INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2017/ M 41017
July, 2017, Torino, Italy

Source	**Gwangju Institute of Science and Technology (GIST)**
Status	**Report**
Title	**[MPEG-I-Visual]: Depth Map Generation from Light Field Epipolar Plane Image**
Author	**Ji-Hun Mun (GIST, jhm@gist.ac.kr)** **Yo-Sung Ho (GIST, hoyo@gist.ac.kr)**

Abstract

Light field images are composed of horizontally or vertically aligned multiple images. In this document, we propose an epipolar plane image (EPI) based depth map estimation method from light field image. Especially, we reduce the number of angular candidate for low computational power consumption and accurate depth map generation from light field EPI.

1 Introduction

Since the Lytro[2] or Raytrix[3] light field camera image has a dense baseline, general stereo matching method is not working well. The purpose of this document is providing a depth estimation method from the light field EPI [1]. For each row line of light field images are stacked in order of image sequences to generate a light field EPI. In order to estimate a depth value for each pixel, we find an optimal angle. Due to the error regions in an initial depth map, we apply weighted median filter at last step.

2 Light-Field EPI Generation

A scene captured with different focusing point is called sub-aperture image in the light field camera system. As indicated in figure 1, light field image composed of image(Ω) and camera(π) plane. Image plane provides spatial image information and camera plane adjusts light ray which transmitted to image plane. Since the camera plane coordinates are varying while capturing a scene, we can get a differently focused image through a light field camera.

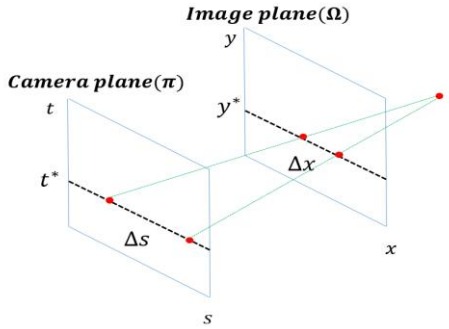

Figure 1. Camera and Image plane in light field camera system

Due to the narrow baseline of each sub-aperture image, conventionally used stereo matching method does not generate an accurate depth map. Instead of using general stereo matching method, we adopt an EPI based depth map generation method. As displayed in figure 2, EPI is created by stacking row lines of each sub-aperture image. From that results, we can easily notice that EPI consist of simple linear structure. Even though a scene has complex and fluctuating pixel intensity, EPI shows a characteristics of light field image. In figure 2, *H* and *W* indicate height and width of sub-aperture image respectively. The width of EPI and sub-aperture image are same, and height of EPI is same with the number of input sub-aperture image.

