RECENT ACTIVITIES FOR 3DTV RESEARCH

Yo-Sung Ho

Realistic Broadcasting Research Center (RBRC)
Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong Buk-gu, Gwangju, 500-712, Korea
Tel: +82-62-715-2211, Fax: +82-62-715-2204
E-mail: hoyo@gist.ac.kr

ABSTRACT

In recent years, various multimedia services have become available and the demand for realistic three-dimensional television (3DTV) is growing rapidly. Since 3DTV is considered as the next generation broadcasting service that can deliver natural, realistic and immersive experiences to users by supporting user-friendly interactions, a number of advanced three-dimensional video processing technologies have been studied. Among them, multi-view video coding (MVC) is the key technology for various applications including free-viewpoint video, free-viewpoint television, 3DTV, immersive teleconference, and surveillance systems. In this paper, we cover recent activities of 3DTV research. After defining the basic requirements for realistic 3D broadcasting services, we explain various multi-modal immersive media processing technologies.

Index Terms— 3DTV, video processing, multi-view video coding, free-viewpoint TV, depth estimation,

1. INTRODUCTION

As realistic and immersive multimedia services are expected to be available soon, the importance of three-dimensional (3D) video is increasingly recognized as the essential part of high-quality visual services [1]. In fact, 3DTV is widely accepted as the next generation television service, because it can provide more realistic and natural viewing experiences to users. In recent years, the ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) has been working on multi-view video plus depth coding related to free-viewpoint TV (FTV) services [2].

For 3DTV and FTV, it is very important to estimate the depth information of the natural scenes accurately. Although various algorithms for depth estimation have been developed in the field of image processing and computer vision, accurate measurement of the depth map from the natural scenes still remains problematic.

In general, we can classify 3D depth estimation methods into two categories: active depth sensing and passive depth sensing methods. Active depth sensing methods usually employ physical sensors to measure the distance. Example of active depth sensing is depth camera. Few years ago, the European ATTEST project showed a possibility of realizing a 3DTV system using a depth camera [3]. Passive depth sensing methods calculate depth information indirectly from 2D images captured by two or more video cameras. Examples of passive depth sensing include shape from stereo [4] and shape from focus [5]. The advantage of indirect depth estimation is that we can create depth maps inexpensively. However, accuracy of the depth maps obtained by passive depth sensing methods is relatively lower than that captured by active depth sensing methods directly.

Recently, Realistic Broadcasting Research Center (RBRC) in Korea has demonstrated full 3D contents for future home-shopping channel broadcasting using a depth camera [6]. However, although the depth cameras can produce useful depth information in real time, there are some in-built problems in the currently available depth camera systems.

In order to increase the quality of depth information and the measuring distance of a depth camera, Researchers proposed a depth map fusion method that combines trifocal cameras and a depth camera [7][8]. This fusion camera system has generated enhanced depth maps estimated by applying a stereo matching algorithm to three-view images with depth information captured by the depth camera. However, this system cannot produce high-resolution depth maps, because it completely depends on the low-resolution depth camera.

With rapid development of camera sensor technologies, future 3D multimedia applications will require depth maps of higher resolution than what current depth cameras can provide. One possible solution for generating high-resolution depth maps is to upgrade the current depth camera or develop a new depth camera. However, due to many challenges in real-time distance measuring systems, upgrading and improvements of depth cameras are very costly and time-consuming.

In this paper, we introduce multi-view camera systems to generate high-resolution depth maps by combining high-resolution multiple cameras and current low-resolution depth cameras. The proposed camera systems produce high-resolution video-plus-depth as a 3D video.

2. PREPROCESSINGS FOR 3D VIDEO GENERATION

2.1. Camera Calibration

Camera calibration is the process to estimate the camera parameters [9]. We extract the corner points from the check-patterned board images shown in Fig. 1. Then we obtain the camera internal and external parameters as the results of camera calibration. The camera internal parameters describe the physical characteristics. The camera external parameters are composed of rotation and translation of the camera. Using these parameters, we can calculate the camera projection matrix.

