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

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

1. Introduction

This document reports experimental results of the depth estimation on 'Cafe' sequence in response to EE1 of 3D video coding [1]. In order to obtain depth videos, we performed the same process as we described in M17488 [2]; depth estimation using DERS 5.0 assisted by semi-automatic data, manual refinement of erroneous parts. The objective and subjective quality evaluation was done by original views and synthesized views.
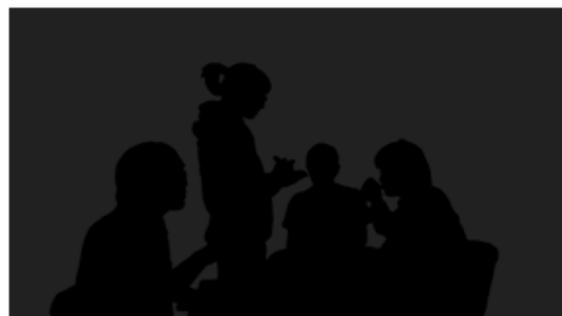
2. Results of Depth Estimation

2.1. Semi-automatic Mode

As described in M17488, we exploited the semi-automatic mode implemented by Nagoya University. Although the semi-automatic depth estimation gives us good quality of depth videos compared to the automatic one, still there exist erroneous regions when the foremost man raises his hand and the woman turns toward the foremost man. Therefore, for more accurate depth estimation, we made additional manual maps by Adobe Photoshop CS4. In other words, the supplementary data, manual disparity map, manual edge map, and manual static map, are obtained for the 1st, 50th, 81st, 91st, 100th, 150th, 183rd, 200th, 250th, and 271st frame of 2nd, 3rd, 4th viewpoint. Figure 1 shows manual disparity maps for the 81st and 183rd frame of 3rd viewpoint.

(a) Original color image (91st frame)   (b) Manual disparity map (91st frame)

(c) Original color image (183rd frame)

(d) Manual disparity map (183rd frame)

Figure 1. manual disparity maps for the 81st and 183rd frame of 3rd viewpoint

2.2. Depth Video Refinement

After performing semi-automatic depth estimation, we refined depth videos degrading rendering quality. Figure 2 shows results of view synthesis. As shown in Fig. 2(a) and Fig. 2(b), erroneous regions near hand cause rendering errors. Therefore, we manually modified the depths. Figure 2(c) shows the refined depth map. As a result, the rendering quality of synthesized image as shown in Fig. 2(d) was increased.

(a) Depth map (4th view)

(b) Synthesized view (3rd view)

(c) Refined depth map (4th view)

(d) Refined synthesized view (3rd view)

Figure 2. Results of view synthesis

2.3. Results of View Synthesis

We performed view synthesis for the 3rd view using two color and depth videos for the 2nd and 4th view. Then, we compared the original view to the synthesized view in terms of PSNR. Figure 3 shows the rendering quality comparison. The average values were 34.1574dB for the previous result and 34.6972dB for the refined result.
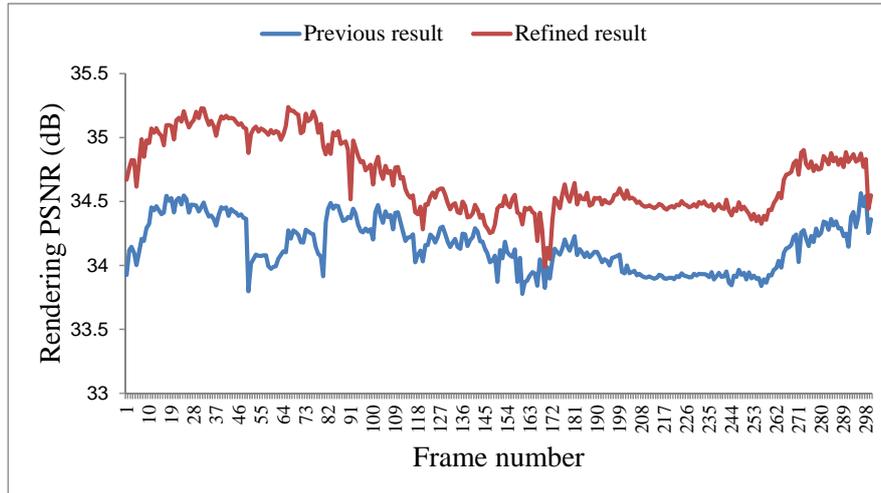

Figure 3. Rendering quality comparison

The final depth map and the synthesized image for the 3rd view are shown in Fig. 4. From the results, we noticed that the refined depth videos guaranteed the good rendering quality of synthesized views.

(a) Depth map (91st frame)

(b) Synthesized image (91st frame)

(c) Depth map (251st frame)

(d) Synthesized image

Figure 4. Results of view synthesis

3. Conclusion

In this document, we reported the depth estimation and view synthesis results for 'Cafe' sequence. In conclusion on EE1, the depth videos guarantee good quality for the view synthesis in terms of average PSNR of the synthesized views and the subjective quality. We are ready to demonstrate these results for the viewing test in this meeting.

4. Acknowledgements

5. References

[1] ISO/IEC JTC1/SC29/WG11, "Description of Exploration Experiments in 3D Video Coding," N11095, Jan. 2010.
[2] ISO/IEC JTC1/SC29/WG11, "Results of EE1 on 'Cafe' Sequence," M17488, April 2010.