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/M16396
April 2009, Maui, USA

Source: GIST (Gwangju Institute of Science and Technology)
Status: Proposal
Title: 3-D Test Sequence – Multi-view Video and Depth Map
Author: Eun-Kyung Lee and Yo-Sung Ho

1. Introduction

This document describes the test sequence provided by Gwangju Institute of Science and Technology (GIST). We present a new camera system combining a time-of-flight depth camera and multiple video cameras to generate a multiview video-plus-depth. In the following, we describe the specification of the test sequence and its camera parameters.

2. Specification of the Test Sequence

The sequence was recorded by using our hybrid camera system, as shown in Fig. 1. The system consists of five multi-view cameras and one depth camera. There is one sync-generator, sending a synchronization signal; this signal is distributed to all PCs. We captured one sequence with the 1-D parallel camera arrangement, where the camera interval is 20cm.

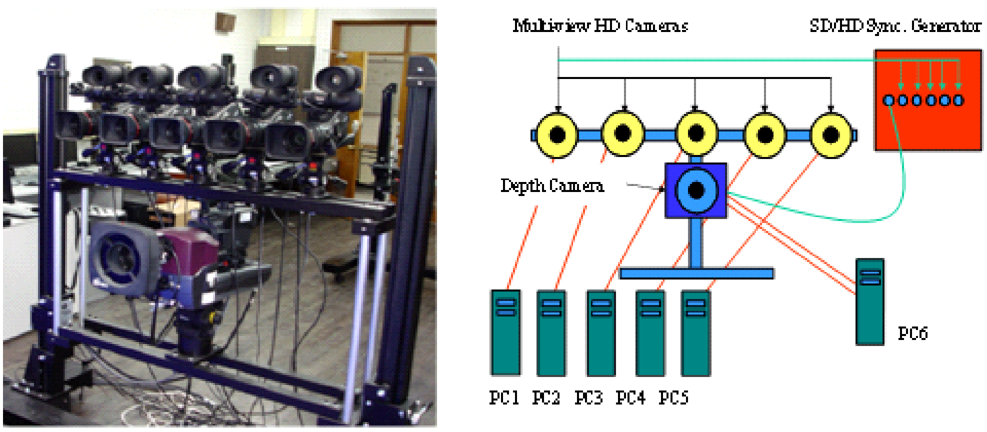

Figure 1: Hybrid camera system

Table 1: Test sequence by GIST

Sequence	Image Property	Camera Arrangement	Depth Specification
Newspaper	1920x1080, 30fps (rectified)	5 cameras with 20cm spacing; 1D parallel	Z_near:2220mm Z_far:4200mm

Table 1 shows the specification of the test data. The original picture size is 1920(H) x 1080(V), and the frame rate is 30 frames/sec. The length of the original sequence is 5 seconds. For the sequence, multi-view rectification and color correction were applied. Then, the images were converted to YUV 4:2:0 format. The Z_near is 2220mm and Z_far is 4200mm for correct interpretation of depth values.

3. Preprocessing

We perform the multiview rectification. The multi-camera arrays have geometric errors, because there are manually built. In order to minimize the geometric errors, we calculate the common baseline, and then apply the rectifying transformation to the multiview image [1]. After the multiview rectification, we correct the color mismatch problem of multiview images using a color calibration method. In general, the color properties of captured images can be inconsistent due to the different camera properties of the multiview camera system [2].. Figure 3 shows the overlapped images with identical vertical coordinates and the equal interval.

Figure 3: Rectified test sequence "Develiry".

4. Camera Parameters

The definition, format, and validation of camera parameters of the sequence is described by referring the document, ISO/IEC JTC1/SC29/WG11 N9595, "Call for Contributions on 3D Video Test Material" [3].

4.1. Intrinsic Parameters

The intrinsic parameters for each camera are represented as follows:

Format of Intrinsic Parameter

focal_length_x	skew_parameter	principal_point_x
0	focal_length_y	principal_point_y
0	0	1

Table 2 shows the rectified intrinsic parameters of the sequence.

