**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 40289**
**April, 2017, Hobart, Australia**

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

## Abstract

This report provides a depth map generation method by using light-field images. Since the light-field camera has dense microlens array, it makes informative depth maps through many sub-aperture images. The generated depth map is visualized through the point cloud in a virtual world. From the point cloud depth map result, we can roughly evaluate the accuracy of generated depth map from light-field images.

# 1 Introduction

Currently, many commercial light-field cameras (Lytro Illum, Raytrix) are widely used in many fields and those are adopted in various image processing researching groups. In order to efficiently use light-field images in multi-view image and 3D image technology, the depth map has to be correctly generated and evaluated.

# 2 Light-Field Camera System

Light field cameras from Lytro and Raytrix are easily available for many customers and research groups. Since the light field camera contains a microlens array in front of the CMOS (Complementary metal-oxide-semiconductor) sensor, it captures many scenes as like a multi-view camera system. In order to capture multi-view images using a multi-view camera system, many numbers of color cameras are needed depending on the number of required views. However, the light field camera can capture multi-view images without the number of cameras. Each light field camera generates different angular images which are the same as different view point images. The Lytro Illum camera which invented in 2015 uses 200,000 microlens array to generate angular images. A dense microlens array receives more light ray information than sparse microlens array. The captured two-dimensional data is analyzed by applying the ray-tracing techniques to compute photographs from acquired light source.
The basic optical configuration of light field camera comprises a photographic main lens, microlens array and photosensor array as indicated in Figure 1. In order to flexibly focus on an object, the main lens will move along the horizontal direction as like a conventional camera system. The ray of light which coming from main lens point on the microlens array. As indicated in Figure 1, microlens separates the ray of light based on the direction angle to create a focused image of main lens. If light field camera adopts a dense microlens array, the micro

lens array will take more various light ray which coming from main lens. Also the wide light ray angle information is transmitted to photosensor array.

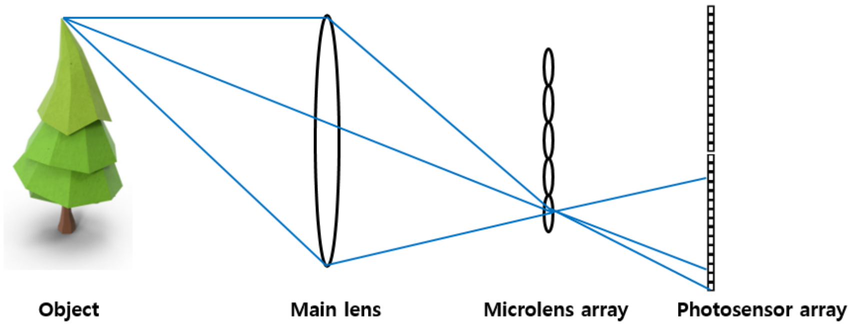


Figure 1. Light field camera system and light ray directions

## 3 Depth Estimation from Light-Field Image

In light field camera system, a number of different viewpoint images are obtained that is known as sub-aperture image. Conventionally used stereo matching algorithm assumes that stereo images have enough baseline distance. However, baseline between the sub-aperture image which taken from a microlens array in light field camera is not enough to apply conventionally used stereo matching algorithms. Since the disparity search range of sub-aperture image is relatively restricted than general stereo image pairs, applying a spatial patch for matching cost computation is not appropriate.
In order to compute a matching cost between sub-aperture images, angular patch based sub-aperture refocusing method is invented [1]. This method shears a 4D light field image in terms of the sub-aperture spatial domain (1).

$$L_\alpha(x,y,u,v) = L\left(x + u\left(1 - \frac{1}{\alpha}\right), y + v\left(1 - \frac{1}{\alpha}\right), u, v\right) \qquad (1)$$

where $L$ is the input light field image and $L_\alpha$ is the refocused light field image. The parameter α is the ratio of the refocused depth to the currently focused sub-aperture image. $x$ and $y$ represent the spatial coordinate, $u$ and $v$ are the angular coordinate. The center of sub-aperture image is located in $(u,v) = (0,0)$ and we compare each sub-aperture image by changing the angular coordinate. Figure 2 shows the shearing result of sub-aperture image in center angular coordinate.

