3D Scene Capturing using Stereoscopic Cameras and a Time-of-Flight Camera

Cheon Lee, Hyok Song, Byeongho Choi, and Yo-Sung Ho, *Senior Member*, IEEE

Abstract — *Since advanced 3D video systems employ multi-view generation techniques using depth data to provide more realistic and comfortable depth impression, many research works are related to depth generation exploiting a time-of-flight (TOF) camera. Most of them use time-consuming depth enhancement methods since the TOF camera generates noisy and unstable depth maps. In this paper, a 3D scene capturing method using stereoscopic cameras and one TOF camera is presented. The proposed method generates two viewpoints' depth videos corresponding to those of color cameras using depth up-converting and 3D warping techniques instead of using complex and time-consuming depth enhancement processes. In addition, the proposed method is implemented with the multi-thread technique to develop a fast system. Experimental results show that the proposed method generates two viewpoints' depth videos at 15 fps and those are used for generating multi-view images*[1].

Index Terms — 3D video, 3D scene capturing, time-of-flight camera, stereoscopic camera, 3D depth warping.

I. INTRODUCTION

Three-dimensional (3-D) video is a promising multimedia technique that leads the next multimedia services and applications. In fact, studies on 3D imaging date back hundreds of years starting with the description of stereopsis by Wheatstone [1]. Recently, with the commercial success of 3D movies in theaters, many 3D products are now releasing in the market such as 3D-TVs, 3D game consoles, 3D laptop PCs, and 3D mobile products. This trend will be continued by the convergence of various technologies; from computer graphics, computer vision, video compression, high-speed processing units, high-resolution displays and cameras [2].

The advanced 3D technologies are focusing on a comfortable 3D imaging and auto-stereoscopic displays using intermediate view generation combining a supplementary data. The depth-image-based rendering (DIBR) [3] is a popular approach for multi-view generation using depth data. Since the depth data describes the distance between the camera and objects in a scene, reconstructing a virtual viewpoint image is possible. Hence, the multi-view-video-plus-depth (MVD) format is widely used for the DIBR technique [4].

This virtual view generation technique gives a potential to reduce the number of viewpoints. For example, instead of transmitting nine-view videos to render with a nine-view 3D display, only two color views with corresponding depth data can be used; rest views are generated at the displays. Exploiting this concept, the moving picture experts group (MPEG) investigated the advanced 3D video systems which employ the multi-view video data to render various viewpoint images. As the first phase, the multi-view video coding (MVC) has been developed for compressing them efficiently by JVT (joint-video-team) [5]. It utilizes the redundancies between views by using inter-view prediction. Afterward, MPEG started standardization on the 3D video coding method of which input data include multi-view depth videos [6], [7]. Through intensive works, both depth estimation and view synthesis methods were investigated by many research institutes [8].

With respect to the recent research activities on 3D video, obtaining accurate depth map is regarded as an important problem to support the advanced 3D video systems. Although the stereo matching algorithms are investigated by huge number of works, estimating precise depth information for a texture-less region, e.g. white wall, is still a hard problem [9]. On the other hand, the hardware based depth sensing such as structured light patterns [10] and depth cameras [11] generate accurate depth data in real-time, whereas due to the expensive cost of equipment, it was hard to use for consumers. In recent years, however, relatively cheap time-of-flight (TOF) cameras with compact size have been introduced. Even some products are applied to a game console.

Recently, fusion camera systems consisted of multiple cameras and one or multiple TOF cameras have been proposed to improve the depth map. Lindner *et al*. [12] and Huhle *et al*. [13] developed fusion camera systems composed of one color camera and one TOF camera to reconstruct a 3D model. Kim *et al.* [14] and Hahne *et al*. [15] used high-resolution stereo cameras and one TOF camera to enhance the unstable depth information. Zhu *et al.* [16] presented a depth calibration method to improve depth accuracy using a depth probability distribution function. Lee *et al*. [17] improved depth quality using the segment-based depth map modification methods. Since the TOF camera is vulnerable to reflecting objects, these works are aimed at improving wrong depth values. Therefore, the complexity of process is too high to realize a consumer product.

We devised a fusion camera system with one TOF camera at center and two color cameras at both sides. The proposed system generates two color videos and their associated depth videos as shown in Fig. 1 to make MVD data. The

[1] This work was supported in part by the IT R&D program of MKE (10035337), and in part by the ITRC support program supervised by NIPA (NIPA-2011-(C1090-1111-0003)).

C. Lee and Y.S. Ho are with the School of Information and Communications, Gwangju Institute of Science and Technology (GIST), Gwangju, Korea (e-mail: leecheon@gist.ac.kr; hoyo@gist.ac.kr).

H. Song and B. Choi are with Multimedia IP Research Center, Korea Electronics Technology Institute (KETI), Seongnam-si, Kyeonggi-do 463-816 Gwangju, Korea (e-mail: hsong@keti.re.kr; bhchoi@keti.re.kr).

