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

ISO/IEC JTC1/SC29/WG11
MPEG2009/M16949
October 2009, Sian, China

Source: GIST (Gwangju Institute of Science and Technology)
Status: Report
Title: 3D Video Test Sequence and Camera Parameters
Author: Yun-Suk Kang, Eun-Kyung Lee, Jae-Il Jung, Jong-Ho Lee, In-Yong Shin and Yo-Sung Ho

1. Introduction

This document reports the test sequence provided by Gwangju Institute of Science and Technology (GIST). We describe the 3D video test sequence and provide its camera intrinsic and extrinsic parameters. The test sequence is composed of 5-view color videos and their corresponding depth videos. They are captured by a multi-depth camera system and then processed. The results of intermediate view synthesis are also presented.

2. Multi-Depth Camera System

To capture the test sequence, we combine five video cameras and five TOF(Time-of-Flight) depth cameras as shown in Fig. 1. The video camera model we use is Basler Pylon GigE that can provide up to HD resolution. The depth camera model is Mesa Imaging SR4000. Two different kinds of cameras are mounted on a camera frame. The fundamental setup is that the first row has five video cameras and the second row has the five depth cameras. These two rows also move to the upper and lower sides. Each interval between neighboring cameras is 65mm.

Fig. 1. Multi-Depth Camera System

For capturing a sequence simultaneously, we connect the video cameras to the acquisition PC through a synchronizer CA-1000 by National Instrument. To synchronize the depth cameras, we modified the software development kit (SDK) provided by Mesa Imaging for synchronized capturing.

3. Data Acquisition and Processing

In this section, we briefly present how to acquire the original data and what we have done to generate the 3D video sequence. Table 1 shows the specification of the captured data.

Table 1: Specification of the original color and depth video

Sequence Name	Cafe (420 frames)
Color Image Property	HD (1920x1080), 30fps
Depth Image Property	QCIF (176x144), 30fps
Camera Arrangement	Five video cameras on the 1st row and five TOF depth cameras on the 2nd row with 6.5cm of horizontal spacing and 8cm of vertical spacing

Five video cameras capture a scene with 30fps. The maximum resolution is HD and the lower resolutions are possible by setting the ROI from the original HD resolution images. Figure 2(a) is the captured five images by the video cameras. Five depth cameras also obtain 3D information of the same scene with 30fps. These depth cameras have two types of output images which are depth images and intensity images. The depth images represent the 3D information of the scene and the intensity images are considered as the grey images of the scene. Despite the depth camera has a small output resolution of QCIF, we can acquire 3D information of the scene in real time. Figure 2(b) shows the depth images of the scene.

However, one problem of the depth camera is that simultaneous capturing is allowed up to three cameras. It is the problem of TOF modulation frequencies. SR4000 has three different modulation frequencies so that we can simultaneously operate up to three cameras. Therefore, we used three depth cameras at positions 1, 3, and 5.

(a) Multi-view color video

(b) Multi-view depth video

Fig. 2. Captured multi-view and multi-depth videos

To generate the 3D video test sequence that is composed of multi-view video and its corresponding depth video, we performed the following procedures. For color images, we

corrected the color difference among views and reduced the geometric error by image rectification. For depth images, we compensated for the lens distortion. Camera parameters for the depth cameras were obtained after this distortion is corrected. We also performed depth value correction and rectified each depth camera. These preprocessed depth image of each view was warped to the position of the upper color camera and used as the initial depth value for stereo matching. Finally, we generated the high-resolution depth maps of each view by segment-based stereo matching.

4. 3D Video Test Sequence and Camera Parameters

Figure 3 shows the 3D video test sequence provided by GIST. It is composed of 5-view HD color and depth videos. The detail specification on the test sequence is described in Table 2. Table 3 shows the results of view synthesis of camera positions 2 and 4 by using the color and depth images at positions 1, 3, and 5. The intermediate view images were generated by VSRS 4.0 and we used the adjacent two view images.

Table 2: Test sequence specification (Cafe)

Sequence length	420 frames
Sequence Status	Rectified and color corrected
Color Image Property	HD (1920x1080), 30fps
Depth Image Property	HD (1920x1080), 30fps
Depth Specification	Z_near: 2500mm Z_far: 4000mm

Fig. 3. Test sequence: multi-view color and depth video

Table 2: View synthesis results

Viewpoints	Camera 2	Camera 4
Average PSNR (dB)	30.54	31.07

The camera parameters of the provided test sequence have the form shown in Eq. 1 and Table 4. In Eq. 1, m and M represent the coordinates of the image point and the world point, respectively. The matrices A and M, and the vector t indicate the intrinsic matrix, the rotation matrix, and the translation vector. Table 5 show the common intrinsic and rotation matrices. Translation vectors are indicated in Table 6.

$$\mathbf{m}=\mathbf{A}[\mathbf{R}\ \ \mathbf{t}]\mathbf{M} \tag{1}$$

Table 4: Form of the camera parameter

Intrinsic Matrix: A			Rotation Matrix: R			Translation Vector: t
A_{11}	A_{12}	A_{13}	R_{11}	R_{12}	R_{13}	t_{11}
A_{21}	A_{22}	A_{23}	R_{21}	R_{22}	R_{23}	t_{21}
A_{31}	A_{32}	A_{33}	R_{31}	R_{32}	R_{33}	t_{31}

Table 5: Camera intrinsic and rotation matrices

Intrinsic Matrix			Rotation Matrix		
1796.4762	0	978.4051	0.9999	0.0034	-0.0056
0	1796.8924	489.3078	0.0032	-0.9994	-0.0349
0	0	1	-0.0058	0.0349	-0.9994

Table 6: Translation vectors

	Camera 1	Camera 2	Camera 3	Camera 4	Camera 5
t_{11}	-259.6527	-324.6527	-389.6527	-454.6527	-519.6527
t_{21}	474.0773	474.0773	474.0773	474.0773	474.0773
t_{31}	2199.5855	2199.5855	2199.5855	2199.5855	2199.5855

5. Conclusion

We have presented the 3D video test sequence generated by GIST. The test sequence is composed of 3-view color videos and their corresponding depth videos. The image resolution is HD. You can download the test sequence from our web site.

6. Acknowledgement

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-2009-(C1090-0902-0017)).

7. References

[1] ISO/IEC JTC1/SC29/WG11 N9595, “Call for Contribution on 3D Video Test Material,” January 2009.