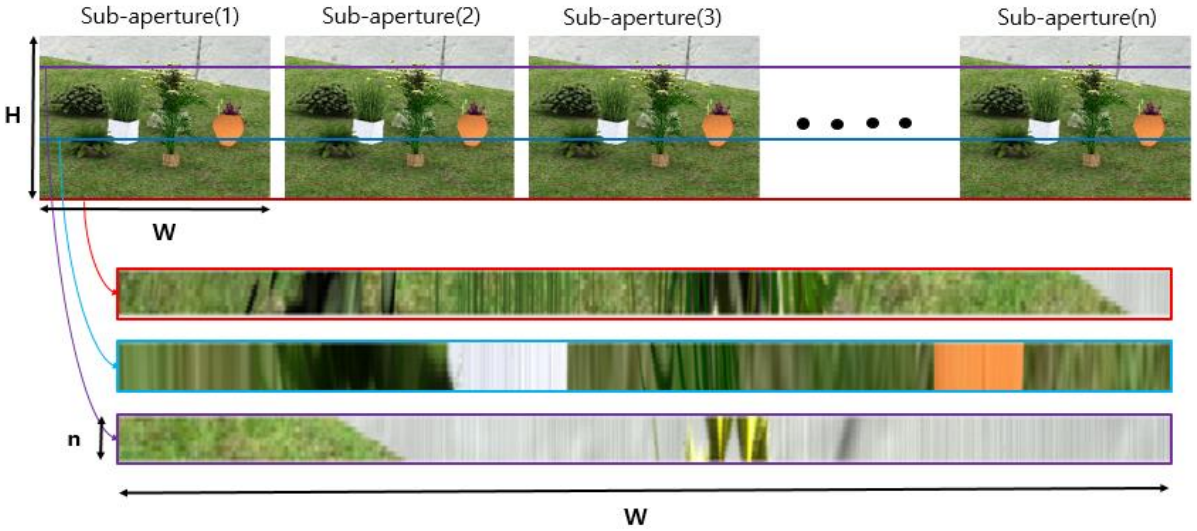

Figure 2. EPI generation from light field images

3 Depth Estimation from EPI

To generate a depth map, light field image shearing based matching cost computation method is introduced in 118th MPEG meeting [4]. However, due to the narrow baseline of light field image, matching cost computation result does not provide a precise depth map result with respect to object boundary region. Since the accuracy of object boundary region affect the estimated depth map quality evaluation, we have to improve the depth map generation method.

In this document, we use a light field EPI to generate a depth map. As represented in figure 1 and 3, an EPI is generated from image plane(x, y) and camera plane(s, t). Since the camera plane encloses a motion data of scene variation, a generated EPI has slope pattern as indicated in figure 3 (b).

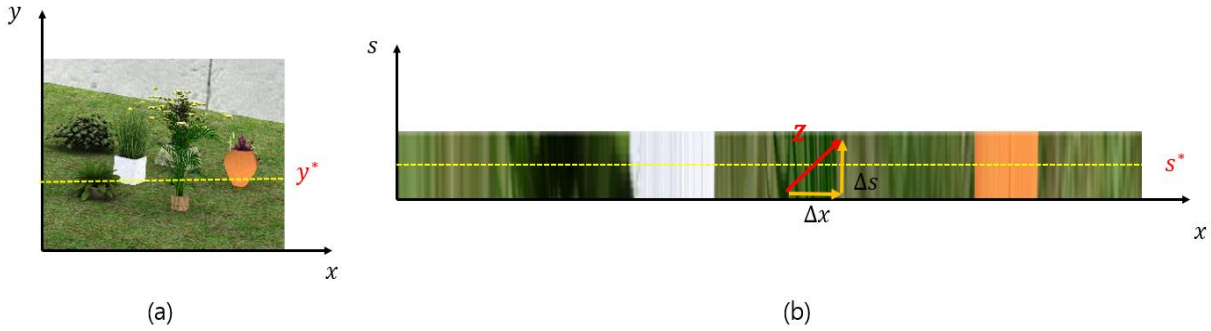

Figure 3. EPI generation from light field images

Before obtaining a depth map from EPI, we have to consider relationship between image and camera plane as shown in figure 4. f indicates a focal length and Z represents real depth value from object and scene plane. Δx and Δs are the coordinate variance of image and camera plane respectively.

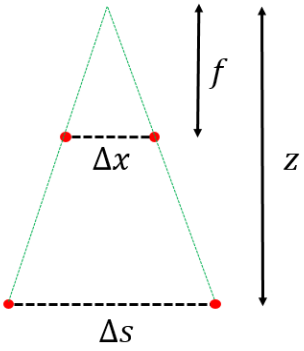

Figure 4. Triangular proportion

When we abbreviates the triangular relationship with respect to Δx, then we can get an equation explained in (1). To obtain a real depth value from (1), that equation is abbreviated in terms of Z. From the equation (2) and figure 3 (b), we can notice that the real depth value Z is acquired from variation of x and s coordinate value. In other words, depth map estimation from EPI is same with finding an accurate directional angle value from EPI.

$$\Delta x = -f\frac{\Delta s}{Z} \tag{1}$$

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

Then we notice that a depth value of each horizontal line of image pixel is estimated by finding an optimal angular value from EPI. Encouraged by this concept, we inspect all of the angular candidates value such as set A={0°, 1°, … 180°} as shown in figure 5.