Contributed Paper
Manuscript received 07/15/11
Current version published 09/19/11
Electronic version published 09/19/11.

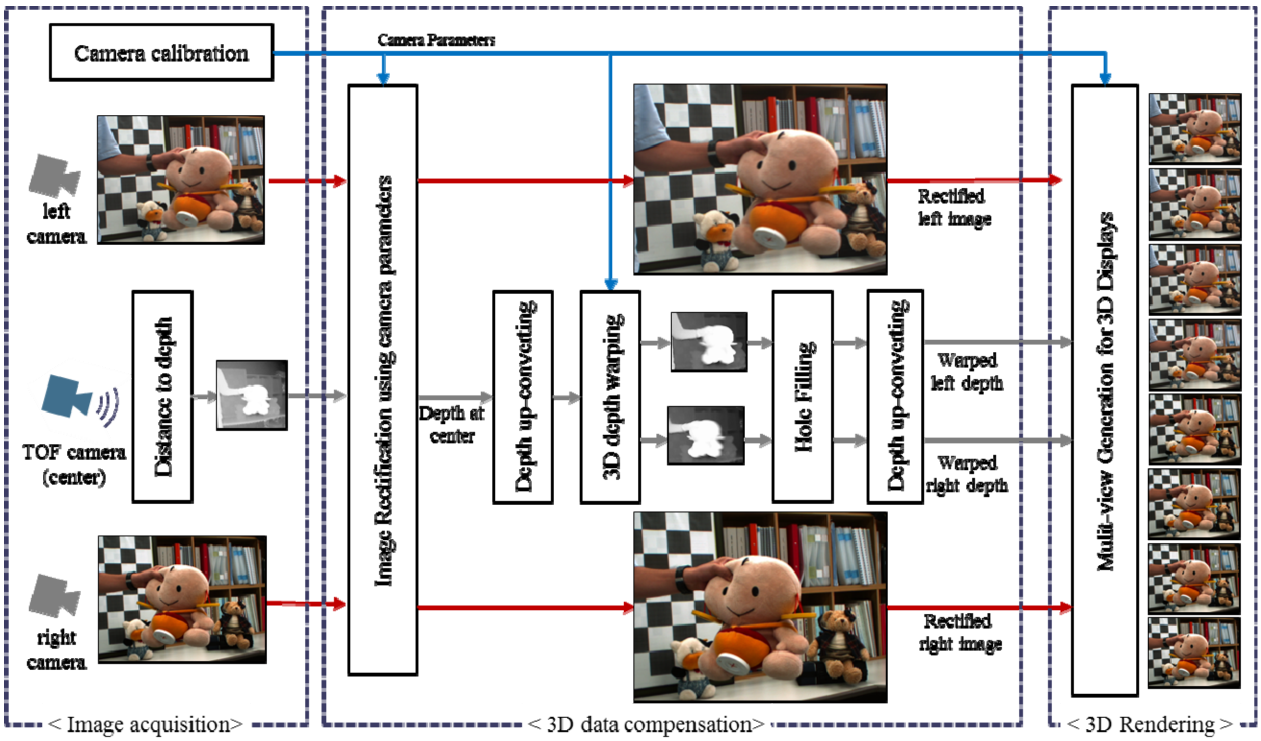


Fig. 1. Overall flows of the multi-view video generation using two color cameras and one TOF camera

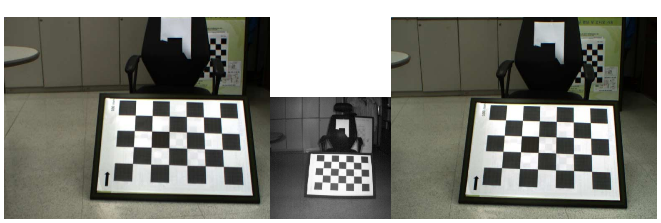
contribution of this work is development of a 3D scene capturing system performing with high-speed using the fusion camera system. This system is applicable to any 3D devices, e.g. 3D camera, 3D camcorder, and 3D mobile device.

The rest of this paper is organized as follows. In section II, the capturing process using the proposed system is described. In section III, two views' depth generating method is explained in detail. In section IV, the implementation of the proposed method and experimental results are presented. Concluding remarks are given in Section V.

II. JOINT STEREO AND DEPTH CAMERA SYSTEM

A. Joint Camera System for 3D Scene Capturing

The proposed joint camera system employs two high-resolution color cameras and one low-resolution TOF camera. The typical baseline distance between stereoscopic cameras is about 50 mm referring to the separation of human eyes. However, if the depth data for their color views are available, the baseline distance can be widened. In addition, since this system generates two depth videos using the 3D depth warping, which will be described in the next chapter, potential artifacts should be minimized. Considering these requirements, the TOF camera is centered by taking into account the volume of the camera, and then the rest color cameras are located at both sides close to the TOF camera.



