end{cases} \tag{6}$$

Since the occlusion affect to cost computation for optimal angular value selection in EPI data, the occlusion has to be handled. To improve the cost computation value accuracy, we replace the occlusion area with neighbor pixel value. The replacing pixel value is determined after computing a variance value within specific window patch.

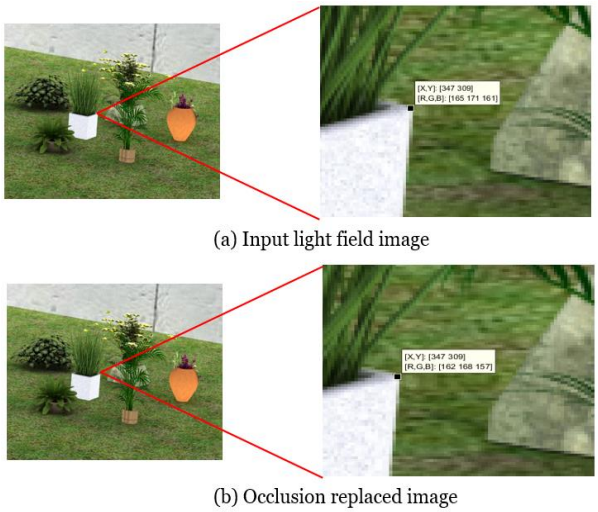
For occlusion replacement, we only compute a variance value on the final occlusion area. The pixel intensity value near the occlusion between current and neighbor view point image has different value since that contain differently projected ray information. To replace that region, we firstly compute an average value within the specific window patch where the center pixel is occlusion pixel. Computing average value around the occlusion is defined in (7). LF indicates a different camera plane light field image, w is the specific window size (e.g. 3×3, 5×5, etc…) and it computed in 4D space x, y, s, and u.

$$\overline{LF_k}(x,y) = \frac{1}{N_k} \sum_{s,u} \sum_{p\in W} LF(x_p, y_p, s_k, u_k), \qquad k = 1,2 \tag{7}$$

After computing an average value, variance value in window patch is computed throughout the (8). The replaced pixel value is determined by selecting a minimum variance value along with the ± 3 neighbor pixel in horizontal direction. By considering a neighbor pixel value, we avoid viewpoint variation flaw in light field image.

$$V_k(x,y) = \frac{1}{N_k} \sum_{s,u} \left(LF(x_i, y, s_k, u_k) - \overline{LF_k}(x,y)\right)^2, \quad k = 1,2 \;\; i = \pm 3 \tag{8}$$

As represented in figure 6, we confirm that after replacing the occlusion pixel value with neighbor pixel value. Since the occlusion has different pixel value with neighbor sub-aperture pixel value, we adopt a pixel value which has smallest variance among the occlusion neighbor pixel.



(a) Input light field image

(b) Occlusion replaced image

Figure 6. Input and occlusion replaced image pixel value comparison

3.2 Candidate reduction for optimal angle selection

Conventionally invented slope-EPI based depth estimation method considers all angular candidate value range from 0° to 180° as indicated in figure 7. However, all angular candidate consideration causes a high computational complexity problem. Currently invented application of light field image such as virtual view synthesis, VR content generation and 3D reconstruction need low computational complexity with high accuracy. Generally, computation complexity from light field EPI –slope based depth estimation method affected by number of angular candidate, maximum length of EPI diagonal and width of EPI data. Especially, we focus on the number of angular candidate for low computational complexity. To handle high computational complexity issue, we propose reducing the number of angular candidates via EPI characteristic analysis.

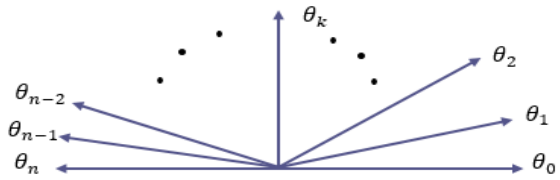


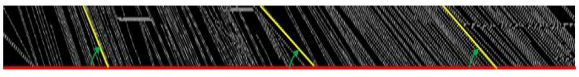
Figure 7. Angular candidate ranges for EPI slope computation

To reduce the angular candidate from original range, we analyze the pattern of EPI data. As indicated in figure 8, each EPI data has similar slope patterns.