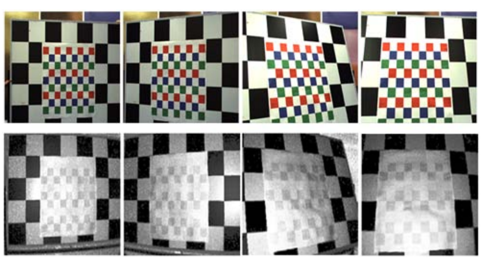

Fig. 1. Images for camera calibration

2.2. Multi-view Image Rectification

Multi-view image captured by multiple cameras has geometric errors because the cameras are manually arranged. These geometric errors are represented as the vertical pixel mismatches and non-uniform disparities between adjacent views shown in Fig. 2(a). Even though we use the same camera model, there are differences between camera internal characteristics. Therefore we have to minimize these geometric errors.

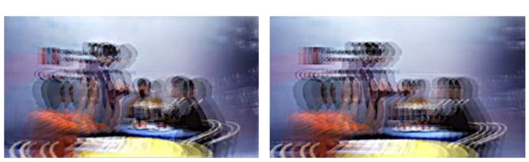

Fig. 2. Multi-view image rectification

Multi-view image rectification is the process to minimization the geometric errors by applying the transformation to each viewpoint. The transformation is calculated using the camera parameters and the error-corrected camera parameters. As a result, the rectified multi-view image has the same vertical coordinates and the same disparity values among viewpoints as shown in Fig. 2(b).

2.3. Multi-view Color Correction

When we use multiple cameras, there exists color mismatch problem among viewpoints. Although we use the same cameras, each camera represents colors differently. Also, cameras have different color characteristics of the scene due to the different illumination conditions. Since this color mismatch problem can deteriorate the quality of the depth map and intermediate view images, multi-view color correction which minimizes this problem is essential.

In general, color correction is performed based on the matching of the histogram or corresponding points of the multiple views. However, there is a color correction method using the standardized color chart. This method provides high accuracy because it is independent on the input images. Therefore, we have to capture the image of the color chart as shown in Fig. 3.

Fig. 3. Color chart images for multi-view color correction

2.4. Lens Distortion Correction

Especially for depth images, there exists a lens radial distortion shown in Fig. 4(a). Lens distortion is that the linear object of the scene is distorted to radial objects in the image by capturing. Because this lens distortion causes the shape mismatch between the color and depth images, it is necessary to correct the lens distortion of the depth image.

In order to correct the lens distortion, the points of distorted linear object is extracted, and the distortion center and distortion parameter are estimated. Then, the corrected image is reconstructed by using the distortion information as shown in Fig. 4(b)

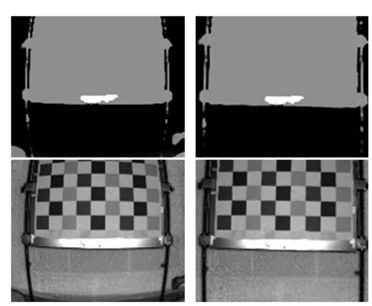

Fig. 4. Lens distortion correction for depth images

3. DEPTH MAP GENERATION USING MULTIPLE DEPTH CAMERA SYSTEM

3.1. Multi-view Camera System

We have constructed a nine-view camera system and captured the test sequence with 1D parallel camera arrangement as shown in Fig. 5 [8]. The system consists of one master PC and eight slave PCs. 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 5cm. The viewing zone is 40 cm in length. The cameras are Point Grey Research Flea with Sony CCD IEEE-1394 camera. The original picture size is 1024(H)×768(V), and the frame rate is 30 frames/sec. Figure 6 shows the acquired multi-view images using the multi-view camera system.

