MULTI-VIEW 3D VIDEO ACQUISITION USING HYBRID CAMERAS WITH BEAM SPLITTER

Gi-Mun Um[1], Taeone Kim[1], Gun Bang[1], Namho Hur[1], Eun-Kyung Lee[2], Jae-Il Jung[2], Yun-Suk Kang[2], Gyo-Yoon Lee[2], and Yo-Sung Ho[2]

[1]Electronics and Telecommunications Research Institute
218, Gajeongno, Yuseong-gu, Daejeon, 305-700, Republic of Korea
E-mail:{gmum, kimm, gbang, namho}@etri.re.kr
[2]Gwangju Institute of Science and Technology (GIST)
261Cheomdan-gwagiro, Buk-gu, Gwangju, 500-712, Republic of Korea
Email: {eklee78, jijung, yunsuk, gyoyoon, hoyo}@gist.ac.kr

ABSTRACT

We present a multi-view 3D video acquisition and its processing system for multi-view 3D television (3DTV). The proposed hybrid camera system consists of three-color cameras and one time of flight (TOF) camera. Since currently available TOF cameras do not provide color images associated with the depth image, we use a beam splitter between the color camera of the center view and the TOF camera to minimize the FOV difference between two cameras. We capture three-view color and one-view depth videos with hardware trigger synchronization. After capturing videos, we perform camera calibration, color correction, lens distortion correction, rectification, depth calibration, and multi-view depth generation for three-view color videos. We showed experimental results using the proposed acquisition system. The proposed system can be used to generate multi-view 3D videos for 3DTV.

Index Terms**—**Multi-view 3D Video, TOF camera, 3DTV

1. INTRODUCTION

With the success of 3D movie Avatar, various applications such as movies, home entertainment, games, teleconference, and TV services are starting to support 3D video. Many countries are attempting experimental 3DTV services [1] or commercial 3DTV services through satellite channels. Most of them are based on the stereoscopic video and viewers should wear glasses for 3D depth perception. In order to provide 3D video without wearing glasses and more flexible depth perception, 3D information such as depth image is required.

For many years, various kinds of depth acquisition techniques have been studied to obtain good 3D depth information of the scene. We can classify them into two different classes. The first one is a passive approach, such as image-based disparity estimation [2]. This approach requires relatively inexpensive color cameras, but has problems at depth discontinuity boundaries due to occlusion. As the second approach, there is an active approach, such as time-of-flight (TOF) camera. There are diverse commercially available TOF cameras from Swiss Ranger [3], PMD Tec. [4], and Prime sense [5]. These TOF cameras provide real-time depth data at video frame rates. However, their resolution of the depth image is much lower than that of the color image, except for Prime sense camera. In addition, since they use infrared (IR) lights for depth acquisition, they have problems in low reflectance areas, such as black regions. Moreover, they provide inaccurate depth values at object boundaries when objects are moving fast in the scene.

In order to alleviate these problems, recent researches are trying to fuse both types of depth acquisition techniques and various types of hybrid camera systems have been proposed [6, 7, 8, 9, 10]. In [6, 7], they used a pair of stereo color cameras and one TOF camera. The TOF camera is located at the center of two color cameras. The camera system in [8] consists of a mirrored stereo HD rig, two satellite HD cameras, and one low-resolution TOF camera on the top. In [9, 10], five color cameras and two TOF cameras were used. Moreover, five HD color cameras and one TOF camera were utilized in [11]. Most of them use 3D warping of the low-resolution TOF depth to high-resolution multi-view color images to obtain high-resolution depth maps. Since the required baseline length between color cameras for auto-stereoscopic 3D display is usually shorter than the real camera baseline length, we should synthesize virtual view images between real cameras. In addition, especially in the more than two camera configuration, the depth quality for the main view (usually center view) is the most important for obtaining high quality multi-view 3D video. Therefore, it is good to design the hybrid camera system to maximize the overlapped area between the main view color camera and TOF camera.

We present a multi-view 3D video acquisition and its processing system that includes a hybrid camera system. It consists of three RGB color cameras and one TOF camera. The proposed hybrid camera system uses an IR coated beam splitter at the center view in order to maximize the overlapped area between the color and TOF cameras.