(a) left camera (b) depth camera (c) right camera

Fig. 2. The pattern images for camera calibration

B. Multi-view Camera Calibration

Since the proposed camera system employs not a single camera but three cameras, the relationship between cameras need to be defined for further image processing such as image rectification, 3D image warping, and virtual view generation. The camera calibration is a process estimating the camera parameters of each camera using the captured pattern images as shown in Fig. 2 [18]. Estimating precise parameters directly affects the accuracy of the 3D warping. In particular, the geometric relationship between cameras is very important because the 3D warping technique utilizes them.

The camera parameters describe the relationship between camera coordinates and world coordinates. They consist of one intrinsic parameter **A** and two extrinsic parameters: the rotation matrix **R** and the translation vector **t**. If a point $\mathbf{M}_w = [X, Y, Z]^T$ in the world coordinate system is projected to the image coordinates, i.e. a pixel $\mathbf{m}=[u \ \ v]^T$, the relationship of two points is defined as Eq. (1).

$$\tilde{\mathbf{m}} = \mathbf{A} \cdot \mathbf{R} \cdot \mathbf{M}_w + \mathbf{A} \cdot \mathbf{t} \tag{1}$$

where $\tilde{\mathbf{m}} = \begin{bmatrix} u \cdot Z & v \cdot Z & Z \end{bmatrix}^T$ is the projected pixel position in the image coordinates in homogenous representation. Extending this concept, the point $\mathbf{M}_w$ is mapped to three pixel positions as:

$$\begin{cases} \tilde{\mathbf{m}}_L = \mathbf{A}_L \cdot \mathbf{R}_L \cdot \mathbf{M}_w + \mathbf{A}_L \cdot \mathbf{t}_L \\ \tilde{\mathbf{m}}_R = \mathbf{A}_R \cdot \mathbf{R}_R \cdot \mathbf{M}_w + \mathbf{A}_R \cdot \mathbf{t}_R \\ \tilde{\mathbf{m}}_D = \mathbf{A}_D \cdot \mathbf{R}_D \cdot \mathbf{M}_w + \mathbf{A}_D \cdot \mathbf{t}_D \end{cases} \tag{2}$$

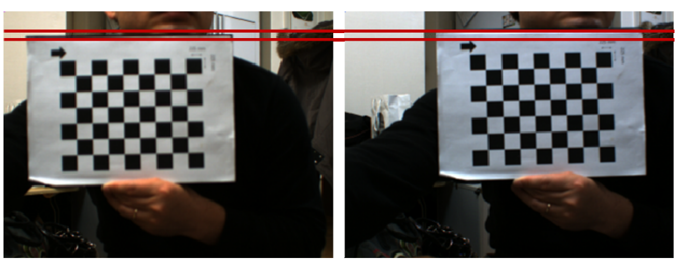
where the subscriptions L, R, and D represent the left and right color cameras and the TOF cameras, respectively.

C. Multi-view Image Rectification

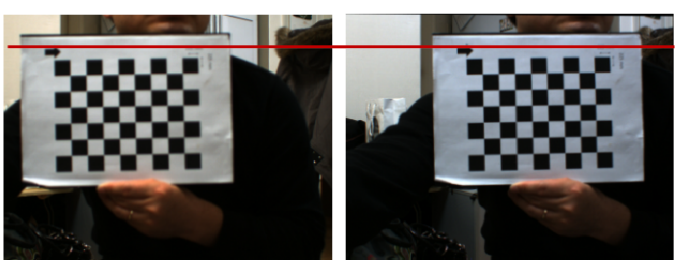
The geometric mismatches between views are induced by misalignment of multiple cameras. These errors induces serious problem in 3D video processing such as depth estimation and 3D image rendering. Therefore, theses geometrical mismatches between views should be minimized; generally, the image rectification is widely used. The image rectification is a compensation process making the epipolar lines aligned [19]. By conducting image rectification, two intrinsic parameters and two rotation matrixes become unified, whereas the translation vectors have different x-axis components due to the displacement of cameras as described in Eq. (3).

$$\begin{cases} \mathbf{A}'_L = \mathbf{A}'_R \neq \mathbf{A}_D \\ \mathbf{R}'_L = \mathbf{R}'_R = \mathbf{R}'_D \\ \mathbf{t}'_L \neq \mathbf{t}'_R \neq \mathbf{t}'_D \end{cases} \quad (3)$$

Differently from the color cameras, the resolution of the TOF camera is relatively small. Hence it is not suitable making the intrinsic parameter of the TOF camera to be unified. Keeping the intrinsic parameter of $\mathbf{A}_D$ as it is, only rotation matrix and translation vector are changed. As a result, the distance between $|\mathbf{t}_L - \mathbf{t}_D|$ and $|\mathbf{t}_R - \mathbf{t}_D|$ become unified; the only different factors are their x-components. Finally, the rectified images are obtained by transforming the image plane from the original image coordinates to the newly defined one. Figure 3 shows the results of image rectification of color images.