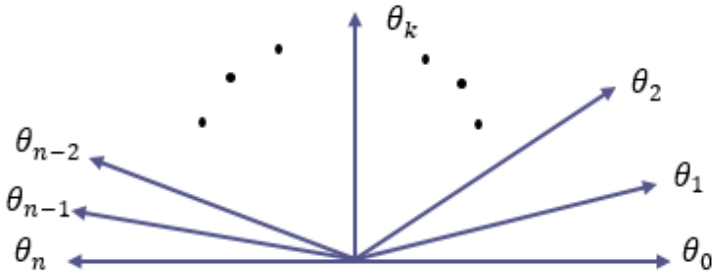

Figure 5. Angular candidates for depth estimation from EPI

For a given EPI data E_l align with the horizontal direction l, and each pixel coordinate $\mathrm{P} = (x, s)$, the optimal angle value is computed via (3). However, measuring all of the angular coordinate to find an optimal angle value is an inefficient procedure in terms of computational complexity.

$$\theta(l, \mathrm{P}) = arg\min_{\theta \in A} C(E_l, P) \tag{3}$$

4 Reduced Angular Candidate for EPI Depth Estimation

Discriminating an optimal angle value among angular candidates is imperative work for EPI depth estimation. Since we have to select an optimal angle from EPI, the initial angular candidate range is same with the set A={0°, 1°, … 180°}. Matching cost equation for optimal angle selection from EPI is defined in (4). Where $N(\theta)$ indicates the number of angular directional neighbor pixels, and E denotes the EPI. E_x and E_y are gradient of x and y direction of EPI. α is weighting factor which controls between EPI and gradient of EPI term.

$$\mathrm{C}(E_l, P) = \frac{1}{N(\theta)} \sum_{P_n \in N(\theta)} \begin{matrix} (1-\alpha)|E(P_n) - E(P)| \\ +\alpha[|E_x(P_n) - E_x(P)| + |E_y(P_n) - E_y(P)|] \end{matrix} \quad (4)$$

Instead of computing all angular candidate set A to find an optimal angle, we reduce the number of angular candidates via averaging a normalized directional tendency. Since each EPI is generated from input light field images that has similar angular pattern as indicated in figure 6, we investigate the tendency of angular direction.

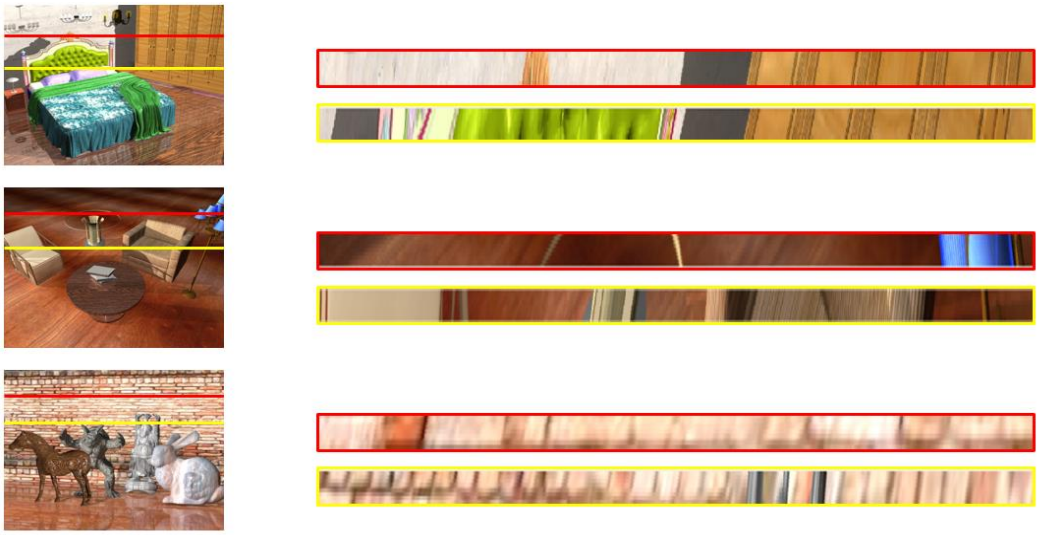

Figure 6. Sample EPI region and angular pattern tendency

In order to exclude a texture smearing and occlusion artifact from RGB channel EPI, we inspect an angular tendency from edge detected EPI. The edge detected EPI only provide an angular patterns using binary value as indicated in figure 7.

