Eye Contact Technique Using Depth Image Based Rendering for Immersive Videoconferencing

Yo-Sung Ho and Woo-Seok Jang
School of Information and Communications
Gwangju Institute of Science and Technology (GIST)
123 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712 Republic of Korea
Email: {hoyo, jws}@gist.ac.kr

***Abstract*— The lack of eye contact severely hinders effective communication in video conferencing. This problem is caused by the disparity between locations of the subject and the camera. 3D video processing techniques, including depth estimation and virtual view synthesis, can be used to correct the eye gaze. They enable an economic and effective setup that can be applicable to large display monitors. In this paper, we have designed a set of color and depth cameras to reduce occlusion regions and improve the depth precision in the less-detailed region. The system setup overcomes some inherent problems of depth sensing.**

***Keywords*— Depth image based rendering, eye contact, videoconferencing.**

I. INTRODUCTION

Videoconferencing is the conduct of conference between two or more participants at different locations using a set of telecommunication systems to transmit audio and video data. Although several videoconferencing systems have been developed, most systems lose eye contact by upper positioning of cameras. This creates some kind of disconnected feeling, reducing effectiveness of interaction. Thus, the eye contact problem is considered as one of the most important issues in the videoconferencing system [1].

Over the past decades, a variety of eye contact techniques have been proposed. However, they require complex hardware configurations for their performances as well as high cost for the system setup. In order to overcome these drawbacks, we can use 3D video processing technologies. In order to solve the eye contact problem, we can use depth estimation and view synthesis methods [2].

Depth information can be estimated by several methods: active or passive sensor-based and hybrid sensor fusion depth estimation. Active sensor-based methods employ physical sensors to measure the depth data directly, while passive sensor-based methods extract the depth data from 2D images indirectly. Hybrid fusion methods combine both approaches to make up for the weakness of each one [3]. Active sensor-based and hybrid fusion methods require expensive physical sensors, such as depth cameras.

Recently, small and cheap active sensors, such as Kinect depth camera, are introduced without high cost burden [4]. Although the Kinect depth data possesses low accuracy, compared to more expensive depth cameras, due to sensor noises and occlusion regions, we can reduce the cost burden for 3D content production. Our hybrid fusion method also utilizes a Kinect depth camera.

II. GAZE CORRECTION

A. System Overview

Figure 1 shows the overall system that includes two stereo camera sets and one Kinect depth camera. In order to capture texture information from the center view, we set up the stereo camera sets on the top and bottom of the display. Each stereo camera set is placed with a small baseline to reduce occlusion regions, which make it difficult to find corresponding points. Only one camera from each stereo camera set is used to generate an eye gaze-corrected image.

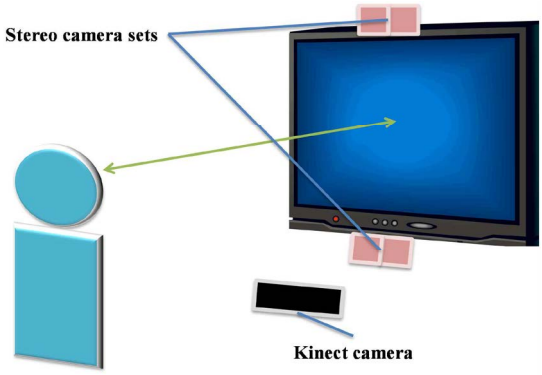


Fig. 1. System setup for eye contact

When we employ two or more cameras, we need camera calibration and color correction. Here we apply camera calibration to obtain camera parameters [5] and histogram matching to adjust color representation [6].

B. Depth Estimation

Kinect based eye contact techniques [1] do not provide high performance 3D information due to inherent problems of the device. We use the Kinect depth data to supplement the crude depth map obtained by stereo matching. After the Kinect depth data is mapped to its corresponding position of the color view by 3D image warping [7], we interpolate the low resolution depth map to the color resolution. The proposed depth estimation is based on the global stereo

matching defined by maximum a posteriori Markov random field (MAP-MRF). The upsampled Kinect depth is utilized as the additional evidence for the energy function of stereo matching. The data enhances the precision and accuracy of the final depth map by allowing large depth variations.