(a) horizontal offset between views: original images



(b) horizontally aligned image: rectified images

Fig. 3. Results of image rectification

D. Capturing Depth Image using TOF Camera

The depth image of the observed scene is obtained by quantizing the real distances into limited levels. The captured depth data is used for the virtual view generation at the display part; hence the depth image should be compatible with disparities of two stereo images. Converting a real depth value Z to a depth image should take the relationship between the baseline distance B and the disparity d into account as:

$$d = \frac{F \cdot B}{Z} \quad (4)$$

The baseline distance is defined as:

$$B = |\mathbf{t}_L - \mathbf{t}_R| \quad (5)$$

For the rectified stereo images, the baseline B is the absolute distance of two x-components of each translation vectors.

A depth level v in the depth image is defined by mapping the real depth value onto the possible levels as Eq. (6).

$$v = (2^N - 1) \cdot \left(\frac{d - d_{\min}}{d_{\max} - d_{\min}} \right) \quad (6)$$

where $d_{\min}$ and $d_{\max}$ represent the minimum disparity and the maximum disparity calculated by Eq. (4) according to the predefined depth range Z_{near} and Z_{far}, respectively. If the bit-depth of the depth image is N, the closest depth value Z_{near} is mapped to the highest value $(2^N\text{-}1)$, whereas the furthest depth value Z_{far} is mapped to the lowest value 0. Figure 4 shows an example the captured depth image with the depth range from 1 m to 3m.

Fig. 4. The captured depth image and its intensity image with depth range [1m ~ 3m].

III. Depth Video Acquisition for Two Views

In order to generate two depth videos corresponding viewpoints of the color cameras, the depth video captured by the TOF camera is transformed using the 3D warping method. Utilizing the depth information captured by the TOF camera, the 3D warping is conducted to each view for color cameras. This chapter describes the procedure of two viewpoints' depth data acquisition from a single TOF camera.

A. Pixel Correspondences between Views

The 3D image warping is a core technique for shifting the viewpoint from reference view to the virtual view by utilizing the provided depth data [3]. Using the depth data, all pixels in the reference image are projected back to the positions in the world coordinates. Then projecting to the virtual viewpoint, pixel correspondences between views are defined.

In order to find the corresponding position in the world coordinate system for a pixel in an image, the inverse equation of Eq. (1) is derived as:

$$\mathbf{M}_w = \mathbf{R}^{-1} \cdot \mathbf{A}^{-1} \cdot [\mathbf{m} \,|\, 1]^T \cdot Z(\mathbf{m}) - \mathbf{R}^{-1} \cdot \mathbf{t} \qquad (7)$$

where the scalar value $Z(\mathbf{m})$ is a real depth value derived by Eq. (6) as:

$$Z(\mathbf{m}) = \frac{1}{\dfrac{D(\mathbf{m})}{(2^N - 1)} \cdot \left(\dfrac{1}{Z_{near}} - \dfrac{1}{Z_{far}} \right) + \dfrac{1}{Z_{far}}} \qquad (8)$$

Note that defining pixel correspondence requires accurate camera parameters and depth data of the reference view. The image resolution is directly related to the intrinsic parameters, but those are not necessarily same between the reference view and the virtual view.

B. Camera Parameter Adjustment for Depth Warping

Ideally, mapping corresponding pixels between views are not constrained with the resolution of cameras. However, if the resolution of the reference view is smaller than the virtual view to be mapped, the warped image has lots of holes due to the lack of source elements. In order to avoid such problem, up-converting on the depth image is useful before the depth warping. Simultaneously, the camera parameters of the TOF camera need to be adjusted.

To define the camera parameters the color cameras for depth warping, the target resolution needs to be defined. In detail, only intrinsic parameters are changed, whereas the rotation and translation vector are unchanged. With the target resolution of depth image for warping, a scaling factor s is defined as:

$$s = \frac{\text{target resolution}}{\text{original resolution}} \qquad (9)$$

Then, the adjusted intrinsic parameter for warping is defined as:

$$\hat{\mathbf{A}} \equiv s \circ \mathbf{A}' \equiv \begin{bmatrix} s \cdot f & 0 & s \cdot p_x \\ 0 & s \cdot f & s \cdot p_y \\ 0 & 0 & 1 \end{bmatrix} \qquad (10)$$

where f is a focal length modified by the image rectification process, and p_x, p_y are the principle points of x-axis and y-axis, respectively.

