Generation of ROI Enhanced Depth Maps Using Stereoscopic Cameras and a Depth Camera

Sung-Yeol Kim, *Student Member, IEEE*, Eun-Kyung Lee, and Yo-Sung Ho, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a new scheme to generate region-of-interest (ROI) enhanced depth maps combining one low-resolution depth camera with high-resolution stereoscopic cameras. Basically, the hybrid camera system produces four synchronized images at each frame: left and right images from the stereoscopic cameras, a color image and its associated depth map from the depth camera. In the hybrid camera system, after estimating initial depth information for the left image using a stereo matching algorithm, we project depths obtained from the depth camera onto ROI of the left image using three-dimensional (3-D) image warping. Then, the warped depths are linearly interpolated to fill depth holes occurred in ROI. Finally, we merge the ROI depths with background ones extracted from the initial depth information to generate the ROI enhanced depth map. Experimental results show that the proposed depth acquisition system provides more accurate depth information for ROI than previous stereo matching algorithms. Besides, the proposed scheme minimizes inherent problems of the current depth camera, such as limitation of its measuring distance and production of low-resolution depth maps.**

***Index Terms*—Depth camera, depth map generation, ROI enhanced depth map, stereo matching.**

I. Introduction

THREE-DIMENSIONAL (3-D) video has been recognized as one of the essential parts for next-generation visual media. As one of the 3-D video representations, it is widely accepted that a monoscopic color video enriched with depth maps [1], which is often called as *video-plus-depth*, provides the groundwork for the envisaging 3-D applications due to backwards-compatibility to current 2-D digital systems and easy adaptability to a wide range different 2-D and 3-D displays. In general, we utilize depth-image-based rendering (DIBR) techniques to synthesize virtual views of a scene from video-plus-depth [2]. Recently, the ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) has also been interested in multi-view video with depth (MVD) [3] that is closely related to free-viewpoint TV (FTV) [4] and 3-D TV [5] to present more natural and realistic viewing experiences in the true dimension.

With respect to the current 3-D TV research activities, it is very important for us to estimate accurate depth information from natural scenes. In the field of computer vision and image processing, state-of-the-art depth estimation methods have been proposed to generate reliable depth maps [6]. However, accurate measurement of depth information from real scenes still remains problematic.

In general, we can classify depth estimation methods into two categories: active depth sensing and passive depth sensing. Passive depth sensing calculates depth information indirectly from 2-D images captured by two or more video cameras. Examples of passive depth sensing include shape from stereo [7], [8], shape from silhouette [9], and shape from focus [10]. The indirect method can provide depth information of far objects with a high resolution, while its accuracy is relatively lower than active depth sensing.

On the other hand, active depth sensing usually employs physical sensors, such as laser sensors, infrared ray (IR) sensors, or light pattern sensors, to directly obtain depth information from natural scenes. Structured light patterns [11] and time-of-flight depth cameras [12] are included in the direct method. Active depth sensing can only generate depths of nearby objects in a lower resolution, but it can produce more accurate depths in a shorter time than passive depth sensing.

Especially, although active range depth cameras, such as Z-Cam, developed by 3DV Systems, Ltd. [13] or NHK Axi-vision HDTV camera [14], can be only applied to capture indoor scenes, we can directly obtain depth maps in real time. A depth camera integrates a high-speed pulsed IR light source with a conventional broadcast TV camera to get color images and their associated per-pixel depth maps from real scenes simultaneously. However, even though the depth camera can produce accurate depth maps directly, there are some inherent technical problems in the current depth camera system.

The first problem is that *a depth map generated by the current depth camera usually includes optical noise*. Optical noise usually occurs as a result of differences in reflectivity of IR sensors according to color variation in objects. The second problem is that *the measuring distance of the current depth camera to capture depth information is limited*. In practice, the depth measuring distance of the current depth camera is approximately from 1m to 4m. As a result, we cannot obtain depth information from far objects. The last problem is that *the current depth camera can only produce low-resolution depth maps*. The image resolution of depth maps acquired by Z-Cam is 720 × 486 maximally.

One possible way to solve the problem of the current depth cameras is to upgrade them or develop a new depth camera.

Manuscript received March 28, 2008; revised June 18, 2008. Current version published December 04, 2008. This work was supported in part by ITRC through RBRC at GIST (IITA-2008-C1090-0801-0017).

