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

ISO/IEC JTC1/SC29/WG11
MPEG2007/m14994
October, 2007 Shenzhen, China

Title: **Depth Map Generation for FTV**

Source: **GIST (Gwangju Institute of Science and Technology)**

Author: **Yo-Sung Ho, Sang-Beom Lee, Kwan-Jung Oh, and Cheon Lee**

1. Introduction

Recently, the depth map is in the spotlight as essential data for 3DTV and FTV/FTV with multi-view video [1]-[3]. Nevertheless, the importance of depth map is not recognized as much as that of multi-view video. The multi-view video can be directly obtained from multiple cameras in general, while the multi-view depth map should be calculated from multi-view video artificially. In fact, the depth camera already exists but it is too expensive and it still has some limitations to represent real depth. Thus, artificial multi-view depth map generation is an importation issue.

This document introduces depth map generation for FTV. The depth map generation is a segment-based approach and uses the 3D warping technique. However, there are still remaining problems such as inaccurate object segments and low temporal correlation. This works are still ongoing.

Fig. 1 Segmented Image for ‘Akko&Kayo’ (view 27)

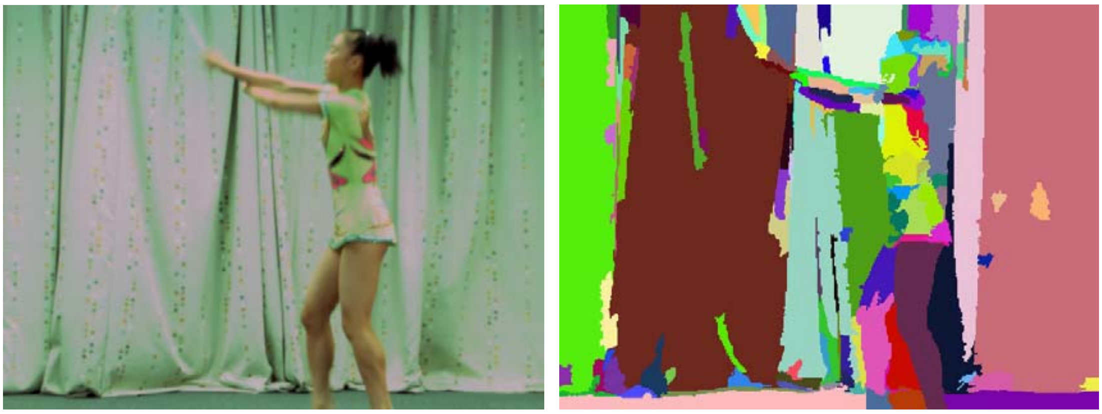

Fig. 2 Segmented Image for 'Rena' (view 50)

2. Depth Map Generation

2.1. Image Segmentation

In this proposal, we utilize the segment-based depth map estimation scheme. In general, the segment-based approach assumes that the whole pixels in one segment have the same depth value. Therefore, the better segmentation results guarantee the higher performance of depth map. In this document, we employ 'Mean Shift-based Image Segmentation' scheme in [4]. Fig. 1 and Fig. 2 show the segmented images for 'Akko&Kayo' and 'Rena'. 'Mean Shift' algorithm shows the more stable results compared to 'Graph-based mage Segmentation' scheme [5].

2.2. Depth Estimation

After the image segmentation, we conduct depth map estimation for each segment. We can directly obtain the depth data using 3D warping technique. To generate the depth image for the center view, both left and right views are considered simultaneously. The most well-known matching functions are SD (Squared intensity Difference) and AD (Absolute intensity Difference). Since these functions are not robust to illumination/color changes between cameras, we use the self-adaptation dissimilarity measure as a matching function [6]. This function adds the absolute gradient difference term to the existing AD and defined as follows

$$C(x,y,d) = (1-\omega)\times C_{MAD}(x,y,d) + \omega \times C_{MGRAD}(x,y,d) \tag{1}$$

ω represents the weighting factor. In Eq. (1), the first and second terms are represented as follows, respectively.