2. MULTI-VIEW 3D VIDEO ACQUISITION SYSTEM

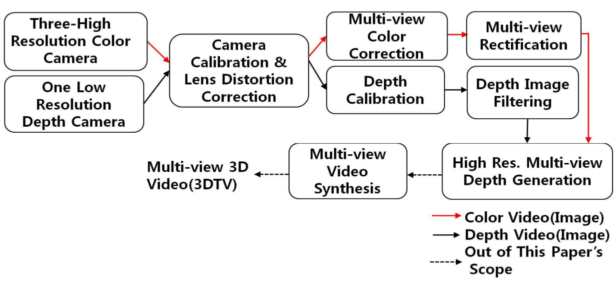


Figure 1. The overall framework of multi-view 3D video acquisition

The overall framework of the proposed system consists of multi-view color and depth capturing, captured video processing such as camera calibration, lens distortion correction, color correction, multi-view image rectification, depth calibration, depth image

978-1-61284-162-5/11/$26.00 © 2011 IEEE.

filtering and high-resolution multi-view depth generation as shown in Figure 1.

2.1. Camera System Setup

We construct the multi-view 3D video capture system consisting of three RGB color video cameras and one TOF camera as shown in Figure 2. We install the TOF camera at the center along with left and right RGB color cameras. The center RGB color camera is installed to the perpendicular direction of the TOF camera as shown in Figure 2. In order to maximize the overlapped region between these two cameras, we also install the IR-coated beam splitter. We use the rectangular shaped camera housing to maintain the locations of the TOF camera, the center RGB color camera, two IR emitters, and the beam splitter as shown in Figure 2. This housing also minimizes the scattering of lights that are reflected or penetrated on the beam splitter. Although we use the beam splitter, there are slight differences in the field of view (FOV) range between two cameras. Therefore, we try to match the center of imaged areas by each camera instead. As multi-view RGB color cameras, we adopt Hitachi's 3CCD HV-F31CL that has a resolution of XGA (1,024 x 768) and supports 15 and 30fps with a hardware trigger mode. We use PMD Tec.'s Camcube 3.0 that has a maximum resolution of 200 x 200 and supports maximum 40fps at the maximum resolution. This model can capture the distances from 0.3 to 7 meters. The wavelength of the infrared illumination unit is 870 nm [12]. Two IR illumination units may be placed horizontally or diagonally. We are also able to control the integration time for balancing noise levels of depth image and motion blur with micro (μ) second unit. For the alignment between color and TOF depth cameras, we adjust the horizontal position and the angle between the beam splitter and the center color RGB camera.

2.2. Multi-view Video Capture System Setup

For capturing three-view color videos and one-view depth video with synchronization, we use two capture servers with RAID HDD system, a synchronization signal generator, and video capturing software. The GUI (Graphical User Interfaces) of the capturing software is shown in Figure 3(a) and Figure 3(b). With this software, we are able to capture three-view color videos and one-view depth videos with hardware synchronization. The first capture server captures three-view color videos and the other capture server captures one-view depth videos. Before capturing, we can also check the alignments among three-color cameras or between TOF camera and color cameras in the server. After doing camera alignments, the TOF camera is connected to the other server for capturing TOF camera images with software as shown in Figure 3(b). There are three kinds of images can be captured from the TOF camera. The first one is a depth image that is quantized gray image with 256 levels. In the depth image, brighter pixel means that corresponding pixel has nearer distance to the camera. The second one is an intensity image that represents a gray level image of the scene. The third one is an amplitude image that reflects the signal strength of the active illumination. This image is used to determine the quality of the depth value [13].

2.3. Captured Video Processing

After capturing three-view color and one-view depth videos, we do video processing as shown in Figure 1 in order to obtain better quality of depth for multi-view 3D video. We first calibrate each camera in order to estimating extrinsic and intrinsic parameters [14]. For camera calibration, we adopt the camera calibration toolbox for MATLAB [15]. We then correct the lens distorted images from the extracted distorted points in the target image using lens distortion correction algorithm in [16].