The authors are with Department of Information and Communications, Gwangju Institute of Science and Technology, Gwangju 500-712, Republic of Korea (e-mail: sykim75@gist.ac.kr; eklee78@gist.ac.kr; hoyo@gist.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2008.2002338

0018-9316/$25.00 © 2008 IEEE

However, due to many challenges in real-time distance measuring systems, upgrading and improvements of depth cameras are very slow and expensive.

As an alternative, a fusion method that combines multiple video cameras and a depth camera has been introduced recently [15]. The hybrid camera system provided enhanced depth maps estimated by applying a stereo matching algorithm onto multi-view image with depth information captured by the depth camera. However, the depth acquisition system cannot produce high-resolution depth maps, because it completely depends on the low-resolution depth camera. In addition, hybrid camera sets integrating a time-of-flight depth camera and a high-resolution video camera have been presented to generate high-resolution range images [16], [17]. However, these previous works have mainly focused on a static 3-D scene reconstruction with the refined depth information acquired from the two-camera setup.

In this paper, we propose a new scheme to generate high-resolution depth maps by combining high-resolution stereoscopic cameras and a current low-resolution depth camera. With the hybrid camera system, we generate region-of-interest (ROI) enhanced depth maps by regarding the depth information obtained by the depth camera as the depth information of ROI in the left image obtained by the left camera of stereoscopic cameras.

Our main contribution is to develop a high-resolution scene depth generation scheme using the features of passive depth sensing and depth camera technology. Since matching failures on textureless or occluded regions, main unsolved problems in passive depth sensing, are more serious near to or in ROI than background, it is very difficult to get accurate ROI depths. In this work, a depth camera is used as a supplement to get accurate depth information for ROI. In addition, we provide an explicit solution to calculate the relative camera information in the hybrid camera system composed of different types of cameras. On one hand, the proposed hybrid camera system provides a practical solution to handle in-built problems of the current depth camera.

The next section will introduce the construction of the proposed hybrid camera system, and then briefly explain the overall framework to generate ROI enhanced depth maps. Section III will present algorithmic solutions for the estimation of depth maps from the images captured by the hybrid camera set. Experimental results will be shown in Section IV. Finally, we will make the conclusion in Section V.

II. The Proposed Hybrid Camera System

A. Construction of Hybrid Camera System

The hybrid camera system consists of high-resolution stereoscopic cameras and a low-resolution depth camera, as shown in Fig. 1. In addition, each camera in the hybrid camera system is connected to a personal computer equipped with a video capturing board. Besides, a clock generator is linked to the camera set to provide synchronization signals constantly. In this paper, we generate the depth map at the left camera using the hybrid camera set. Basically, we capture four synchronized 2-D images in each frame with the proposed hybrid camera set: left and right images from the stereoscopic cameras, and a color image and its associated depth map from the depth camera.

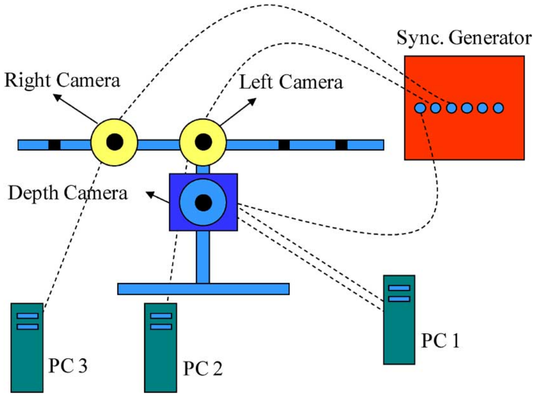


Fig. 1. Hybrid camera system.

B. Overall Framework

Fig. 2 describes the overall framework of the proposed ROI enhanced depth map generation. In order to clearly explain the methodology of the proposed scheme, we define image terminologies used in this paper in advance. In the defined image terminologies, a color image and a depth map naturally have the same resolution as the depth camera. On the other hand, the other images have the same resolution as the stereoscopic cameras.