We assume a single line, which represented by red line in figure 7. Based on that line, we measure all angular directions at the bottom row pixels. The angular direction tendency examination is executed only on the bottom row, due to the angular direction is started from bottom of edge detected EPI. To correctly measure the angular direction without the effeteness of object boundary or occluded region, we normalize the all angular tendency. By using this method, we can reduce the angular search range for depth map generataion.

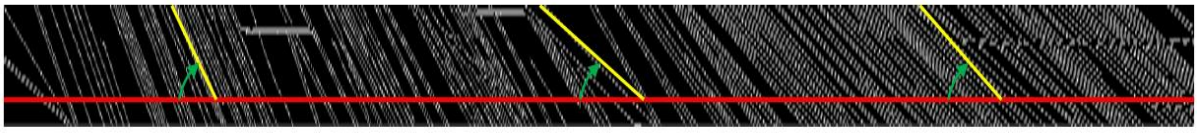

Figure 7. Angular direction measurement from edge detected EPI

5 EPI Depth Estimation Results and Evaluation

The depth estimating equation for an optimal angle value selection is defined in (5), where $A_{Reduced}$ indicates the reduced angular candidates. If we find an optimal angle value from (5), the corresponding depth value can be computed through the equation (6). Depending on the resolution of light field sub-aperture image, the number of EPI is varying. Even though we exclude a texture smearing and occlusion artifact in angular candidate reduction step, initially estimated depth map involves noisy pixel value as exhibited in figure 8.

