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/M15594
July 2008, Hannover, Germany

Source: GIST (Gwangju Institute of Science and Technology)

Status: Proposal

Title: Enhancement of Temporal Consistency for Multi-view Depth Map Estimation

Author: Sang-Beom Lee and Yo-Sung Ho

1. Introduction

Recently, many researchers have concentrated on the acquisition of depth information. Especially, the depth map is essential data for 3DTV using multi-view video [1]-[3]. 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 computed by using multi-view video.

As one of the passive 3D depth sensing methods, the stereo matching algorithm is well-known. The task of stereo matching is the computation of 3D data from 2D stereoscopic images. Since two images are obtained from slightly different perspectives, the position of a pixel in one view is horizontally displaced in the other view.

We can expand the stereo matching algorithm by adding more views. If we utilize three or more views, we can obtain more accurate depth map. However, since the conventional algorithms performed for each frame separately, we notice that the results have low temporal consistency. In other words, the resultant depth maps have the low consistency of depth values at the same region but in different time.

In this document, we describe the depth map estimation scheme enhancing temporal consistency. The whole process of depth map estimation is based on segments and we use a temporally weighted matching function to consider the previous depth value.

2. Enhancement of Temporal Consistency

In this section, we describe a segment-based depth map estimation scheme [4]. The whole procedure is the same as the previous scheme except for the matching function. After we divide the center image into several segments, we aggregate matching costs for each segment using 3D warping and segment-based matching techniques. In the refinement process, we apply a segment-based belief propagation algorithm.

2.1. Modified Matching Function

When we calculate matching costs, the well-known matching functions are MAD (mean absolute difference) and MSE (mean square error). However, since these functions are not robust to illumination changes between views, we cannot guarantee the good result.

Figure 1 shows the initial depth maps for the different matching function. As shown in Fig. 1(a) and Fig. 1(b), three views have the different average luminance, especially at the background area. Since the lighting condition is not constant for the different views, the initial depth map using MAD has the poor result. On the other hand, if we use MGRAD as a matching function the result can be robust to the different lighting condition.

(a) Original image (View 4) (b) Reference images (View 3, 5)

(c) Initial depth map (MAD) (d) Initial depth map (MGRAD)

Fig. 1. Initial depth maps for the different matching function

Therefore, we add MGRAD (mean gradient absolute difference) to the existing MAD. The matching function is defined by

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

where ω represents a weighting factor. In Eq. (1), ω can be determined by the difference of average luminance between two views. It is defined by

$$\omega = \left| DC_C - DC_{ref} \right| \tag{2}$$

where DC_c and DC_{ref} represent the average luminance for the center and the left or right view, respectively. Therefore, the larger the difference of average luminance is, the more we reflect the gradient maps for aggregating the matching costs.

2.2. Temporally Weighted Matching Function

As mentioned above, since the conventional depth estimation methods separately estimate the depth value for frame by frame, the resultant depth sequence has a low temporal consistency. Furthermore, even if the refinement is performed, there still exist errors. Therefore, we describe a new matching function that refers to the depth value of the previous frame when estimating the depth of the current frame. The matching function adds the weighting function considering the depth value of the previous frame. The temporally weighted matching function is defined by

$$C(x,y,d) = C_{MAD}(x,y,d) + \omega * C_{MGRAD}(x,y,d) + C_{temp}(x,y,d) \tag{3}$$

$C_{temp}(x,y,d)$ can be defined by

$$C_{temp}(x,y,d) = \lambda \,|\, d - D_{prev}(x,y) \,| \tag{4}$$

where λ represents the slope of the weighting function and $D_{prev}(x,y)$ represents the previous depth value.

Figure 2 illustrates the graph of matching functions. In Fig. 2, the dotted line represents the previous matching function and the chain line represents the weighting function considering the previous depth value. The solid line represents the temporally weighted matching function.

As shown in Fig. 2, in case of the previous depth value is around 70, the probability that the current depth value is around 70 is very high. Therefore, we apply the weighting function that increases the matching score when the distance between the current and the previous depth value are larger. Finally, we obtain the depth sequence with high temporal consistency and low error regions since the current depth value that is similar to the previous depth value is determined.