Table 2: Rectified intrinsic matrix

2059.383013	0.0	928.543706
0.0	2057.422796	503.546216
0.0	0.0	1.0

4.2. Extrinsic Parameters

Table 3 and Table 4 show the rectified extrinsic parameters of the sequence. In these tables, “i” means *i*th camera, R_{mn} means (m, n) coordinate of rotation matrix for each camera for the sequence, and t_k means *k*th coordinate of translation vector for each camera for the sequence. Rotation matrix and translation vector is described by the following formats.

The external transformation is performed by rotation **R** and translation **t** from a 3D point M_c in the camera coordinate system to the related 3D point M_w in the world coordinate system by

$$M_w = \mathbf{R} M_c + \mathbf{t} \tag{1}$$

The external transformation from a 3D point M_w in the world coordinate system to the related 3D point M_c in the camera coordinate system is performed by

$$M_c = \mathbf{R}^{-1} M_w - \mathbf{R}^{-1}\mathbf{t} \tag{2}$$

The extrinsic parameters for each camera are represented as follows:

Format of Rotation Matrix

R_{11}	R_{12}	R_{13}
R_{21}	R_{22}	R_{23}
R_{31}	R_{32}	R_{33}

Format of Translation Vector

t_1
t_2
t_3

To verify that camera parameters and world to camera equation are consistent a 3D-point $\mathbf{M}_w$ is projected on all camera planes. The following 3D-Point $\mathbf{M}_w$ is given (see below on how to get this) in world coordinates. The 3-dimensional world point is now mapped into a 2-dimensional camera point by:

$$s * \mathbf{m} = \mathbf{A} * (\mathbf{R}^{-1} \mathbf{M}_w - \mathbf{R}^{-1}\mathbf{t}) \tag{3}$$

s is an arbitrary scaling factor to make the third coordinate of **m** equal to one. Note that **R** is the rotation of the camera relative to the world coordinates and **t** is given in world coordinates.

Table 3: Rectified rotation matrix

0.9997	0.0014	0.0251
-0.0004	0.9992	-0.0396
-0.0252	0.0396	0.9989

Table 4: Rectified translation vector for each camera

	translation		
Camera	t_x	t_y	t_z
0	50.26184	32.58658	23.03765
1	30.42681	32.58658	23.03765
2	10.59178	32.58658	23.03765
3	-9.24325	32.58658	23.03765
4	-29.07829	32.58658	23.03765

5. Experimental Results

In order to check the validity of this sequence, we have tested the quality of synthesized image using reference software. In the following subchapters, we demonstrate the experimental results.

5.1. Depth Estimation

The proposed sequence consists of 5 view videos. Among them, we estimated depth video of two views, view 2 and view 4. Figure 2 shows a snapshot of an estimated depth image on view 3.

Figure 2: Original test sequence "Delivery"

5.2. View Synthesis

Using the estimated depth video, we synthesized two intermediate views. Figure 3 is a demonstration of synthesized image. The left image is the original image for viewpoint 2 and the right image is its synthesized image using 'VSRS' software. As you can see, the quality of the synthesized image is good. The average PSNR value of synthesized images is higher than 30 dB.

Figure 3: Original image of view 2(left) and synthesized image (right)

6. Conclusion

We have explained the test sequence provided by GIST. Using our hybrid camera system, we have obtained the sequence and the corresponding depth map named "Delivery". You can download them at our web site:
The ftp address, User ID and Password will be announced at the meeting.

7. 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 IITA(Institute for Information Technology Advancement) (IITA-2009-C1090-0902-0017).

8. References

[1] Jae-Il Jung and Yo-Sung Ho, "Color Correction Method Using Gray Gradient Bar For Multi-view Camera System," International Workshop on Advanced Image Technology (IWAIT), pp. 1-6, 2009.

[2] Yun-Suk Kang and Yo-Sung Ho, "Geometrical Compensation Algorithm of Multiview Images for Arc Multi-Camera Array," Lecture Notes in Computer Science (LNCS), 5353, pp. 543-552, 2008.

[3] ISO/IEC JTC1/SC29/WG11 N9595, "Call for Contributions on 3D Video Test Material," January 2008.