- *Left image* is an image captured by the left camera.
- *Right image* is an image captured by the right camera.
- *Color image* is an image captured by the depth camera.
- *Depth map* is a depth map captured by the depth camera.
- *Initial disparity map* is a disparity map generated by a stereo matching algorithm with the left and right images.
- *Initial ROI disparity map* is a disparity map generated by a 3-D image warping operation with the depth map.
- *ROI disparity map* is a disparity map generated by a hole-filling algorithm with the initial ROI disparity map.
- *ROI enhanced disparity map* is the final disparity map.
- *ROI enhanced depth map* is the final depth map.

At the preprocessing stage, we calculate their relative camera information to the position of the depth camera, and reduce optical noise in the depth map using a depth data enhancing technique. In addition, the left and right images are rectified and color-segmented. Then, we apply a stereo matching algorithm on the rectified left and right images to obtain an initial disparity map of the left image.

Thereafter, in order to generate the initial ROI disparity map, we carry out 3-D image warping to move the depths captured by the depth camera into the world coordinate, and then reproject the warped depths onto the left camera. Next, depth holes in the initial ROI disparity map are removed by a hole-filling algorithm to generate an ROI disparity map.

Then, we create the ROI enhanced disparity map by merging the initial disparity map generated by stereo matching with the ROI disparity map. Finally, an ROI enhanced depth map is obtained with the final disparity map via a disparity to depth conversion.

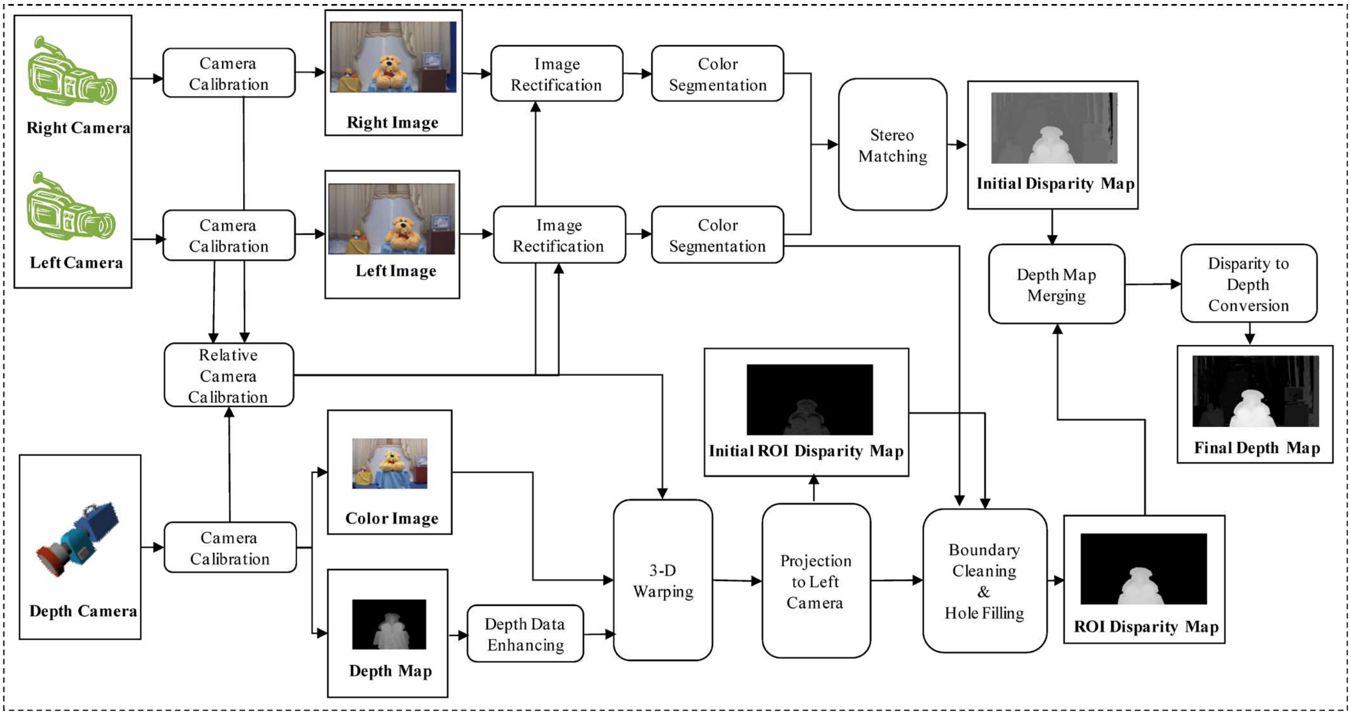


Fig. 2. Overall framework of the proposed depth map generation.

