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/M18512
October 2010, Guangzhou, China

Source: GIST (Gwangju Institute of Science and Technology)
Status: Report
Title: 3DV EE1 Results on Café
Author: 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 several steps; depth estimation using DERS 5.0 assisted by semi-automatic data, bilateral depth filtering, manual refinements. The resultant depth data have improved the accuracy.

2. Results of Depth Estimation

In the last meeting, we had presented the refined depth data in response to EE1, but those data generated visual artifacts in the synthesized images. As a result, the evaluation on Café sequence was ‘Slight artifacts’. To solve this problem, we modified them with additional steps as described in Fig. 1. In principal, we generate the initial depth map using DERS 5.0 with semi-automatic method, then we perform a post-processing for depth data. Next, we refine some regions inducing visual artifacts by manual manner.

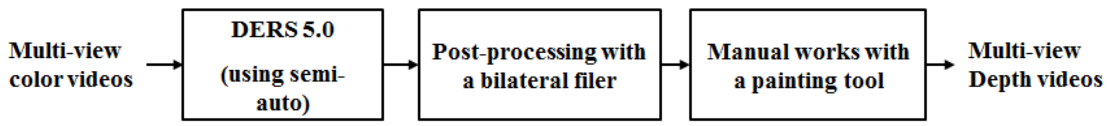
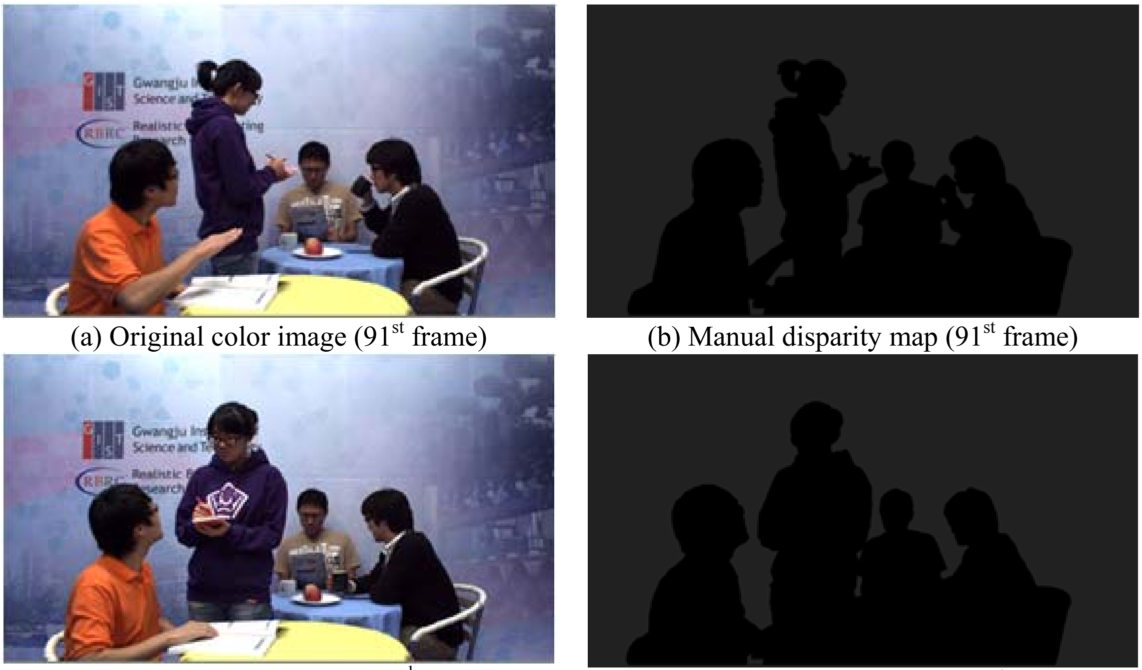


Fig. 1. Depth generation for Café sequence

2.1. Semi-automatic Mode

As described in M17788 [2], we exploited the semi-automatic mode proposed 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 a painting tool. 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)