The up-converting on the depth image before warping is conducted by the factor of 2; hence the size of depth image becomes 400x400. The target resolution of the color views for depth warping is down-converted to 320x240 from the original resolution 640x480, which is selected to minimize the holes by warping. Hence, the scaling factor s for color cameras is 0.5. Therefore, the adjusted intrinsic parameters are as:

$$\begin{cases} \hat{\mathbf{A}}_L = \hat{\mathbf{A}}_R = 0.5 \circ \mathbf{A}'_L = 0.5 \circ \mathbf{A}'_R \\ \hat{\mathbf{A}}_D \equiv 2 \circ \mathbf{A}_D \end{cases} \qquad (11)$$

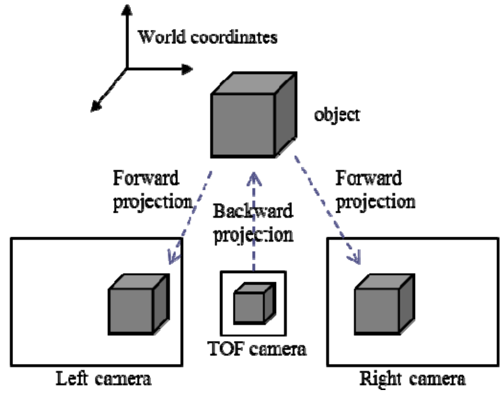


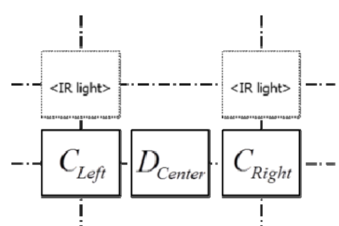
Fig. 5. Depth warping procedure

C. 3D Depth Warping using Adjusted Parameters

The 3D depth warping consists of two steps with using the adjusted camera parameters as illustrated in Fig. 5. The first step is finding the corresponding position in the world coordinates using Eq. (7) and the up-converted depth image by backward projection, and then project to the color cameras' viewpoints, respectively. If a projected pixel to left or right color viewpoint is out of the image frame, its corresponding depth value is neglected. Likewise, if two pixels in the depth image are projected to the same pixel of a color viewpoint, greater one is assigned since it is closer to the camera. Undefined pixels of the color viewpoints are the hole regions induced by the occluded areas in the depth viewpoint.

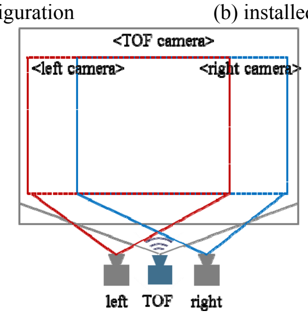
D. Hole Filling on Warped Depth Image

The warped depth images from center to left and right views have holes due to the viewpoint shifting. The holes consist of small holes and large holes; the former is generated by the rounding errors during warping, the latter is generated by occluded area at the reference view. The small holes are filled with a simple median filter using a 3x3 mask. The large holes are filled with the background depth values since most likely those are the background regions.



(a) system configuration

(b) installed system



(c) FOV (field-of-view) of each camera

Fig. 6. Stereo-depth camera system

Fig. 7. Demonstration of 3D scene capturing with the implemented software: the camera system is at the left side and the capturing software is displayed at the right side.

IV. Implementation and Results

A. The Proposed Joint Stereo Camera System

The proposed 3D cameras are configured as shown in Fig. 6(a) and 6(b). Two color cameras capture a VGA (640x480) image in real-time, i.e. 30 fps with a CCD sensor, and the TOF camera at the center captures depth information in a resolution 200x200 in real-time with a CCD sensor and two IR lights. The baseline distance between two color cameras is 104mm; hence the baseline distance between one color camera and the TOF camera is 52mm. The focal lengths of the color and the TOF camera are 8mm and 12.8mm, respectively. With this lens composition, a scene is captured as shown in Fig. 6(c).

B. Software Implementation

The proposed 3D scene capturing method as presented in Fig. 1 is implemented in C++ program language. Since the objective of the proposed method is development of a fast capturing system, the multi-threading technique is used as shown in Fig. 7. All five threads are performed simultaneously, but consuming time of each are different each other. *Thread*_1 operates capturing two color videos simultaneously in real-time. *Thread*_2 rectifies two captured images using their associated transformation matrixes. *Thread*_3 converts depth information into a depth image using Eq. (6) and rectifies it using the associated transformation matrix. *Thread*_4 and *Thread*_5 warp the captured depth image into the left and right cameras' viewpoints, respectively.

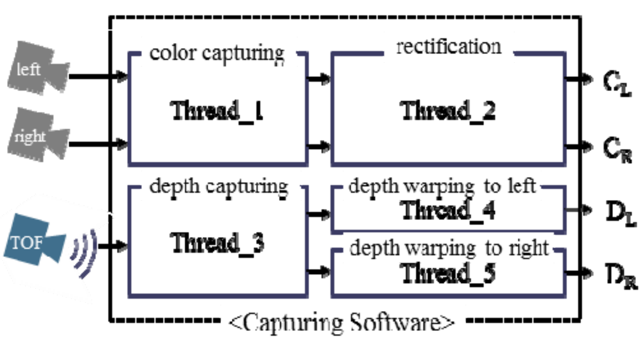
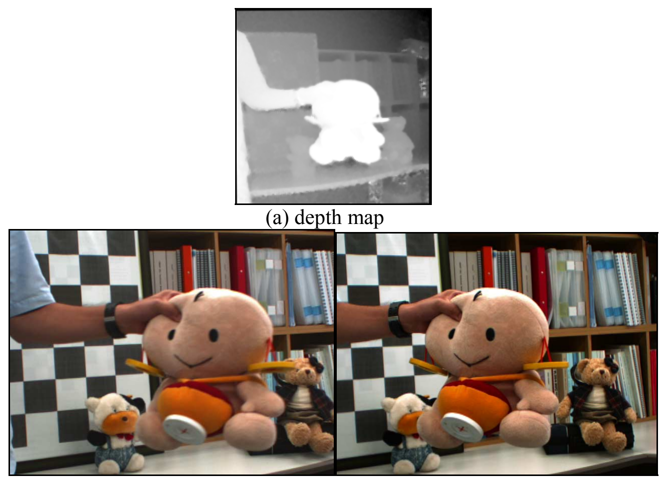