Fig. 5. Multi-view camera system

Fig. 6. Multi-view images by multi-view camera system

3.2. Hybrid Camera System

The proposed hybrid camera system is composed of one depth camera and five high-definition (HD) video cameras as shown Fig. 7 [10]. Those multiple video cameras are arranged in a one-dimensional array to construct a multi-view camera system. In addition, a clock generator sends a synchronization signals constantly to each camera and its corresponding personal computer equipped with a video capture board. The proposed camera system captures multi-view images by the multiple video cameras and a depth map from the depth camera at each sampling time.

Fig. 7. Hybrid camera system

To obtain depth maps for multi-view images, we perform a 3D warping operation onto each multi-view camera using the depth map measured by the depth camera. The warped depth data is used as an initial depth at each camera position. After we segment each multi-view image, we assign the depth value of the warped depth data in each segment as the initial depth of the segment.

We separate moving objects in each image and detect occlusion and disocclusion regions to improve the depth accuracy of object boundaries. Then, the depth of each segment is refined by a color segmentation-based stereo matching method for the foreground and background regions. Finally, we obtain multi-view depth maps by performing a pixel-based depth map refinement using a proposed cost function in each segment.

In order to evaluate the performance of the proposed method, we have constructed a hybrid camera system with five HD cameras and one depth camera as shown in Fig. 7. The measuring distance for depth information of the depth camera was from 0.50m to 5.00m. The baseline distances among multi-view HD cameras were 20cm.

Figure 8 shows the test multi-view image and depth map sequences captured by the hybrid camera system. The resolution of the multi-view images was 1920×1080, and the resolution of the depth maps was 720×486. Figure 9 shows the final multi-view color images and their depth maps

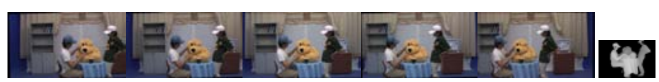

Fig. 8. Multi-view images by multi-view camera system

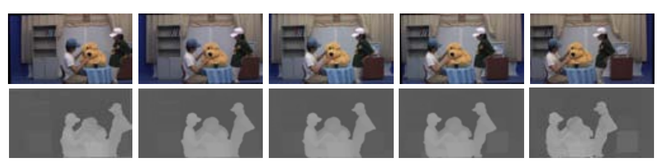

Fig. 9. Multi-view images by multi-view camera system

3.3. Multi-Depth Camera System

Figure 10 shows a multi-depth camera system that is composed of five color cameras and five depth cameras. The model of the color cameras is Basler Pylon GigE and the depth camera is Swiss Ranger SR4000 [11]. Two types of cameras are mounted on the frame, and they can shift to left and right, or up and downward. The horizontal and vertical distances between adjacent cameras are 6.5cm and 6cm, respectively. Multi-view color cameras are connected to the control PC through the synchronizer, and each depth camera is connected directly to the control PC.

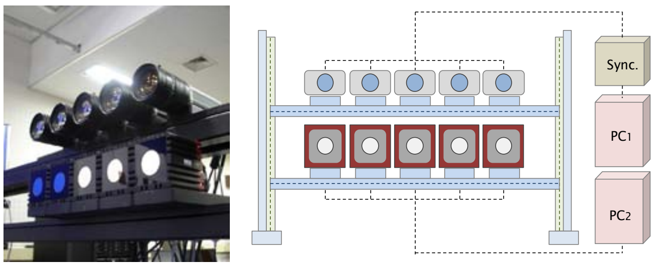

Fig. 10. Multi-depth camera system

Figure 11 shows the image set captured by the multi-depth camera system. The resolutions of the color and depth images are 1920x1080 and 176x144, respectively. Although we have three depth cameras, we can use three of them simultaneously since there are three different modulation frequencies to avoid interference between depth camera.