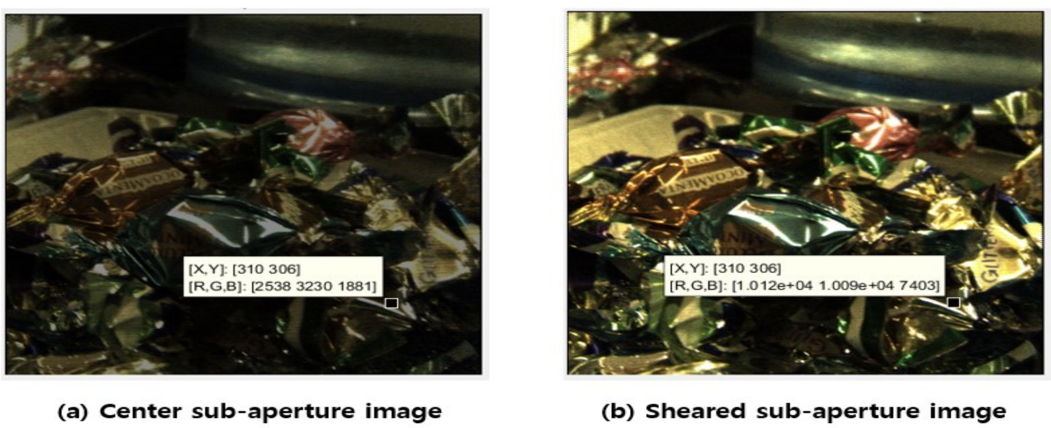


Figure 2. Before(a) and after(b) sub-aperture shearing

Even though the spatial coordinate of sheared image and center sub-aperture image is same, the pixel intensity has a different value as represented in Figure 2. Since the sub-aperture images are obtained by microlens array, sheared image has different angular light ray intensity.
In order to compute a matching cost of sub-aperture images, SAD (sum of absolute differences) and image gradient factor are jointly used. The matching cost volume is defined in (2).

$$
\begin{aligned}
C(x) &= \alpha C_{SAD}(x) + (1-\alpha) C_{Grad}(x) \\
C_{SAD}(x) &= \sum_{u \in V} \sum_{x \in W} \operatorname{argmin}(|L(u,x) - L_{\alpha}(u,x)|) \\
C_{Grad}(x) &= \sum_{u \in V} \sum_{x \in W} [\operatorname{argmin}(\Delta_x(L, L_{\alpha}, x, u)) + \operatorname{argmin}(\Delta_y(L, L_{\alpha}, x, u))]
\end{aligned}
\tag{2}
$$

where $V$ indicate the set of angular coordinate $(u, v)$ and $W$ is a rectangular region which used for matching cost computation. $\Delta_x$ and $\Delta_y$ represent $x$ and $y$ directional gradient image respectively.
From (2) we generate matching cost volumes which need to be refined to improve the accuracy of disparity value. In order to refine the cost volumes, the guided image filter [2] is adopted. The guided image filter smooths the matching cost volume based on the input image (center sub-aperture image). It preserves the edge region and alleviates the scattered unstable matching cost value while performing the filtering. As a sequential procedure, the winner-takes-all (WTA) strategy is applied to determine the minimum matching cost value. The final disparity map is obtained by using a weighted median filter on WTA result. Since the WTA result includes many noises or improper disparity value, the weighted median filter removes that kinds of errors for accurate final disparity map.
Figure 3 exhibits center sub-aperture image which used as a reference sub-aperture image of other sub-aperture image and depth map which corresponding to each sub-aperture image.

Figure 3. Light field center sub-aperture image and generated depth map

# 4 Depth Map Visualization using Point Cloud

Through the point cloud visualization, we can subjectively evaluate how the depth map which obtained in Chapter 3 is accurate. Since the used light field test image does not provide a ground truth image, we have to evaluate the depth map quality subjectively. In the near future, the point cloud data will be used as a supporting tool for virtual viewpoint image rendering and data fusion. Figure 4 exhibits point cloud results based on the estimated depth map. In Figure 4 (a), the flowers are represented mainly two parts. As indicated in the point cloud results, we can notice that the below flower region is more closely located in camera viewpoint than above part. The point cloud of Figure 4 (b) also correctly expressed based on the estimated depth map.