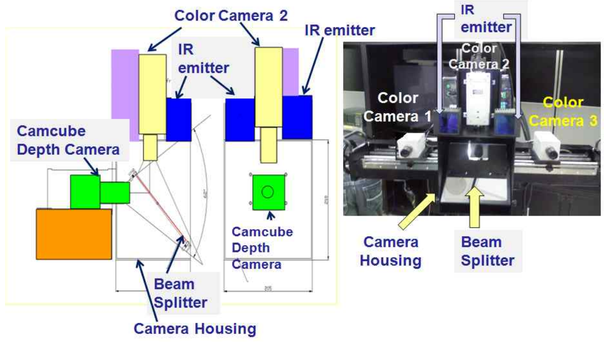
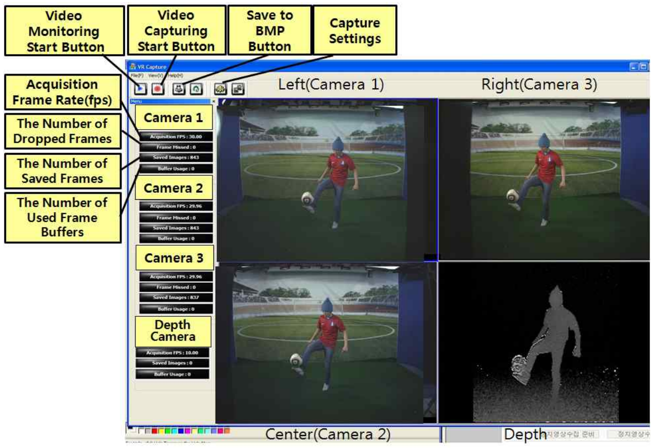
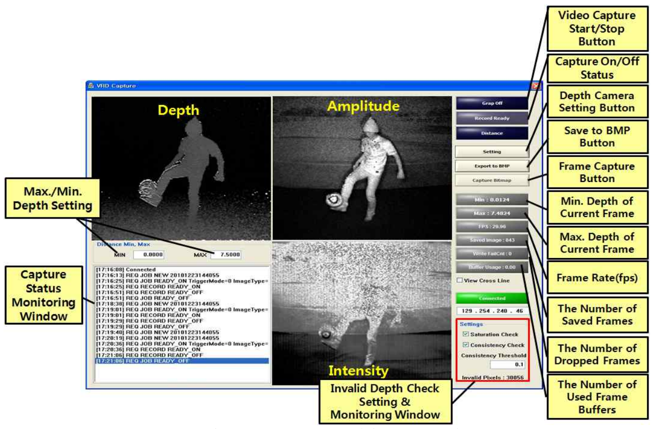


Figure 2. The structure of the proposed hybrid camera system



(a) Three-view color video capturing and depth video monitoring



(b) TOF camera video capturing

Figure 3. The GUI of color and TOF video capturing software

2.3.1. Color Video Processing

As a next step, color correction for three-view color images is performed using the method in [17] that uses Macbeth 24-color checker as shown in Figure 4. We analyze the differences between corresponding patches, and calculate the correction curve using non-linear regression. This process is applied to the left and right images, and the center image is regarded as a reference view. With the calculated curve, we generate lookup tables for each view and each color channel. The color values of the main scene are corrected by mapping with the tables.

After correcting the color differences among views, we then rectify three-view color images using the technique in [18]. Figure 5 shows an example of the rectified images. We set the center camera as the reference and then estimate the left and right camera parameters to have the identical characteristic to the reference camera except x-axis' positions. The transform is calculated by using the original and estimated camera parameters and then it is applied to each image. In order to remove unmapped region at the image boundary, we increase the focal length and crop the image to the proper resolution.

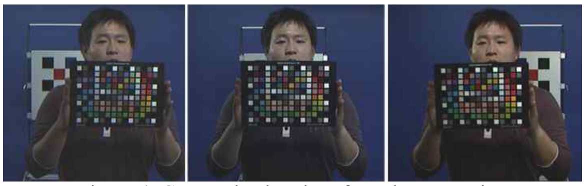


Figure 4. Captured color chart for color correction

Figure 5. Example of rectified multi-view images

2.3.2 Depth Calibration

The depth values in the depth video from the TOF depth camera are very sensitive to noises. Their sources are diverse including physical limitation of hardware and specific object properties, etc. Therefore, depth data are noticeably contaminated with random and systematic measurement errors dependent on reflectance, angle of incidence, and environmental factors like temperature and lighting. To reduce those errors, we employ a depth calibration method as in [6].

For the depth calibration in indoor environments, we compute the depth of the planar checker pattern within the limited space by increasing the distance from the checker pattern to the depth camera using our system as shown in Figure 6. Thereafter, we make a 4D look-up table (LUT) that maps the depth value from the TOF camera to the 3D positions of the multi-view RGB color video cameras and. We then calibrate the 3D position of each pixel in the depth video from the x and y positions of the feature points, and the real depth values obtained from the disparities between stereo pairs of multi-view images. More details can be found in [6].