III. Generation of ROI Enhanced Depth Map

A. Preprocessing

Since we are employing two different types of cameras to construct the hybrid camera system, it is necessary to calculate relative camera information using camera calibration [18]. In order to get relative camera information, we first carry out the camera calibration algorithm three times. Hence, we can get three projection matrices for the three cameras as (1)

$$\begin{aligned} P_s &= K_s[R_s|t_s] \\ P_l &= K_l[R_l|t_l] \\ P_r &= K_r[R_r|t_r] \end{aligned} \tag{1}$$

where P_s is the projection matrix of the depth camera generated by its camera intrinsic matrix K_s, rotation matrix R_s, and transition matrix t_s. The term P_l and P_r indicate the projection matrices of the left and right cameras generated by their camera intrinsic matrices K_l and K_r, rotation matrices R_l and R_r, and transition matrices t_l and t_r, respectively.

Then, the left and right images are rectified using an image rectification algorithm [19]. Hence, the projection matrices P_l and P_r of the left and right images are changed as (2)

$$\begin{aligned} \widetilde{P}_l &= K_l' \left[R_l'|t_l\right] \\ \widetilde{P}_r &= K_r' \left[R_r'|t_r\right] \end{aligned} \tag{2}$$

where K_l' and K_r' are the changed camera intrinsic matrices for the left and right cameras by image rectification, respectively. The term R_l' and R_r' are the changed rotation matrices for the left and right cameras, respectively.

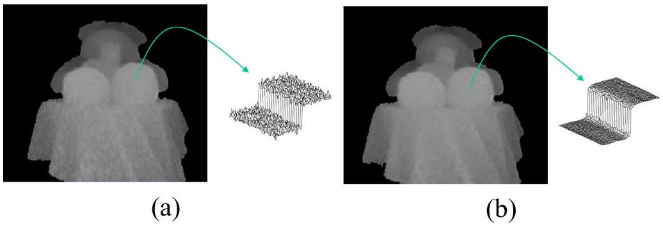


Fig. 3. Depth data enhancement. (a) Before; (b) after.

Thereafter, we convert the rotation matrix R_s of the depth camera into the identity matrix I by multiplying inverse rotation matrix R_s^{-1}. Thereafter, we convert the transition matrix t_s of the depth camera into the zero matrix O by subtracting the transition matrix t_s. Hence, we can define the new relative projection matrices for the left and right cameras on the basis of the depth camera as (3)

$$\begin{aligned} P_s' &= K_s[I|O] \\ \widetilde{P}_l' &= K_l' \left[R_l'R_s^{-1}|t_l - t_s\right] \\ \widetilde{P}_r' &= K_r' \left[R_r'R_s^{-1}|t_r - t_s\right] \end{aligned} \tag{3}$$

where P_s', $\tilde{P}_l'$, and $\tilde{P}_r'$ indicate the modified projection matrices of the depth camera, the left camera, and the right camera, respectively.

On one hand, it is necessary to reduce optical noise in the depth map. To minimize the noise, we use a depth data enhancing algorithm [20] that combines downsampling, bilateral filtering, and linear interpolation techniques. Fig. 3 shows the improved depth map after removing optical noise.

B. Generation of Initial Disparity Map

In order to create the initial disparity map, a stereo matching algorithm based on color segmentation is applied to the rectified

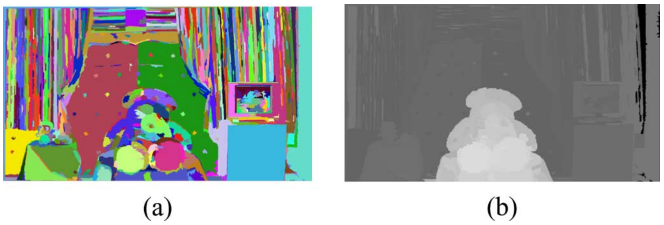


Fig. 4. Initial disparity map generation. (a) Color segmentation; (b) Initial disparity map.

left and right images. First, the rectified left and right images are color-segmented using a graph-based image segmentation algorithm [21]. In color segmentation, we carry out bilateral filtering prior to color segmentation to create more consistent color segments by removing noises included in the rectified left and right images. Fig. 4(a) shows the results of color segmentation for the rectified left image.