$$\theta^*(l,\mathrm{P}) = arg \min_{\theta \in A_{Reduced}} C(E_l, P) \tag{5}$$

$$d(o) = -f \cdot \tan(\theta^*(l,p)) \tag{6}$$

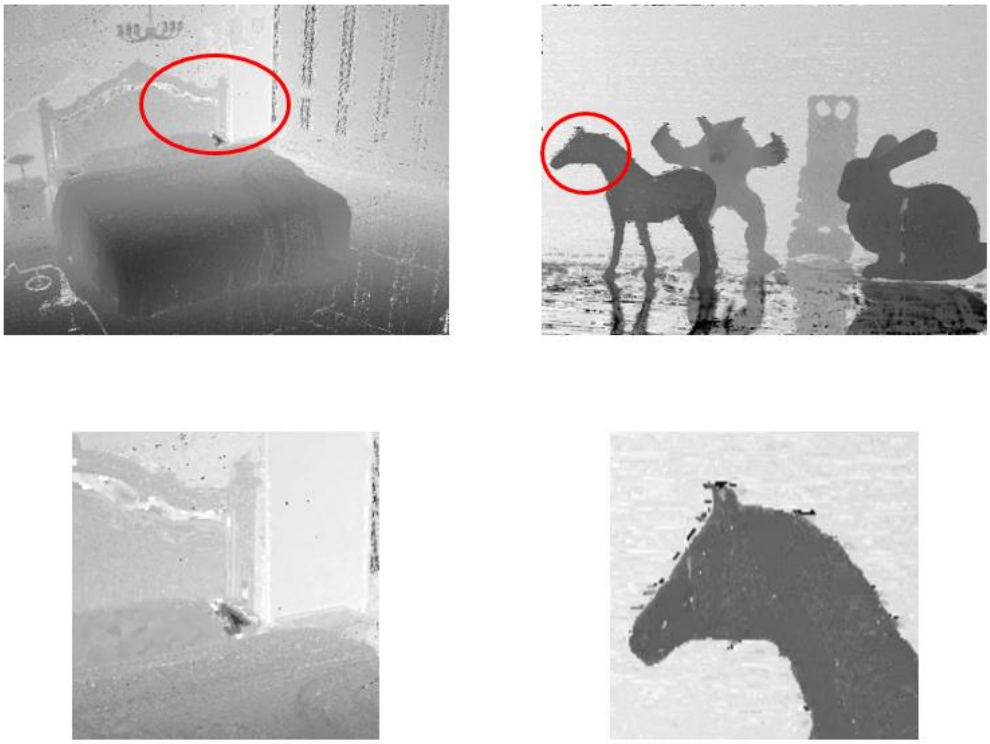

Figure 8. Initial depth map and noise area

To eliminate error regions included in initial depth map, we adopt the weighted median filter [5]. It can remove the error regions while preserving the object boundary regions. As a filter weight, we use joint bilateral filter defined in (7). Where N_w is the number of pixel in specific window kernel($w_{bf} \times w_{bf}$), and σ_s and σ_c adjust spatial and color similarity respectively. In this document, we use constant parameter values for all test sequences: {N_w, σ_s, σ_c}={9, 0.8, 1}.

$$W_{bf} = \frac{1}{N_w} \exp\left(-\frac{|i-j|^2}{\sigma_s^2}\right) \exp\left(-\frac{|I_i - I_j|^2}{\sigma_c^2}\right) \tag{6}$$

Figure 9 exhibits initially estimated depth map, post processed results, and general method (full angular candidate searching) results with 4 different test sequences. As we can notice from that results, post processing reduces the error region near the object boundary area. Also, it fills

object inside hole depth value while preserving depth intensity. In figure 9, the left images are initially estimated depth map result, middle ones are the proposed method post processing results, and right images indicate general method with post processing result.

Figure 9. Initial depth map(left), proposed method + post processing (middle), general method + post processing (right)

To verify the efficiency of reduced angular candidate based depth estimation method, we check the running time of proposed and general method. Table 1 shows the running time comparison results of general EPI based depth estimation method (M1) and reduced angular candidate based method (M2). Running time is measured without post processing due to the fair comparison between general and proposed method. Used all test sequences have same resolution 600×800.

(sec)

	M1	M2
(a) Bedroom	687	281
(b) Livingroom	782	310
(c) Plant	649	233
(d) Sculptures	771	264

Table 1. Running time comparison results

Also, quality of estimated depth maps is evaluated by comparing estimated depth map with ground truth image. We check the bad pixel rate (BPR) for depth map accuracy measurement. If the difference of pixel intensity between estimated depth map and ground truth is larger than 1, we define that pixel is bad pixel. The BPR evaluation results are indicated in table 2. We also check the BPR of general method with proposed method. As shown in table 2, general method generates more accurate depth value than proposed method. To derive fair comparison results, we include post processing to general and proposed method respectively.

	M1	M2
(a) Bedroom	13.2%	14.1%
(b) Livingroom	11.9%	12.3%
(c) Plant	12.2%	13.2%
(d) Sculptures	12.7%	13 %

Table 2. BPR comparison results

6 Conclusion

Depth map generation from light field image is prospective task for view synthesis and generating an immersive video content. Generally, to obtain a depth map from light field image, matching cost computation techniques are widely used. However, that kind of methods have high computational complexity problem. In this document, we propose a light field EPI based depth estimation technique. Instead of inspecting all angular directional candidates for an optimal angle selection, we reduce the angular candidates by inspecting an angular directional tendency. From the experiment results, we verify that the proposed method takes less computational time than general method. However, the accuracy of proposed method results has lower BPR performance than general method.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) Grant funded by the Korea Government(MISP) (No. 2011-0030079)

References

[1] M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, “Depth from Combining Defocus and Correspondence using Light Field Camers,”

[2] Jon Karafin, Didier Doyen, Gauthier Lafruit, “Lytro Cinema Dense Light Field Video Sequence Specification,” ISO/IEC JTC1/SC29/WG11 MPEG2017/M40357, April 2017, Hobart, AU.

[3] https://jpeg.org/downloads/plenoworkshop/

[4] J.H. Mun, and Y.S. Ho, “Light-Field Depth Map Generation and Visualization,” ISO/IEC JTC1/SC29/WG11 MPEG2017/M40289, April 2017.

[5] W. Wu, L. Li, and W. Jin, “Disparity Refinement based on Segment-tree and Fast Weighted Median Filter,” ICIP, pp. 3449-3453, Sept. 2016.