2.3.3. Depth Image Filtering

After doing depth calibration, we perform the depth map filtering for the removal of salt noise. Since the depth, value with salt noise is unusually very large compared to the depth values of correct neighboring depth pixels. Therefore, these unexpected depth pixels should be removed and be substituted by new depth values. In order to do this, we utilize an averaging filter together with median filter. We firstly apply the averaging filter to the original depth image.

We also apply a median filter to the original depth image independently. With the filtered two depth images, we finally obtain the improved depth image by summarizing two depth images using the following rule. If the differences between the depth value of the current pixel and those of both left and right neighboring pixels are larger than a certain threshold, we select the depth value from the median filtered depth image. Otherwise, we select depth value from the average filtered depth image. Figure 8 shows a result of the depth image filtering. We can see that very bright erroneous pixels in the depth image are removed in the filtered depth image.

2.3.4. 3D Warping of Depth

After removing noisy pixels in the depth images, we then generate the initial depth images for high-resolution multi-view images by performing 3D warping of the depth values to the obtained from the TOF depth camera.

Firstly, we project pixels of the depth images into the 3D world coordinate using the obtained depth values. Then we then re-project the 3D points into each view image. Figure 9 shows an example of 3D warping result using the acquired depth image.

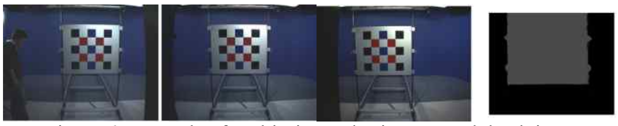


Figure 6. Example of multi-view color images and depth image

(a) Original (b) After filtering

Figure 7. TOF depth image filtering

(a) Color image (b) 3D warped depth

Figure 8. Example of 3D warping result

2.3.5. Stereo Matching with Warped Initial Depth

After getting warped depth images for each high-resolution color video, we generate the final depth video for each view by stereo matching. In the process of stereo matching, we utilize the warped depth values as initial depth values for the stereo matching. By doing this, we can reduce the search range for finding the matched regions or points. In addition, depending on the reduced search range, we can overcome the mismatching problem in the textureless regions. To reduce mismatched depth values along object boundaries, we detect moving objects using color difference between frames. More details of stereo matching technique that we use can be found in [8].

3. EXPERIMENTAL RESULTS

In this section, we present the experimental results using the test contents captured by the proposed acquisition system. Figure 3 shows an example of captured three-view color images and one depth image for the center view. We can see that the FOV of the color image of the center view is quite similar to that of the TOF depth camera although the FOV of TOF depth camera is a little smaller than that of color camera. Figure 5, Figure 7, and Figure 8 show the multi-view rectification result, depth filtering result, and 3D warping result for the test sequence in Figure 3, respectively. Figure 9 shows that finally generated high-resolution depth images by two different depth estimation techniques for the rectified color image of the center view. The depth image shown in Figure 9(a) is obtained by DERS used in the MPEG 3DV [19], and the depth image in Figure 9(b) is obtained by the proposed depth estimation technique. We can see that the results by the proposed technique maintain the object boundaries well whereas the DERS has many false matches at the boundaries of the object due to similar color at the background. Figure 10 shows the high resolution depth images generated by the proposed technique for the three different scenes . Table 1 presents that the comparison of the average peak signal-to-noise ratio (PSNR) of the synthesized images generated by the DERS technique and the proposed technique, respectively. We can see that the synthesized views produced by proposed technique show higher PSNR values than those obtained by the DERS technique.

4. CONCLUSION

We have presented a multi-view 3D video acquisition and processing system for multi-view 3DTV. The proposed hybrid camera system consists of three-color cameras and one TOF camera. We are able to minimize the difference of FOV between color and depth cameras using an IR coated beam splitter. With the proposed technique, we obtained better quality depth images than the DERS technique; in particular, around object boundaries and homogenous regions.

As a future work, we are going to reduce motion blur in the depth image around object boundaries and increase time consistency of the depth image in the background of the scene.

5. ACKNOWLEDGEMENTS

This work was supported by the IT R&D Program (Development of Multiview 3D Compatible UHDTV Broadcasting Technology) of KCC, Korea.

6. REFERENCES

[1] K. Yun, K. Kim, N. Hur, S.I. Lee and G.H. Park, “Efficient Multiplexing Scheme of Stereoscopic Video Sequences for Digital Broadcasting Services,” *ETRI Journal*, vol. 32, no. 6, pp.961-964, December 2010.