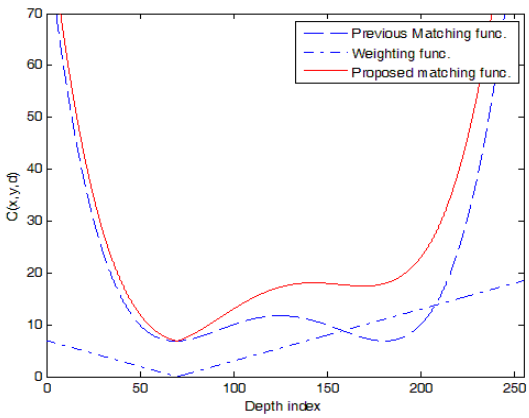


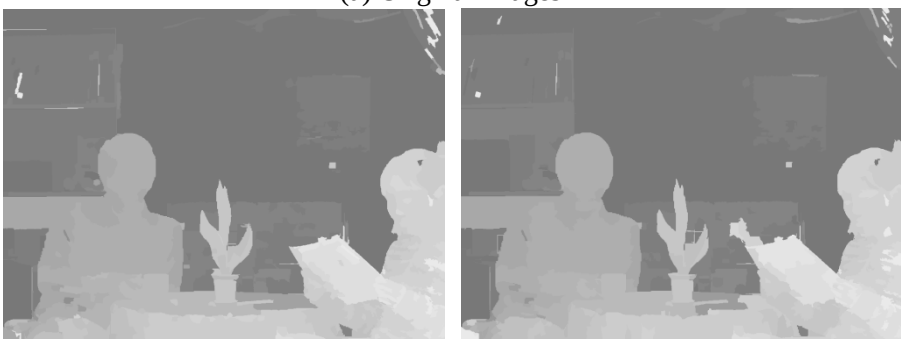
Fig. 2. Graph of the weighted matching function

2.3. Depth Sequence Estimation Results

Figure 3 and Fig. 4 demonstrate the depth estimation results for ‘Newspaper’ provide by Gwangju Institute of Science and Technology [5] and ‘Pantomime’ provided by Nogoya university [6]. These results are obtained by applying the temporally weighted matching

function for t=6 and t=7. As shown in Fig. 3 and Fig. 4, the proposed scheme efficiently removes the errors in the background area.

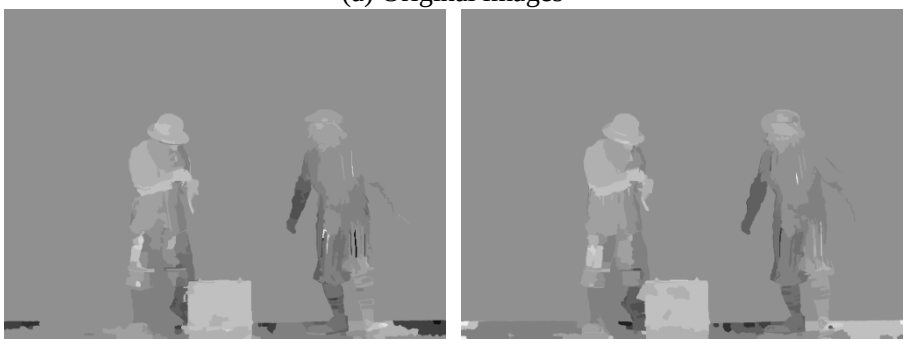
(a) Original images



(b) Depth maps

Fig. 3 Depth sequence for ‘Newspaper’ (view 4)

(a) Original images



(b) Depth maps

Fig. 4 Depth sequence for ‘Pantomime’ (view 4)

3. Summary

In this document, we described a multi-view depth map estimation scheme that modifies the matching function to enhance the resultant reliability and the temporal consistency. We adopt the weighting factor for compensating the different illumination environment. Also, we applied a temporally weighted matching function to consider the previous depth when calculating the matching score for the current frame. Experimental results have showed that the proposed scheme efficiently increased the temporal consistency.

Acknowledgements

This work was supported in part by ITRC through RBRC at GIST (IITA-2008-C1090-0801-0017).

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] Y. Ho, S. Lee, K. Oh and C. Lee, "Depth Map Generation for FTV," ISO/IEC JTC1/SC29/WG11, M14994, Oct. 2007.

[5] Y. Ho, E. Lee and C. Lee, "Multiview Video Test Sequence and Camera Parameters," ISO/IEC JTC1/SC29/WG11, M15419, April 2008.

[6] M. Tanimoto, T. Fujii and N. Fukushima, "1D Parallel Test Sequences for MPEG-FTV," ISO/IEC JTC1/SC29/WG11, M15378, April 2008.