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

ISO/IEC JTC1/SC29/WG11
MPEG2010/M17223
January 2010, Kyoto, Japan

Source: GIST (Gwangju Institute of Science and Technology)
Status: Report
Title: Results of EE1 on 'Cafe' Sequence
Author: Cheon Lee, Minkoo Kang, and Yo-Sung Ho

1. Introduction

This document reports experimental results of depth estimation for "Café" sequence as a response to EE1 of 3D video coding [1]. "Café" sequence is distributed by GIST on the end of November 2009. It consists of five Full-HD videos having good depth impressions. Satisfying the requirement of EE in N10925, we conducted depth estimation using the reference software version DERS 5.0 and synthesized intermediate images using them. We evaluated the objective quality by comparing differences between the original view and the synthesized view.

2. Settings of Depth Estimation Software

"Café" sequence is a new sequence captured by five Full-HD cameras. Fig. 1 shows still images of each view. Table 1 describes the settings of DERS. Although the semi-automatic depth estimation method shows the best performance, we did not use it because we have no auxiliary data. Therefore, we used the automatic method including "half-pel" and "3x3 block matching" methods. Remarkable point of this sequence is the wide range of disparity from 30 to 90.

(a) View1 (b) View2 (c) View3 (d) View4 (e) View5

Fig. 1. Still images of 'Café' sequence

In order to fulfill the requirement of EE1 in N10925, we estimated depth maps of three views such as "view2," "view3," and "view4" where the left most view is named as "view1". The "view2" is estimated from "view1" and "view3". The other views are generated in the same manner. The number of the estimated frames per a view is 300.

There are two configurations: 2-view and 3-view. In 2-view case, we synthesized "view3" using both the depth maps of "view2" and "view4" as described in Fig. 2. In order to make a stereo video of "2-view" case, we chose "view2" as an original view. In "3-view" case, we synthesized two intermediate views such as "view2.75" and "view3.25". Synthesized images of "view2.75" are generated by using depth videos of "view2," "view3". In the same manner, we generated the images of "view3.25". Using both generated intermediate views, we make a stereo video.

Table 1. Configuration of DERS

Disparity Range	[30, 90]
Search Range	[30, 90]
Precision	Half-pel
Smoothing coefficient	1.0
Matching block	3x3
Depth estimation mode	Automatic
Temporal Enhancement	Off
Image Segmentation	Off

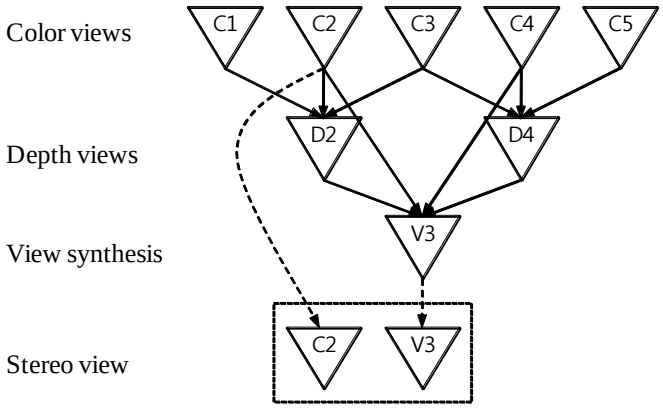


Fig. 2. Procedure of depth estimation and view synthesis for 2-view configuration

In addition, we have conducted experiments for the 3-view configuration. Unlikely with the 2-view configuration, this configuration has three color views and their associated depth videos. Using those data, we generated two synthesized views as desribed in Fig. 6(left). The camera parameters of two intermediate viewpoints are calculated by a linear interpolation. For the 9-view display, we need to prepare 9 viewpoint images, but we can generate only three depth views because of the limitation of depth estimation software. By this reason, we could synthesize only 7 views among the viewpoints described in N10925. Therefore, two viewpoints, view2 and view4, are inlcuded with ogirinal data in the 9-view display data.

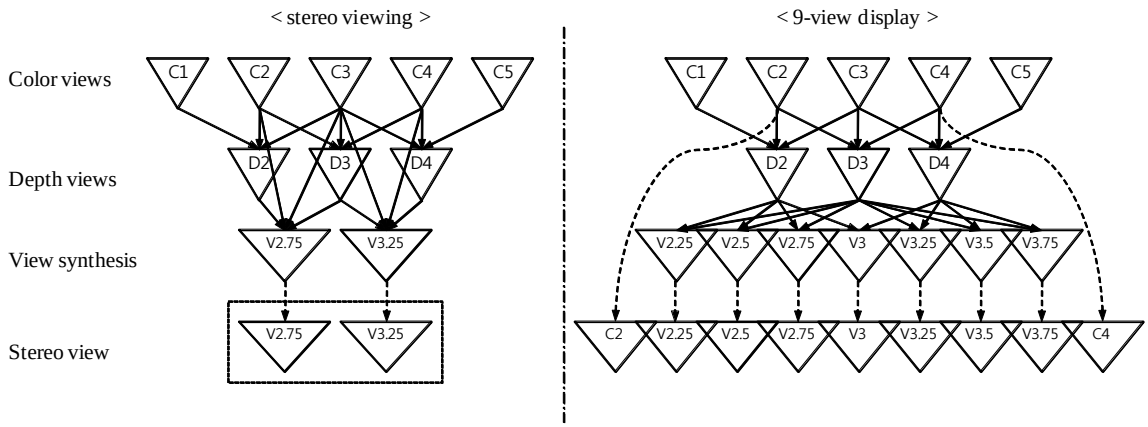


Fig. 3. Procedure of depth estimation and view synthesis for 3-view configuration

3. Results of Depth Estimation

3.1. Overall quality of the generated depth map

Fig. 3 shows results of depth estimation. As shown in Fig. 4(b), the representation of depth is clear and acceptable enough. The flickering depth values in time domain still exists since we did not use the "temporal enhancement" method, but it is negligible.