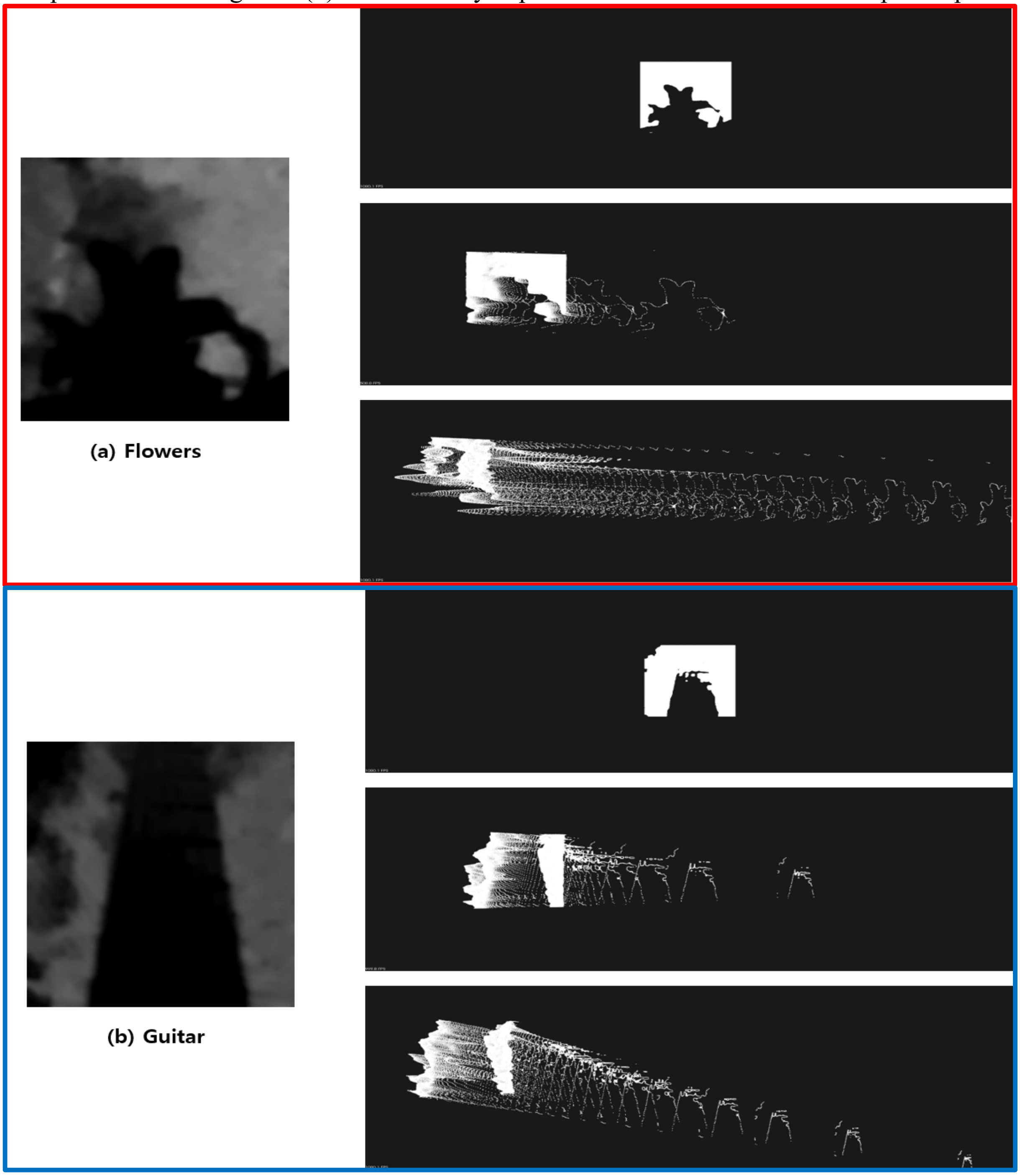


Figure 4. Depth map and visualization result via point cloud

## 5 Conclusion

Generally, a virtual viewpoint image and multi-view are generated by using a depth map which obtained through a stereo matching approach between stereo image pairs. However, super multi-view (over 80 viewpoints) and other camera system have a restriction in terms of the geometrical condition to setup the capturing system. Instead of using a multi-view camera system for depth generation, the light field camera based stereo matching algorithm and view synthesis method have been invented. Especially, we generate a depth map by using a patch-based matching cost computation and guided image filter for matching cost refinement. To subjectively evaluate an accuracy of generated depth map quality, the depth map is expressed by the point cloud data. From the point cloud data, we can notice that the depth map is appropriately generated from a set of sub-aperture images.

## Acknowledgement

This research was supported by the 'Cross-Ministry Giga KOREA Project' of the Ministry of Science, ICT and Future Planning, Republic of Korea (ROK). [GK16C0100, Development of Interactive and Realistic Massive Giga- Content Technology]

## References

[1] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep. CSTR 2005-02, vol. 2, no. 11, 2005.

[2] K. He, J. Sun, and X. Tang, "Guided image filtering," IEEE Trans. on PAMI, vol. 35, no. 6, pp. 1397-1409, Oct. 2012.