Figure 8. Angular candidate ranges for EPI slope computation

Our angular candidate reduction procedure is composed with 3-steps. Firstly, extract the edge region via Canny edge detection from each EPI data as indicated represented in figure 9 (a). Even though all angular gradients are not detected, slope pattern of EPI can be easily noticeable. From the extracted slope edge pattern, we assess the slope angle for each bottom line pixels as expressed in figure 9 (b) red line. It shows the sample slope angle assessment procedure from EPI edge map.

(a) EPI edge detection result



(b) EPI edge slope pattern assessment

Figure 9. EPI edge detection and assessment

The slope is measured by using the Hough transform method, since that can compute the angle between two lines which crossing each other. Line curve fitting algorithm is applied on two-lines equation to find a coordinates where two lines are crossed. From the crossing point coordinate, we can measure the angle between two-lines.

Angular values are measured for all edge detected map, and the measured angle compose a histogram as defined in (9). H_e is the histogram about all measured angular value in EPI edge map, and k indicates total number of histogram level.

$$H_e = \sum_{i=1}^{k} P_e(i), \qquad Angle_{min} \le k \le Angle_{max} \tag{9}$$

Even though we measure the angle from the edge detected map in figure 9 (b), not all the edge data is extracted accurately. Since the original EPI includes slope patterns for each bottom line pixels, all the angular pattern should be considered to define a properly reduced angular search range. In order to compensate the inaccurately detected edge values in histogram H_e, the range of histogram value is adjusted by normalizing the H_e. The reduced angular search range is adjusted based on mean and variance of histogram as defined in (10).

$$\begin{aligned} C_{i,j} &= \frac{R_{i,j} - E(R)}{\sigma_R}, \\ E(R) &= \frac{1}{N}\sum_{k=1}^{N} R_k \\ \sigma_R &= \sqrt{\frac{1}{N}\sum_{k=1}^{N}(R_k - E(R))^2} \end{aligned} \tag{10}$$

where $C_{i,j}$ is the adjusted minimum and maximum angular search range and $R_{i,j}$ represents originally measured minimum and maximum angular search range in histogram. $E(R)$ and σ_R indicate mean and variance of histogram respectively. To compute a mean value, summation of all angular value frequency R_k is divided by total number of angular candidate levels N. By adjusting an angular search range using mean and variance of histogram value, improperly detected edge region in EPI is compensated.

3.3 Cost computation

Based on the reduced angular candidates, we compute a cost value for depth map generation. Depth values selection for each pixel are same with finding an optimal angular value as explained in (3). In order to find an optimal angle from the angular candidates, we search the minimum cost value as defined in (10). Minimum cost value for optimal angle at each camera and image plane $k_\alpha = (s_\alpha, t, u, v)$. Apparently, the cost is computed within the reduced angular candidates Rc.

$$\theta(k_\alpha) = arg \min_{\theta \in Rc} C(\theta, k_\alpha) \tag{10}$$

Similar with the generally used matching cost computation in stereo matching procedure, we define a cost computation function (11) which uses x and y directional gradient in gray scale and RGB channel EPI data. In order to consider the consistency between RGB channel and gradient image, we combine that terms using weight factor α. In this paper, we set the α as 0.3 to enforce the effect of color channel EPI data for photo consistency.

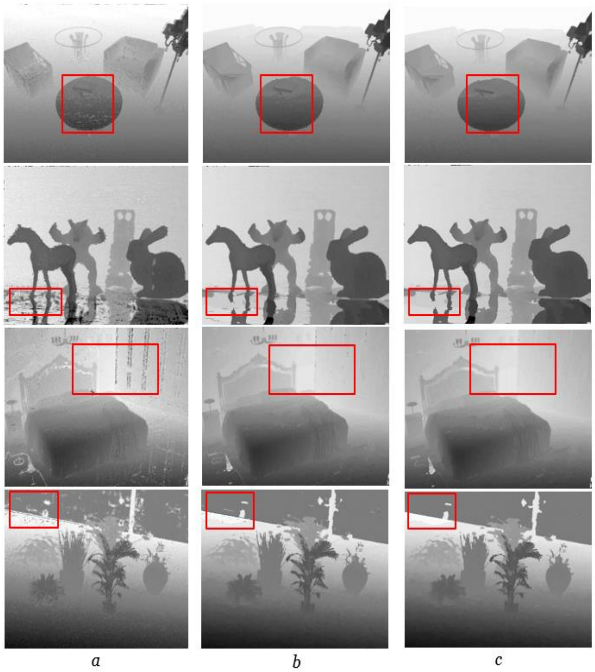
$$C(p) = \sum_{p_n \in N(\theta)} \alpha \left| E^g_{x,y}(p_n) - E^g_{x,y}(p) \right| + (1-\alpha) \left| E(p_n) - E(p) \right| \quad (11)$$