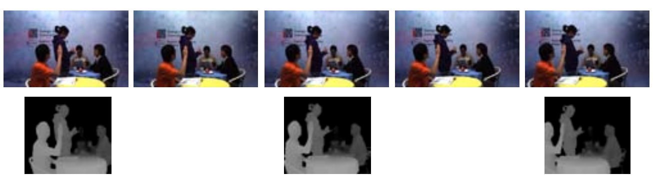

Fig. 11. Captured images by multi-depth camera system

For these images, we perform camera calibration, multi-view image rectification, and multi-view color correction to the color images. We also perform lens distortion correction, camera calibration, and depth correction to the depth images. We use the depth camera images as initial disparity values to increase the accuracy of the stereo matching. In order to compensate for the resolution difference between color and depth images, we use 3D warping.

The backprojected depth values of each pixel by using the depth of the depth camera images itself are then projected onto each color image positions. After obtaining the initial depth information, we generate the depth map of each view using the initial depth. Because there are a number of holes in the warped depth images, we first estimate the initial depth values at the hole regions. Then, we generate multi-view depth maps of the scene using stereo matching based on the initial depth values as shown in Fig. 12.

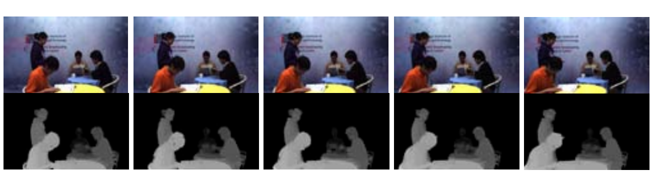

Fig. 12. Generated multi-view depth maps

4. CONCLUSION

In this paper, we have introduced the preprocessing technologies and explained three types of the multi-view camera systems to generate the multi-view HD depth video corresponding to the HD color. We have acquired multi-view sequences from our camera systems and generated the final depth maps using the stereo matching methods. Experimental results have shown that our systems produced multi-view high-resolution and high-quality depth maps using our systems. Finally, we have generated high-quality 3D contents video for realistic 3D broadcasting services. Therefore, our proposed system could be useful for various 3-D multimedia applications.

ACKNOWLEDGMENT

This research was supported by MKE under the ITRC support program supervised by NIPA (NIPA-2010-(C1090-1011-0003))

REFERENCES

[1] C. Fehn, R. Barre, and S. Pastoor, “Interactive 3-DTV-concepts and key technologies,” Proceedings of the IEEE, vol. 94, no. 3, pp. 524-538, March 2006.
[2] ISO/IEC JTC1/SC29/WG11 N8944, “Preliminary FTV model and requirements,” April 2007.
[3] A. Redert, M. op de Beeck, C. Fehn, W. IJsselsteijn, M. Pollefeys, L. van Gool, E. Ofek, I. Sexton, and P. Surman, “ATTEST – Advanced three-dimensional television system technologies,” Proceedings of International Symposium on 3D Data Processing Visualization and Transmission, pp. 313-319, June 2002.
[4] C. Zitnick and T. Kanade, “A cooperative algorithm for stereo matching and occlusion detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 7, pp. 675-684, August 2000.
[5] S. Nayar and Y. Nakagawa, “Shape from focus,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 8, pp. 824-831, 1994.
[6] S. Kim, S. Yoon, and Y. Ho, “Realistic broadcasting using multi-modal immersive media,” Lecture Notes in Computer Science, vol. 3768, pp. 164-175, November 2005.
[7] G. Um, K. Kim, C. Ahn, and K. Lee, “Three-dimensional scene reconstruction using multi-view images and depth camera,” Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XII, vol. 5664, pp. 271-280, January 2005.
[8] J. Zhu, L. Wang, R. Yang, and J. Davis, “Fusion of time-of-flight depth and stereo for high accuracy depth maps,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 231-236, June 2008.
[9] Camera calibration toolbox program for Matlab provided by Caltech, http://www.vision.caltech.edu/bouguetj/calib doc/
[10] E. Lee, Y. Kang, Y. Jung, Y. Ho, “3-D video generation using hybrid camera system,” International Conference on Immersive Telecommunications, pp. T5(1-6), June 2009.
[11] Y.S. Ho and Y.S Kang, “Multi-view depth generation using multi-depth camera system,” Proceedings of 3D Systems and Applications, pp. 67-70, May, 2010.