Fig. 8. Software implementation using multi-threads

The demonstration of the implemented software is shown in Fig. 8, where the joint camera system is at the left side and the capturing software is at the right side. Three frames are sampled at every five frames from the leftmost frame. Clearly, the positions of the man's arm between the color and depth images are identical; which means that the scene is captured near real-time, i.e. 15 fps.

C. Results of 3D Scene Capturing

We used 8-bits for depth image acquisition; hence N is 8 in Eq. (6) and (8). Figure 9 shows the captured stereo images and a depth image by two color cameras and one TOF camera,



(a) depth map

(b) left image (c) right image

Fig. 9. Three captured by the joint camera system

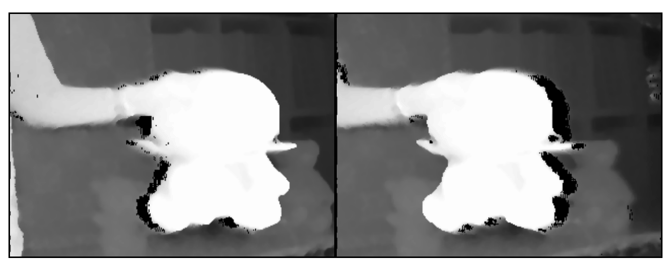


Fig. 10. Two warped depth images with holes

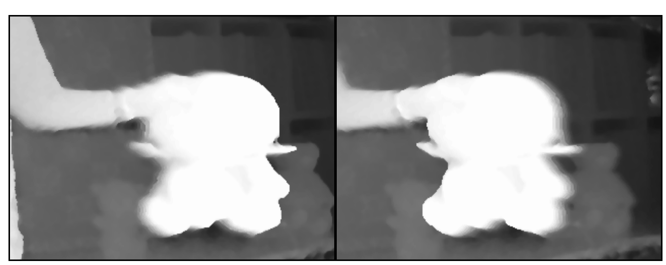


Fig. 11. Hole filled depth images

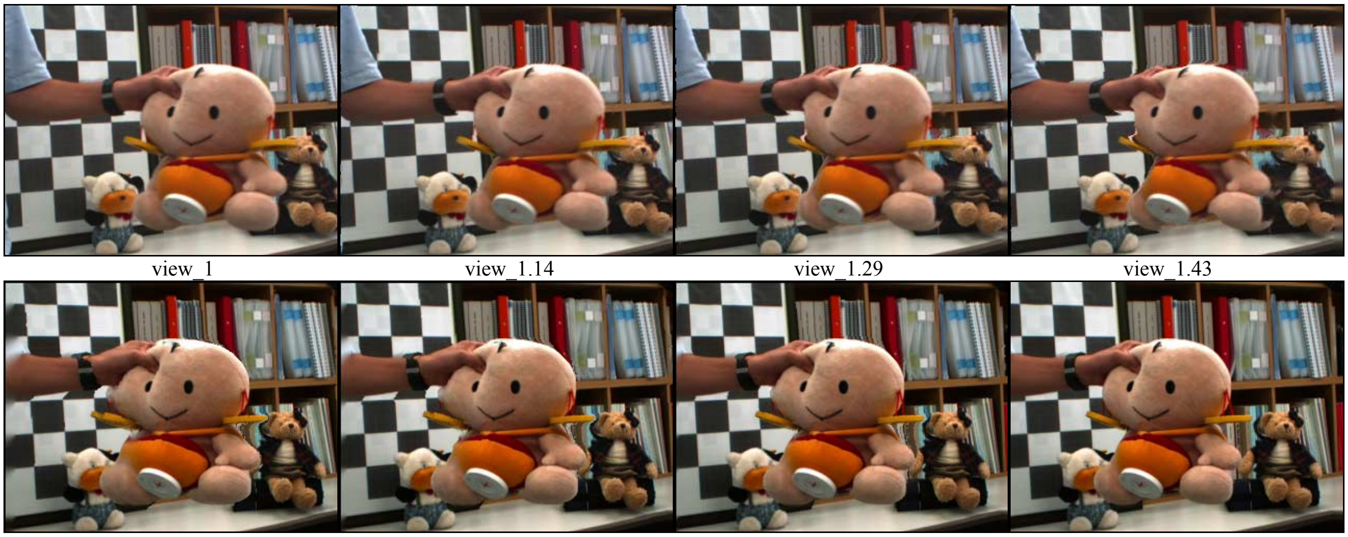


Fig. 12. Generated multi-view images; 6 intermediate virtual views are generated. The top-left and the bottom-right images are the original views captured by stereo camera.

respectively. Although the size of the depth image is relatively smaller than that of color image, its FOV is larger than the color image. Consequentially, the depth warping was performed from the center TOF camera to the left and right cameras as shown in Fig. 10. Hole regions painted in black are revealed due to the occluded regions at the center view. Figure 11 shows the hole filled depth images using the proposed method explained above.