4. Experimental Results

To verify the performance of our proposed method, we conduct the experiment using four CG datasets [10]. Since those datasets provide ground truth depth image, the quality of generated depth map can be numerically evaluated in terms of bad pixel rates (BPR). The BPR indicate that an estimated depth value has a bigger than 1 difference with ground truth depth value. Since the proposed method eliminates occlusion regions for depth map quality improvement, we compare the BPR at the non-occluded and whole region.

The window size for mean and variance value computation in overall procedure is set to 5×5. Even though we can adopt larger window size, but that size does not correctly find out the characteristics of object boundary for occlusion elimination procedure. To improve the depth map quality at final step, we apply the weighted median filter [11]. The weighted median filter removes the remaining small hole region in the final depth map.

The proposed method result is compared with other state-of-the-art light field depth estimation methods. Zhang et al [9] use EPI data for depth estimation from light field image. But, they do not consider the number of angular candidates and also the occlusion area is not eliminated before computing a cost value. Wanner et al [8] propose a depth estimation method using EPI tensor. Even though the tensor provides an angular value of EPI for each pixels, that method does not handle the occlusion issue which appear in input light field image. Since those methods estimate depth map from the light field EPI data, we compare the performance of our algorithm with their methods. Fig. 10 shows the experimental results of proposed method and other conventional methods.

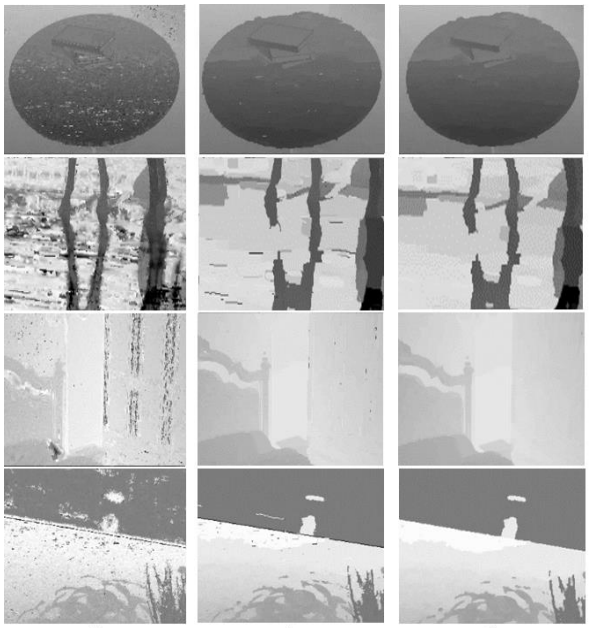


a *b* *c*

Figure 10. Experimental results of depth estimation

Experiments are performed with CG dataset which composed of 9×9 light field images. The range of camera plane image affect to occlusion issues in light field depth estimation procedure, since wide camera plane provides informative pixel value when eliminating occlusion areas in pre-processing. Fig. 10 a is the experiments results of Wanner method, Fig. 10 b indicates the Zhang method experiment results and Fig. 10 c is the proposed method results.

To clarify the efficiency of proposed algorithm, we enlarge the occlusion region which represented in Fig. 10 experimental results on Fig. 11. From the experimental result of occluded region, we notice that the proposed method outperforms other state-of-the-art light field depth estimation methods. Since the pixel values of occlusion regions are replaced with neighbor pixel values, it provides an accurate cost value when finding an optimal angular value selection.



a *b* *c*

Figure 11. Enlarged occlusion region

Not only the result images but also we compare the numerical performance with conventional method. Table 1 shows the BPR comparison results. From the BPR comparison result we notice that the Wanner method has lower depth value accuracy than other depth estimation method in occluded region.

Table1. BPR comparison results (%)

Sequences		Wanner[8]	Zhang[9]	Proposed
Livingroom	Non-occ	8.8	7.4	6.8
	All	14.7	12.3	11.2
Sculptures	Non-occ	9.3	7.7	7.3
	All	15.2	12.7	12.3
Bedroom	Non-occ	9.7	7.1	6.6
	All	14.8	12.1	12.8
Plant	Non-occ	8.2	7.3	6.9
	All	13.5	11.8	12.5

The objective of proposed method is improving the depth map quality while reducing the computational time. Because of that reason, we measure the computational time of proposed method with other methods. Table 2 shows the computational time measurement results.