[2] A. Klaus, M. Sormann and K. Karner, “Segment-Based Stereo Matching using Belief Propagation and a Self-Adapting Dissimilarity Measure,” *Proc. 18th Int’l Conf. on Pattern Recognition*, vol. 3, pp.15-18, 2006.

[3] http://www.mesa-imaging.ch/prodview4k.php

[4] http://www.pmdtec.com/products-services/pmdvisionr-cameras/pmdvisionr-camcube-20/, 2009.

[5] http://www.primesense.com, 2010.

[6] J. Zhu, L. Wang, R. Yang, and J. Davis, “Fusion of Time-of-Flight Depth and Stereo for High Accuracy Depth Maps,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2008*, pp. 1-8, June 2008.

[7] S.Y. Kim, E.K. Lee, Y.S. Ho, “Generation of ROI Enhanced Depth Maps Using Stereoscopic Cameras and a Depth Camera,” *IEEE Trans. on Broadcasting*, 54 , paper 4, pp. 732-740, 2008.

[8] B. Bartczak and R. Koch, “Dense Depth Maps from Low Resolution Time-of-Flight Depth and High Resolution Color Views,” *Lecture Notes in Computer Science,* vol. 5876, pp.228-239, 2009.

[9] E.K. Lee, and Y.S. Ho, “Generation of High-Quality Depth Maps Using Hybrid Camera System for 3-D Video,” *Journal of Visual Communication and Image Representation*, vol. 22, no. 1, pp. 73-84, no. 1, January 2011.

[10] Y.S. Kang, and Y.S. Ho, “High-quality Multi-view Depth Generation Using Multiple Color and Depth Cameras,” International Workshop on Hot Topics in 3D, pp. 1405-1408, 2010.

[11] A. Frick, B. Bartczack and R. Koch, “3D-TV LDV Content Generation with a Hybrid TOF-Multi Camera Rig,” *Proceedings of 3DTV CON*, vol. 1, June 2010.

[12] *PMD[vision] Camcube 3.0 datasheet*, 2010.

[13] *PMD SDK 2 version 2.2.0 Programming Manual*, PMD Tec., 2010.

[14] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330-1334, November 2000.

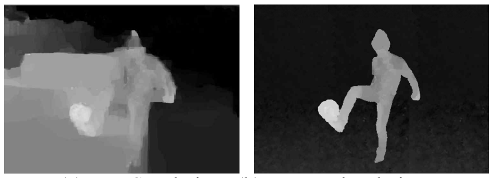
[15] http://www.vision.caltech.edu/bouguetj, Camera Calibration Toolbox for MATLAB.

[16] A. Wang, T. Qiu, and L. Shao, “A Simple Method of Radial Distortion Correction with Centre of Distortion Estimation,” *Journal of Mathematical Imaging and Vision*, vol. 35, no. 3, pp. 165-172, July 2009.

[17] J.I. Jung and Y.S. Ho, “Color Correction Method Using Gray Gradient Bar For Multi-view Camera System,” *Int’l Workshop on Advanced Image Technology (IWAIT*), pp. MP.C4(1-6), January 2009.

[18] Y.S. Kang and Y.S. Ho, "Geometrical Compensation for Multi-view Video in Multiple Camera Array," *Proc. of International Symposium on Electronics and Marine (ELMAR)*, pp. 83-86, Sept. 2008.

[19] ISO/IEC JTC1/SC29/WG11, “Reference Software of Depth Estimation and View Synthesis for FTV/3DV,” M15836, October 2008.

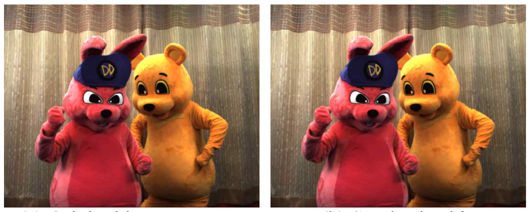


(a) DERS technique (b) Proposed technique

Figure 9.The comparison of the generated depth images

(a) Soccer I (b) Pantomime (c) Soccer II

Figure 10.Generated depth images by the proposed technique



(a) Original image (b) Synthesized image

Figure 11. The comparison of the original image and synthesized image from generated depth images

Table 1. Objective quality comparison of view synthesis results

Sequence \ Depth Gen. Technique	Average PSNR (dB)	
	DERS	Proposed
Soccer I	28.35	29.84
Pantomime	27.59	29.13
Soccer II	29.12	31.08