In stereo matching, we determine the disparity of each segment in the left image by computing the sum of absolute difference (SAD) with its corresponding region in the right image. Then, we refine the estimated disparity of each segment by considering disparities of its neighboring segments. Fig. 4(b) shows the initial disparity map of the left image.

In this paper, since we mainly focus on enhancing the ROI depths in the initial disparity map with the proposed hybrid camera system, any kind of high-performance stereo matching algorithm is acceptable to obtain the initial disparity map.

C. 3-D Image Warping

In order to generate the initial ROI disparity map, we use a 3-D image warping operation. In the step of 3-D image warping, we move the depths acquired by the depth camera to the world coordinate, and then reproject the warped 3-D data into the left camera. When $D_s(p_{sx}, p_{sy})$ is the depth information at the pixel position (p_{sx}, p_{sy}) in the depth map, we can regard the pixel $p_s(p_{sx}, p_{sy}, D_s(p_{sx}, p_{sy}))$ as a 3-D point. The corresponding point p_l of the left image is calculated by (4)

$$p_l = \tilde{P}_l' \cdot P_s'^{-1} \cdot p_s \tag{4}$$

where $\tilde{P}_l'$ and $P_s'^{-1}$ are the relative projection matrix of the left camera and the inverse relative projection matrix of the depth camera, respectively. Here, $p_l(p_{lx}, p_{ly}, 1)$ has the corresponding pixel position (p_{lx}, p_{ly}) of the pixel p_s in the left image.

In addition, the depth information $D_l(p_{lx}, p_{ly})$ of p_l is calculated by (5)

$$D_l(p_{lx}, p_{ly}) = \tilde{t}_{lz} + D_s(p_{sx}, p_{sy}) \tag{5}$$

where $\tilde{t}_{lz}$ indicates the third value of the relative transition matrix of the left camera.

Fig. 5 shows the initial ROI disparity map. When we compare the depth map with the initial ROI disparity map, we can notice that the body region is extended to fit with the high-resolution left image. We can also notice that holes occur in the initial ROI disparity map due to the warping operation.

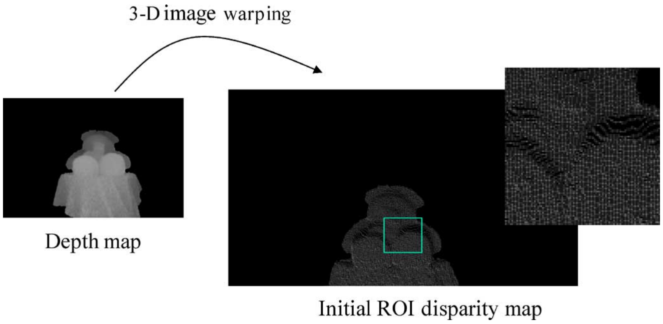


Fig. 5. Generation of initial ROI disparity map.

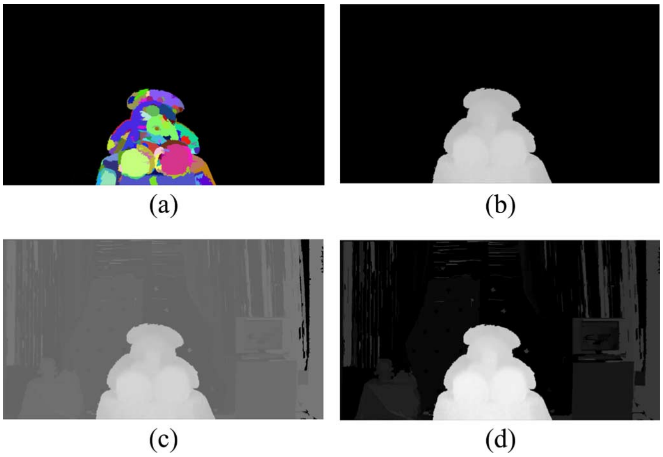


Fig. 6. ROI enhanced depth map generation. (a) Color segmentation set for ROI; (b) ROI disparity map; (c) final disparity map; (d) final depth map.