(a) Color image (b) Estimated depth image

Fig. 4. Color image and estimated depth image of 'view2'

We synthesized the intermediate images of "view3" using two depth maps "view2" and "view4" and associated color images. To compare the quality of the intermediate images, we calculated the PSNR values. It was 34.89 dB for luma component. Fig. 5 shows all PSNR values.

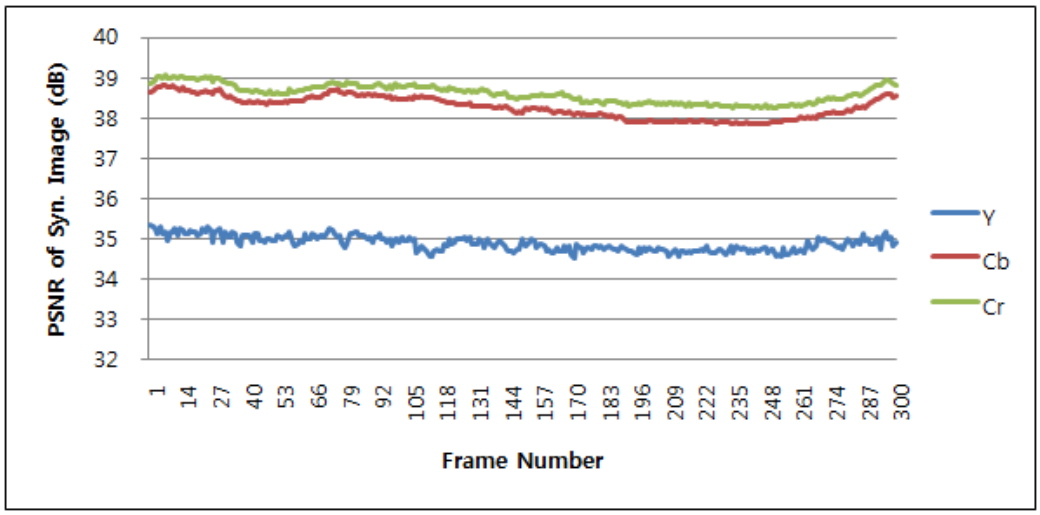
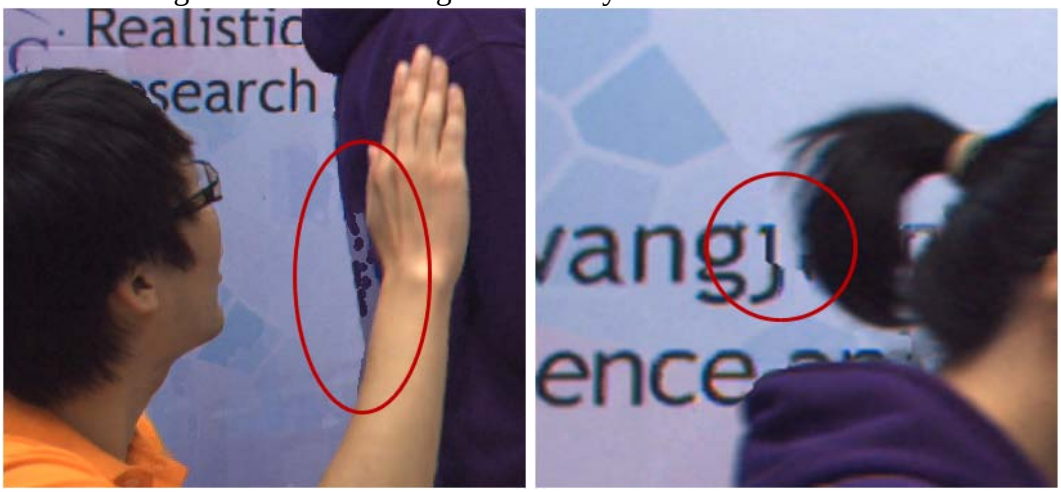


Fig. 5 PSNR values of the synthesized images

3.2. Visual artifacts in the synthesized image

Overall quality of the synthesized images is quite acceptable. Comparing to the current test sequences, this sequence shows clear depth boundaries and shapes. Nevertheless, we found some visual artifacts in the synthesized images. Fig. 6 shows two kinds of artifacts. The first is synthesis errors on the background objects as shown in Fig. 6(a). This artifact appeared throughout the whole frames. We think this does not come from the depth map, but false view synthesis method. The other artifact is shape distortion around object boundaries as shown in Fig. 6(b). When the depth boundaries are not overlapped with that of the color image, this shape distortion occurs. Thus, if the depth estimation is perfect for the depth discontinuity region, this problem would not appear. This problem also occurs throughout the whole images like the synthesis errors.



(a) Synthesis errors on background objects (b) Shape distortions

Fig. 6. Visual artifacts

3.3. Preparation of viewing test

Following the description of EE1, we prepared the test materials of EE1. We made two files for two configurations in “avi” format. The length of the file is 300.

4. Conclusion

In this document, we reported that the performance of the depth estimation on newly distributed “Café” sequence. Overall quality of the generated depth images looks satisfying although the visual artifacts still appear. In conclusion, “Café” sequence is acceptable as a reference sequence. We recommend including this sequence to the test sequence of EE4.

5. Acknowledgements

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency" NIPA-2010-(C1090-0902-0017)

6. References

[1] ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding,” N10925, October 2009.