Note that the positions of objects in the warped depth images are coincident with that of the color images. In practical, those positions are not precisely overlapped due to radial errors in the TOF sensor. However, we neglected them since the objective of this work is to design a fast capturing system. The improvement for this problem is included in further works.

D. Generated Multi-view Images

The output data of the proposed system are two color videos and their corresponding depth videos. With the camera parameters, those are used for generating multi-views as shown in Fig. 12. The top-left image the original left image and the bottom-right image is the original right image, other intermediate six images are the generated images between two original views. Overall procedures of multi-view generation are followed as described in [20], and the used hole filling method is the image inpainting [21]. As a result, throughout the eight views, the doll is located at a distance from one another. These multi-view images can be input data of the auto-stereoscopic displays, or can be rendered on the typical stereoscopic displays by choosing two proper viewpoint images.

E. Complexity Evaluation

The experiments for complexity evaluation were performed by a workstation, which consists of a Zeon E5630@2.53GHz processor, 32GB DDR2 RAM, and Windows 7 64-bit. Tested components are the capturing color video by stereo cameras, capturing depth video by the TOF camera, and the depth warping to each color's view. We tested 300 frames in total and calculated the speed of each part. Table I shows the results of the capturing speed. The color images for two views were captured at 30 fps, and the depth image for one view was captured at 35 fps. However, the warped depth images for two views were generated at 15 fps. It is because that the depth warping employs the matrix operation for each pixel including the backward projection and the forward projection. On the other hand, the hole filling process on the warped depth image was much faster than the depth warping.

V. CONCLUSION

In this paper, the 3D scene capturing method using a TOF camera is presented. The described method generates two views' depth map using one view's depth data using a

TABLE I
CAPTURING SPEED OF EACH PROCESS

Parts		Detailed Parts	Average Speed (ms)	Total Speed (ms)	Frame Rate (fps)
Part 1	Capturing Color Video by Stereo Camera	Image capturing by color cameras	17.900	33.540	29.815
		Image rectification	15.640		
Part 2	Capturing Depth Video by TOF Camera	Depth image capturing by TOF camera	28.230	28.963	34.529
		Depth image rectification	0.733		
Part 3	Depth Warping to Each Color's view	Depth warping from TOF to color views	64.340	66.750	14.981
		Hole filling on depth image	2.410		

centered TOF camera. In order to design a high-speed capturing system, we used the depth data as captured and up-converted them for 3D depth warping. After warping to the color viewpoints, holes are filled using a median filter and background depth values. The proposed capturing processes are implemented using the multi-thread technique. Experimental results showed that the positions of objects in the warped depth images are coincident with those of the color images. The warped two depth videos are generated at 15 fps. The output data are used for generating multi-view images for 3D displays. Therefore, the proposed method is able to apply to various 3D multimedia projects.

REFERENCES

[1] C. Wheatstone, "On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision," *Philos. Trans. R. Soc.,* vol. 54, pp. 196-199, 1838.

[2] A. Kubota, A. Smolic, M. Magnor, T. Chen, and M. Tanimoto, "Multi-View Imaging and 3DTV – Special Issue Overview and Introduction," *IEEE Signal Processing Magazine,* vol. 24(6), Nov. 2007.

[3] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV," *In Proceedings of the SPIE, Stereoscopic Displays and Virtual Reality Systems XI*, volume 5291, pages 93–104, 2004.

[4] A. Smolić and P. Kauff, "Interactive 3-D Video Representation and Coding Technologies," *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery,*, vol. 93, pp. 98-110, Jan. 2005.

[5] A. Smolić, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards," in *IEEE ICME*, pp. 2161-2164, 2006.

[6] ISO/IEC JTC1/SC29/WG11, Applications and Requirements on 3D Video Coding," *in MPEG output document* N11829, Jan. 2011.

[7] ISO/IEC JTC1/SC29/WG11, Call for Proposals on 3D Video Coding Technology," *in MPEG output document* N12036, March 2011.

[8] ISO/IEC JTC1/SC29/WG11, Report on Experimental Framework for 3D Video Coding," *in MPEG output document* N11631, Oct. 2010.

[9] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," *International Journal of Computer Vision,* vol. 47, pp. 7-42, April 2002.