C. *View Synthesis for Eye Contact*

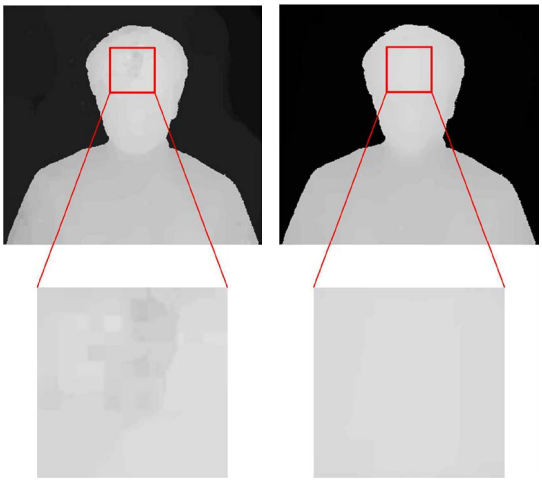
View synthesis is used to generate an eye contact image. We calculate camera parameters at the eye contact position. Intrinsic parameters are adjusted by computing averages of the focal length and the principal point values from the top and bottom cameras used for view synthesis. Intermediate rotation parameters are calculated by average of Euler angles extracted from the rotation matrix of the top and bottom views [8]. The translation parameters are determined by the camera centers of two views and intermediate rotation parameters.

In order to generate an eye gaze corrected image, we project the whole texture information of the original images to the target position using the depth information. Consequently, blending of the texture information and hole filling operations are performed.

III. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed eye gaze correction system, we have compared our proposed method to a stereo matching based method. The resolution of the color image is 1920×1080 pixels while that of the Kinect depth data is 640×480 pixels. The Kinect depth data is represented by 16-bit.

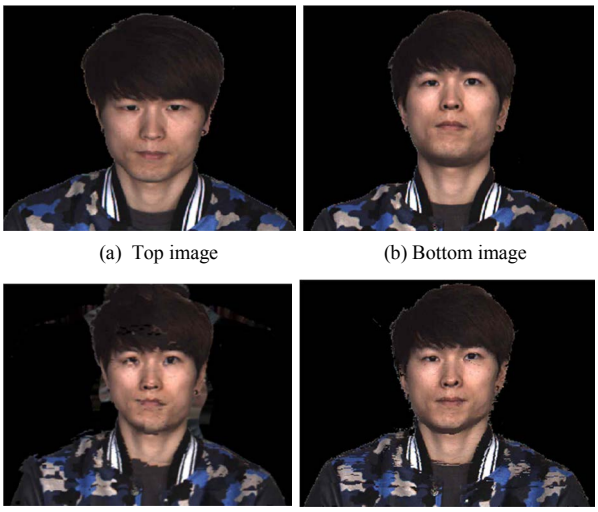
Figure 2 shows two depth maps. From the enlarged images of their results, we observe that the proposed method can represent the depth information more accurately. Furthermore, our method produces better image quality in textureless regions, such as in the face.



(a) Stereo matching based method (b) Proposed method

Fig. 2. Depth map comparison

Figure 3 shows eye gaze corrected images using view synthesis. These results demonstrate that the hybrid depth estimation method is more effective for eye gaze correction.



(a) Top image (b) Bottom image

(c) Stereo matching based method (d) Proposed method

Fig. 3. Results of eye gaze correction

IV. CONCLUSION

Eye contact is critical for immersive videoconferencing. In this paper, we propose an eye gaze correction method using two stereo camera sets and one depth camera. The camera configuration is employed to reduce the problems of each depth sensor. The proposed depth estimation algorithm improves depth precision, especially in less-detailed regions. Experimental results show that the proposed system is more effective in generating natural eye contact images.

ACKNOWLEDGMENT

This research was funded by the MSIP(Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2014.

REFERENCES

[1] C. Kuster, T. Popa, J.C. Bazin, C. Gotsman, M. Gross, "Gaze correction for home video conferencing," ACM Transaction on Graphics, vol. 31, no. 6, pp. 1-6, Nov. 2012.

[2] Video and Requirement Group, "Call for Proposals on 3D Video Coding Technology " N12036, ISO/IEC JTC1/SC29/WG11, 2011.

[3] E.K. Lee and Y.S. Ho, "Generation of multi-view video using a fusion camera system for 3D displays," IEEE Trans. on Consumer Electronics, vol. 56, no. 4, pp. 2797-2805, Nov. 2010.

[4] L. Xia, C.C. Chen, and J.K. Aggarwal, "Human detection using depth information by Kinect" Computer Vision and Pattern Recognition Workshops , pp. 15-22, June 2011.

[5] Z. Zhang, "A flexible new technique for camera calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, Nov. 2000.

[6] U. Fecker, M. Barkowsky, and A. Kaup, "Improving the Prediction Efficiency for Multi-View Video Coding Using Histogram Matching," Picture Coding Symposium, pp. 2-17, 2006.

[7] Y.S. Kang, and Y.S. Ho, "High-quality multi-view depth generation using multiple color and depth cameras," IEEE International Confernce on Multimedia and Expo, pp. 1405-1410, July 2010.

[8] M. Day, "Extracting Euler Angles from a Rotation Matrix," https://d3cw3dd2w32x2b.cloudfront.net/wp-content/uploads/2012/07/euler-angles1.pdf