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/m15120
January, 2008 Antalya, Turkey

Title: **Virtual View Synthesis for FTV**

Source: **GIST (Gwangju Institute of Science and Technology)**

Author: **Yo-Sung Ho, Sang-Tae Na, and Kwan-Jung Oh**

1. Introduction

Free viewpoint television (FTV)[1][2], one of the most challenging technologies, enables unrestricted spatio-temporal navigation within a scene captured using multiple views. The multiple views can be virtual views as well as real views from multiple cameras. In order to obtain the virtual views, we need corresponding depth data and camera parameters. In this document, we propose a method of virtual view synthesis for FTV.

2. Virtual View Synthesis for FTV

In general, if we have multi-view video and its corresponding depth data, we can generate virtual views. When we project pixel positions in the reference view to the target view, the positions are determined by depth values of the corresponding position in depth video and camera parameters of the reference view and the target view. Therefore, in order to generate high quality of virtual views, we need precise depth values and camera parameters.

2.1. 3D Warping Technique

In order to generate virtual views, we have applied a 3D warping technique. The perspective projection matrix for 3D warping can be represented by

$$PM = A[R \mid t] \tag{1}$$

where A, R, and t denote the intrinsic matrix, rotation matrix, and translation vector, respectively, and these values are called as camera parameters. We can project pixel positions from the image coordinate to the 3D world coordinate using the projection equation. Eq. (2) is the projection equation, which consists of the depth data and Eq. (1), and it can be transformed to Eq. (3).

$$P_{ref}(x,y,1)\cdot D = A[R\,|\,t]\cdot \tilde{P}_{WC}(x,y,z,1) \quad (2)$$

$$P_{WC}(x,y,z) = R^{-1}\cdot A^{-1}\cdot P_{ref}(x,y,1)\cdot D - R^{-1}\cdot t \quad (3)$$

where D denotes the depth data, P is the pixel position on the 3D world coordinate or the homogenous coordinate in the reference image coordinate system, and $\tilde{P}$ indicates the homogenous coordinate in the 3D world coordinate system. After the projection, the pixel positions in the 3D world coordinate are mapped into the positions in the desired target image by Eq. (4) that is the inverse form of Eq. (1).

$$P_{target}(x,y,1) = A\cdot R\cdot (P_{WC}(x,y,z) + R^{-1}\cdot t) \quad (4)$$

Then, we can get the right pixel positions in the target image with respect to the pixel positions in the reference image. After that, we copy the pixel values from the pixel positions on the reference image to the projected pixel positions on the target image.

2.2. Virtual Camera Parameters and Hole Filling in View Synthesis

In order to synthesize virtual views, we need camera parameters of references views and virtual views. However, camera parameters for virtual views are not given. Thus, we need to find proper camera parameters. Since virtual views are in-between views of the two reference views provided, we assume that the virtual camera parameters are divided values of the given camera parameters by the number of virtual views linearly.

During the view synthesis process, holes are generated due to the imperfect camera parameters and depth data as well as occlusion and disocclusion regions. The holes, generated by occlusion and disocculsion, can be filled using two reference views. Figure 2 shows that holes are generated in the opposite side of the foreground depending on the position of reference views, so we can remove the holes. Besides, relatively small holes are also generated due to the mismatching caused by the improper camera parameters and depth data. For those small holes, we have applied median filtering. Finally, for holes existing around edge regions, we have used neighboring pixel values that are the nearest nonzero values from the holes.

(a) Original view 0

(b) Original View 1

Fig. 1 Original Views

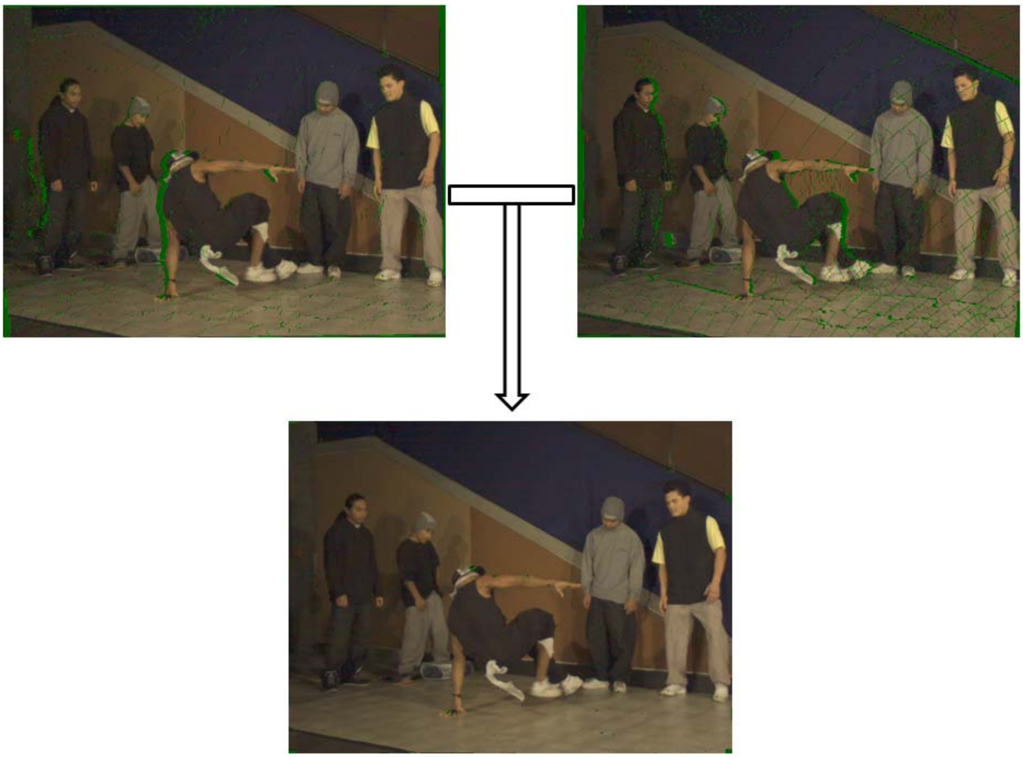

Fig. 2 Hole filling using two virtual views

Fig. 3 Synthesized Virtual Views between View 0 and View 1

3. Summary

In this document, we have proposed a method for virtual view synthesis. After we applied the 3D warping technique using neighboring camera parameters to find virtual camera parameters, we synthesized virtual views using reference views, corresponding depth data and camera parameters of the virtual views.

4. Acknowledgements

This work was supported in part by the Information Technology Research Center (ITRC) through the Realistic Broadcasting Research Center (RBRC) at Gwangju Institute of Science and Technology (GIST), and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

5. References

[1] A. Smolic and P. Kauff, "Interactive 3d video representation and coding technologies," IEEE Special Issue on Advances in Video and Delivery, Dec. 2004.

[2] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and Thomas Wiegand, "3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards", ICME 2006, IEEE International Conference on Multimedia and Expo, July 2006