Wanner method compute a tensor to estimate depth value from light field images and Zhang method consider all angular candidates for depth estimation. Contrary to our method and Zhang method, Wanner method compute a tensor. As a result of the computation time measurement, Wanner method shows faster than Zhang method in terms of in terms of depth estimation procedure. Since our method reduce the angular candidate using EPI pattern analysis, it shows 3.8 times faster result than Zhang method in terms of computational time. From the experimental results, it has been confirmed that the proposed method efficiently reduces the number for angular candidate while improving the accuracy of depth map.

Table2. Computational time comparison results(sec)

Sequences	Wanner[8]	Zhang[9]	Proposed
Livingroom	322	782	315
Sculptures	285	771	270
Bedroom	308	687	285
Plant	314	649	239

5. Conclusion

In this paper we propose depth estimation method from light field EPI data. In order to handle the occlusion problem which caused by view point variation, we replace the occlusion pixel value with neighbor sub-aperture image pixel value. To check the proper pixel value for replacement, average and variance value are computed within the specific window size. In addition, angular candidate reduction method also proposed. Since each EPI data has their own slope pattern, we can efficiently reduce the angular candidate by analyzing the EPI pattern. Instead of considering all angular candidate range, we only compute cost values within the reduced angular candidates range. The experimental results show that our proposed depth estimation from light field EPI generates accurate depth map than other methods, also it was performed in a short time than other conventional methods.

Acknowledgements

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) and 'Brain Korea 21 Plus Project' of the Ministry of Education & Human Resources Development, Republic of Korea (ROK) [F16SN26T2205].

References

[1] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Bart, A. Adams, M. Horowitz, and M. Levoy, "High Performace Imaging using Large Camera Array," SIGGRAPH, vol. 24, no. 3, pp. 765-776, Aug. 2005.

[2] https://illum.lytro.com.

[3] https://www.raytrix.de

[4] D. Cho, S. Kim, and Y. –W, Tai, "Consistent Matting for Light Field Images," ECCV, pp. 90-104, Sept. 2013.

[5] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu, "Light Field Stereo Matching using Bilateral Statistics of Surface Cameras," CVPR, pp. 1518-1525, June 2014.

[6] H. Lin, C. Chen, S. B. Kang, and J. Yu, "Depth Recovery from Light Field using Focal Stack Symmetry," ICCV, Dec. 2015.

[7] M. Tao, P. P. Srinivasan, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi, "Depth from Shading, Defocus, and Correspondence using Light Field Angular Coherence," CVPR, pp. 1940-1948, June 2015.

[8] S. Wanner and B. Goldluecke, "Globally Consistent Depth Labelling of 4D Lightfields," CVPR, pp. 41-48, June 2012.

[9] Y. Zhang, H. Lv, Y. Liu, H. Wang, X. Wang, Q. Huang, X. Xiang, and Q. Dai, "Light-Field Depth Estimation via Epipolar Plane Image Analysis and Locally Linear Embedding," TCSVT, vol. 27, no. 4, pp. 739-747, April 2017.

[10] Synthetic data sets, http://cseweb.ucsd.edu/~viscomp/LF/dataset.zip.

[11] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," TPAMI, vol. 35, no. 2, pp.504-511, Feb. 2013.

Author Biography

Ji-Hun Mun received his B.S in electronic engineering from the Chonbuk National University, Chonbuk, Korea (2013) and his M.S. in Science of Information and Communication from the Gwangju Institute of Science and Technology, Gwangju, Korea (2015). Now he takes Ph.D. course in Science of Information and Communication from the Gwangju Institute of Science and Technology, Gwangju, Korea. He works has focused on the video codec H.265/HEVC, 3D video technology and stereo matching.

Yo-Sung Ho received his B.S. and M.S in electronic engineering from the Seoul National University, Seoul, Korea (1981) and his Ph.D. in electrical and computer engineering from the University of California, Santa Barbara (1990). He worked in Philips Laboratories from 1990 to 1993. Since 1995, he has been with the Gwangju Institute of Science and Technology, Gwangju, Korea, where he is currently a professor. His research interests include image analysis, 3D television, and digital video broadcasting.