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

ISO/IEC JTC1/SC29/WG11
MPEG2008/m15119
January, 2008 Antalya, Turkey

Title: **Segment-based Multi-view Depth Map Estimation for FTV**

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

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

1. Introduction

The depth map is currently in the spotlight as essential data for 3DTV and FTV/FTV with multi-view video [1]-[3]. Nevertheless, the importance of the depth map is not recognized as much as that of multi-view video. In order to reconstruct intermediate images at virtual viewpoints, we need the depth information. The multi-view video can be obtained from multiple cameras directly in general, while the multi-view depth map should be calculated from multi-view video.

Many works have been carried out for the acquisition of 3D depth information. As one of the passive 3D depth sensing methods, the stereo matching method is well-known. The task of stereo matching is the computation of 3D data from 2D input images. Recently, many algorithms start to use multi-view images to obtain more accurate depth information, instead of stereoscopic images.

This document proposes a method of segment-based multi-view depth map estimation for FTV. The whole process of depth map estimation is based on segments and we use the 3D warping technique. However, there are still remaining problems, such as low temporal correlation.

2. Depth Map Estimation

2.1. Image Segmentation

In this section, we describe a segment-based depth map estimation scheme. In general, the segment-based approach assumes that all 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 a ‘mean shift-based image segmentation’ scheme [4] before depth estimation. Figure 1 shows segmented images for ‘Akko & Kayo’. The mean shift algorithm shows the more stable results, compared to the ‘graph-based image segmentation’ scheme [5].

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

2.2. Initial Depth Map Estimation

After an image is segmented, we conduct depth map estimation for each segment. We can obtain the depth value using the 3D warping technique directly. In order to estimate the depth image for the center view, we consider both left and right views simultaneously. The well-known matching functions for depth estimation are SD (squared intensity differences) and AD (absolute intensity differences). Since these functions are not robust to illumination and color changes between cameras, we use a self-adaptation dissimilarity measure as the matching function [6]. This function adds the mean absolute gradient difference term to the existing AD. It is defined by

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

where ω represents a weighting factor. In Eq. (1), the first and second terms are represented by

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

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

where M represents the number of pixels in the segment S_k and (x',y') represents the position of the warped left or right view. Figure 2 shows the initial depth map for ‘Akko & Kayo’ view 27.

2.3. Refinement Using Belief Propagation

The belief propagation (BP) algorithm has been applied successfully for stereo matching. BP is an iterative inference algorithm that propagates messages in the network [7]. While the conventional BP approach is a pixel-based method, we utilize a segment-based BP algorithm in the refinement process. As shown in Fig. 2, most erroneous depth values are removed and we can finally obtain the refined depth map.

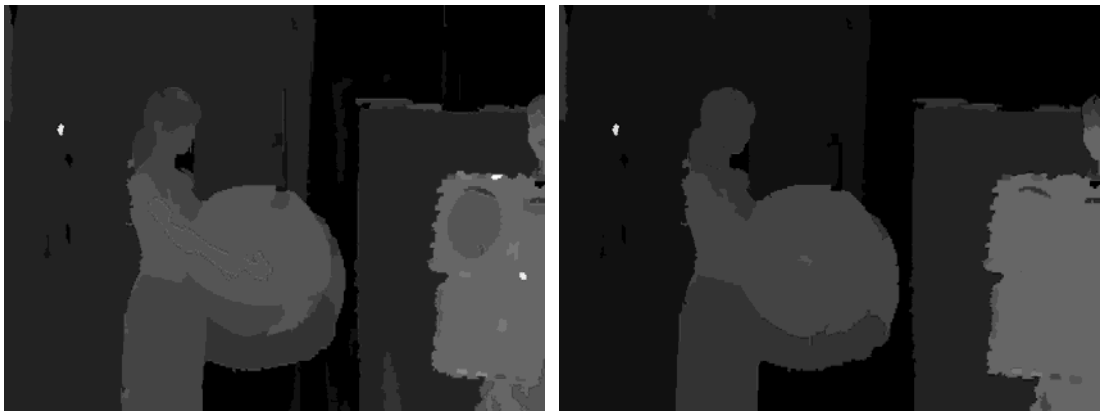

Fig. 2 ‘Akko & Kayo’ view 27 (left: initial depth map, right: refined depth map)

3. Summary

In this document, we described an algorithm of segment-based multi-view depth map estimation for FTV. Image segmentation and 3D warping techniques are used for depth map estimation. However, there are still remaining problems, such as low temporal correlation in the depth map sequence.

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.

References

[1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, “3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards”, IEEE International Conference on Multimedia and Expo (ICME),, 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] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Graph-Based Image Segmentation," International Journal of Computer Vision, vol. 59(2), pp. 167–181, 2004.

[6] 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.

[7] S. Jian, Z. Nan-Ning and S. Heung-Yeung, "Stereo matching using belief propagation," IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 25. No. 7, July 2003.