Fig. 2. Semi-automatic manual data for the 81st and 183rd frame of 3rd viewpoint

2.2. Post-processing with a Bilateral Filter

As a post-processing for depth data, we employed a bilateral filter designed for a depth [3]. Using both the color and depth image, we obtain a refined depth image. Let the reference color image be *I*, the generated depth image be *D*, and bilateral filter radius be *r*. For a typical pixel p = {x, y}, assume $\vec{u}_p$ = {x − r, · · ·, x + r}, $\vec{v}_p$ = {y − r, · · ·, y + r}, we determine the refined depth value *C* as:

$$C(x,y) = \underset{d \in \vec{d}_p}{\arg\min} \frac{\sum_{u\in\vec{u}_p}\sum_{v\in\vec{v}_p} W(u,v)\ \ C(u,v,d)}{\sum_{u\in\vec{u}_p}\sum_{v\in\vec{v}_p} W(u,v)} \tag{2}$$

where

$$W(u,v) = \exp(\ \frac{\left\| I(x,y), I(u,v) \right\|^2}{2\sigma_R^2})\ \exp(\ \frac{(x\quad u)^2 + (y\quad v)^2}{2r^2}) \tag{3}$$

$$C(u,v,d) = \min(\lambda L, | D(u,v) - d |) \tag{4}$$

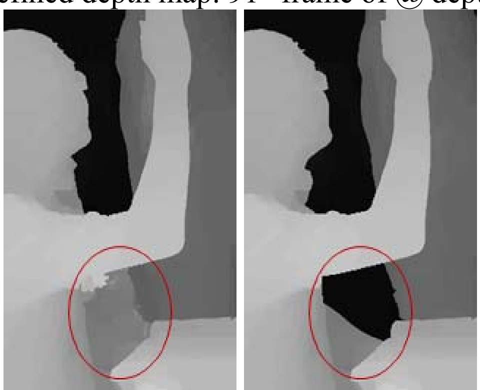
In the experiments, we use the bilateral filter radius r=5 and the standard deviation σ_R=10. Figure 3 shows the filtered depth map. Object boundaries of the refined depth map are clearer then that of the DERS 5.0.

Fig. 3. Post-processing using bilateral filter: depth map by DERS 5.0(left), refined depth map(right)

2.3. Depth Video Refinement

Although the depth data are refined by the post-processing step, some erroneous depth values are exist in the depth map which induces visual artifacts. To remove them from the depth map, we refined depth map manually using a painting tool. Figure 4 shows results of manual depth refinement. As shown in Fig. 4(a) and Fig. 4(b), erroneous depth values are removed clearly.

(a) Refined depth map: 91st frame of @ depth_café_2



(b) Refined depth map: 132nd frame of @ depth_café_2

Fig. 4. Manually refined depth data

2.4. Results of View Synthesis

We performed view synthesis for the 3rd view using two color and depth videos of the 2nd and 4th view. Then, we compared the original view to the synthesized view in terms of PSNR. In order to avoid the common hole problem as reported in our technical contribution [4], we used the modified hole filling method with the boundary noise removal method. Figure 5 shows the rendering quality comparison. The average value of the previous depth data reported in 93rd meeting is 33.04 dB and the average value of the refined depth data is 33.14 dB.

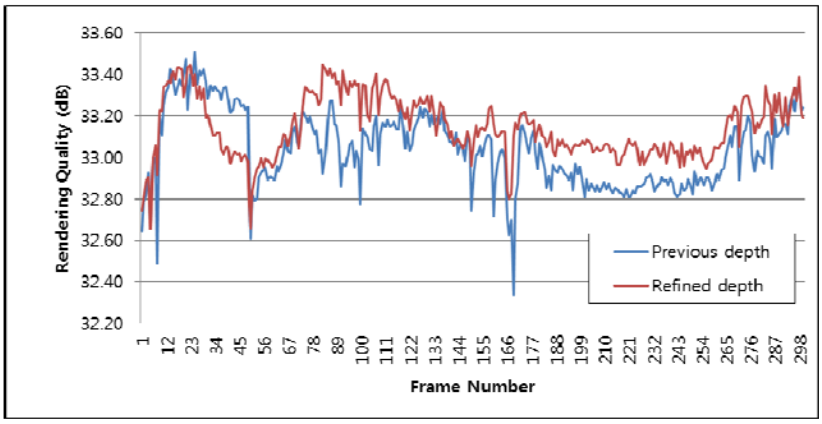


Fig. 5. Rendering quality of the synthesized images

The final depth map and the synthesized image for the 3rd view are shown in Fig. 6. 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

Fig. 6. 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

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-(C1090-1011-0003)).

5. References

[1] ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding,” N11477, July 2010.
[2] ISO/IEC JTC1/SC29/WG11, “Results of EE1 on ‘Cafe’ Sequence,” M17788, July 2010.
[3] Q. Yang, L. Wang, and N. Ahuja, “A constant-space belief propagation algorithm for stereo matching,” in *Proc. of CVPR*, pp. 1–8, 2010.
[4] ISO/IEC JTC1/SC29/WG11, “Common-hole Filling for Boundary Noise Removal in VSRS,” m18514, Oct. 2010.