$$C_{MAD}(x,y,d) = \frac{1}{M}\sum_{(x,y)\in S_k}\left|I_1(x,y) - I_2(x',y')\right| \tag{2}$$

$$C_{MGRAD}(x,y,d) = \frac{1}{M}\sum_{(x,y)\in S_k}\{\left|\nabla_x I_1(x,y) - \nabla_x I_2(x',y')\right| + \left|\nabla_y I_1(x,y) - \nabla_y I_2(x',y')\right| + \left|\nabla_{-x} I_1(x,y) - \nabla_{-x} I_2(x',y')\right| + \left|\nabla_{-y} I_1(x,y) - \nabla_{-y} I_2(x',y')\right|\} \tag{3}$$

M represents the number of pixels in the segment S_k and (x', y') represents the position of the warped left or right view. Fig. 3 shows the final depth images for 'Akko&Kayo' view 27 and 'Rena' view 50.

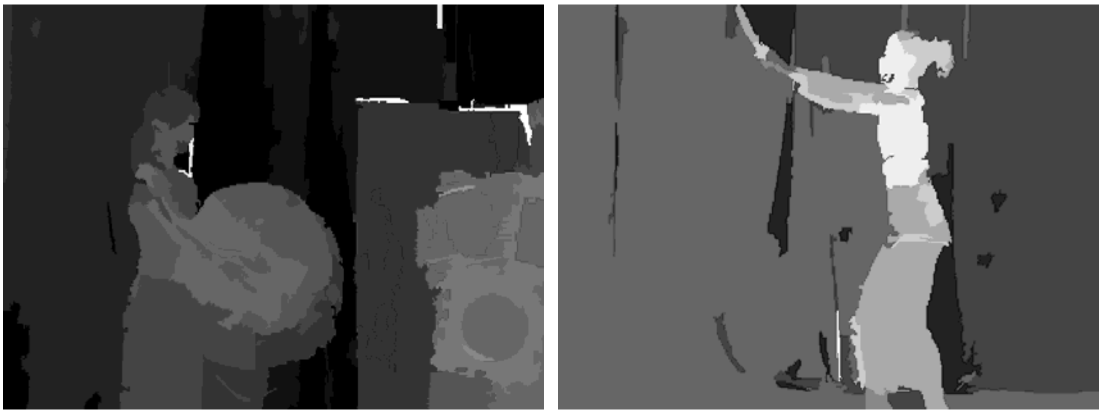

Fig. 3 Final Depth Images (left: 'Akko&Kayo' view 27, right: 'Rena' view 50)

3. Summary

In this document, depth map generation for FTV was reported. The image segmentation and 3D warping techniques were used for depth map generation. However, there are still remaining problems such as inaccurate object segments and low temporal correlation. This works are still ongoing. The experiment demonstrated the final depth map.

4. Acknowledgements

This work was supported in part by the Information Technology Research Center (ITRC) through the Realistic Broadcasting Research Center (RBRC) at Gwangju Institute of Science and Technology (GIST), and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

5. References

[1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and Thomas Wiegand, "3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards", ICME 2006, IEEE International Conference on Multimedia and Expo, Toronto, Ontario, Canada, July 2006.

[2] A. Smolic, and P. Kauff, "Interactive 3D Video Representation and Coding Technologies", Proc. of the IEEE, Special Issue on Advances in Video Coding and Delivery, vol. 93, no. 1, Jan. 2005.

[3] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "MVC: Experiments on Coding of Multi-view Video plus Depth," ITU-T and ISO/IEC JTC1, JVT-X064, Geneva, CH, July 2007.

[4] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," IEEE: PAMI, vol. 24(5):603–619, May 2002.

[5] A. Klaus, M. Sormann and K. "Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure," ICPR 2006, vol. 3, pp. 15-18, 2006.

[6] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Graph-Based Image Segmentation," International Journal of Computer Vision, vol. 59(2), pp. 167–181, 2004.