[10] D. Scharstein and R. Szeliski, "High-accuracy Stereo Depth Maps using Structured Light," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* vol. 1, pp. 195-202, June 2003.

[11] G. J. Iddan and G. Yahav, "3D Imaging in Studio (and elsewhere...)," in *Proc. of SPIE Videometrics and Optical Methods for 3D Shape Mearuements*, pp. 48-55, 2001.

[12] M. Lindner, A. Kolb, and K. Hartmann, "Data-fusion of PMD-based Distance-information and High-resolution RGB-images," *International Symposium on In Signals, Circuits and Systems*, vol. 1, pp. 1-4, 2007.

[13] B. Huhle, S. Fleck, and A. Schilling, "Integrating 3D Time-of-flight Camera Data and High Resolution Images for 3DTV Applications," *3DTV-Conference,* pp. 1-4, 2008.

[14] Y. M. Kim, D. Chan, C. Theobalt, and S. Thrun, "Design and Calibration of a Multi-view TOF Sensor Fusion System," *Proc. of CVPR Conference*, 2008.

[15] U. Hahne and M. Alexa, "Depth Imaging by Combining Time-of-Flight and On-demand Dtereo," *in Dynamic 3D Imaging, DAGM Workshop, Dyn3D 2009,* pp. 70-83, 2009.

[16] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of Time-of-Flight Depth and Stereo for High Accuracy Depth Maps," *IEEE Computer Society Conference on In Computer Vision and Pattern Recognition*, pp. 1-8, June 2008.

[17] E. K. Lee and Y. S. Ho, "Generation of Multi-view Video using a Fusion Camera System for 3D Displays," *IEEE Transactions on Consumer Electronics,* vol. 56, pp. 2797-2805, Nov. 2010.

[18] R. Y. Tsai, "Versatile Camera Calibration Technique for High-accuracy 3D Machine Vision Metrology using Off-The-Shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation,* vol. RA-3, pp. 323-344, Aug. 1987.

[19] Y. S. Kang and Y. S. Ho, "Geometrical Compensation for Multi-view Video in Multiple Camera Array," *International Symposium ELMAR*, 2008.

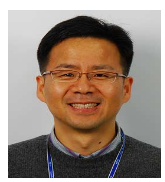
[20] C. Lee and Y. S. Ho, "View Synthesis using Depth Map for 3D Video," Asia-Pacific Signal and Information Processing Association, pp. 350-357, Oct. 2009.

[21] A. Telea, "An Image Inpainting Technique based on The Fast Marching Method," *Journal Graphics Tools,* vol. 9, pp. 25-36, May 2004.

BIOGRAPHIES

Cheon Lee (S'06) received his B.S. degree in Electronic Engineering and Avionics from Korea Aerospace University (KAU), Korea, in 2005 and M.S. degree in Information and Communication Engineering at the Gwangju Institute of Science and Technology (GIST), Korea, in 2007. He is currently working towards his Ph.D. degree in the Information and Communications Department at GIST, Korea. His research interests include digital signal processing, video coding, 3D video coding, 3D television and realistic broadcasting.

Hyok Song received his B.S. degree in Control and Instrumentation Engineering in 1999 and M.S. degree in Electronic Engineering in 2001 from Kwangwoon University (KWU), Korea. He joined Korea Electronics Technology Institute (KETI), where he was involved in the development of multi-view video, stereo vision, video recognition and other video systems. He is currently a Senior Researcher of MultimediaIP research center. He also is currently pursuing for the Ph.D. degree in the Department of Electronic Engineering at KWU. His research interests include digital image processing, and its application, especially such as 3DTV, stereo vision system.



Byeongho Choi received the B.S and M.S degrees in Electronic engineering from the University of Hanyang, Republic of Korea, in 1991 and 1993, respectively, and Ph.D. degree in the Department of Image Engineering from the University of Chungang, Republic of Korea, in 2010. From 1993 to 1997, he had worked for LG Electronics Co. Ltd as a junior researcher. In 1997, he joined Korea Electronics Technology Institute (KETI), where he was involved in the development of multi-view video, stereo vision and other video systems. He is currently a Managerial Researcher of SoC Research Center. His research interests include digital image processing, and its application, especially such as 3DTV, stereo vision system.

Yo-Sung Ho (M'81-SM'06) received both B.S. and M.S degrees in electronic engineering from Seoul National University (SNU), Korea, in 1981 and 1983, respectively, and Ph.D. degree in Electrical and Computer Engineering from the University of California, Santa Barbara, in 1990. He joined Electronics and Telecommunications Research Institute (ETRI), Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor New York, where he was involved in development of the advanced digital high-definition television (AD-HDTV) system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korea direct broadcast satellite (DBS) digital television and high-definition television systems. Since 1995, he has been with the Gwangju Institute of Science and Technology (GIST), where he is currently a professor in the Information and Communications Department. His research interests include digital image and video coding, image analysis and image restoration, advanced coding techniques, digital video and audio broadcasting, 3D television, and realistic broadcasting.