D. Generation of ROI Enhanced Depth Map

ROI of the left image and the initial ROI disparity map do not match correctly on the region of ROI boundaries. The main reason of the mismatch is the differences in reflectivity of IR sensors in the depth camera according to color values. Besides, the incorrectness of the camera calibration result can be the cause of the mismatch. In this paper, we solve the mismatch problem with the color segmented left image and the initial ROI disparity map.

In order to correctly detect ROI of the left image, we match the color segmented left image with the initial ROI disparity map. Then, we construct the color segment set for ROI from color segments of the left image by (6)

$$R(s_i) = \begin{cases} 1, & if \frac{n(A(s_i))}{n(s_i)} \geq 0.5 \\ 0, & otherwise \end{cases} \tag{6}$$

where $R(s_i)$ indicates whether the i^{th} color segment s_i of the color segmented left image is included in ROI of the left image or not. When $R(s_i)$ is 1, the corresponding color segment is included in the color segment set for ROI. The term of $n(s_i)$ is the total count of pixels in s_i, and $n(A(s_i))$ is the total count of pixels on the region of initial ROI disparity map $A(s_i)$ that is matched with the region of s_i. Fig. 6(a) shows the color segment set for ROI.

After ROI detection, we refine the initial ROI disparity map from the color segment set. We get rid of outside pixels of ROI

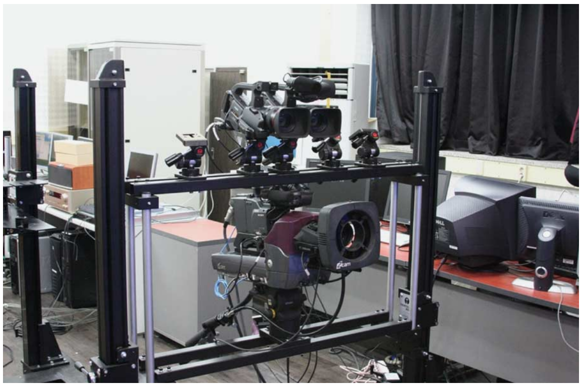


Fig. 7. Construction of the hybrid camera system.

from the initial ROI disparity map. As a result, we can get the refined initial ROI disparity map. Then, we fill holes in the ROI disparity map with the pixels generated by linearly interpolating with their neighboring pixels [22]. The hole-filling algorithm is performed as a unit of the color segment by (7)

$$I\left(R(x,y)\right)_k = \frac{1}{n} \cdot \sum_{i=0}^{W} \sum_{j=0}^{W} I\left(R(i,j)\right)_k \tag{7}$$

where $I(R(x,y))_{\mathrm{k}}$ is the interpolated pixel value at the (x,y) position of the k^{th} color segment in the refined initial ROI disparity map R using the valid neighboring pixel value $I(R(i,j))_{\mathrm{k}}$ in the k^{th} color segment. The term n is the valid number of pixels within a W × W window. Since the hole-filling algorithm is performed in a color segment, the valid depth pixels in its neighboring segments will not affect to fill the holes in the target color segment. Fig. 6(b) shows an ROI disparity map generated by the color segment-based hole-filling algorithm.

Next, we combine the initial disparity map with the ROI disparity map to generate an ROI enhanced disparity map. The ROI enhanced disparity map F is created by replacing the depth information of ROI in the initial disparity map H with the depth information of the ROI disparity map R by (8)

$$I\left(F(i,j)\right) = \begin{cases} I\left(H(i,j)\right), & if \quad I\left(R(i,j)\right) == 0 \\ I\left(R(i,j)\right), & otherwise \end{cases} \tag{8}$$

where $I(F(i,j)), I(H(i,j)$, and $I(R(i,j))$ are the depth values at the (i,j) position in F, H, and R, respectively. Fig. 6(c) shows an ROI enhanced disparity map.

Finally, disparity values in the ROI enhanced disparity map are converted into their depth values using disparity to depth conversion [23]. Fig. 6(d) shows the ROI enhanced depth map generated by the hybrid camera system.

IV. Experimental Results

In order to evaluate our scheme, we constructed a hybrid camera system with two HD cameras as stereoscopic cameras and one Z-Cam as the depth camera, as shown in Fig. 7. In our experiment, the distance that we can measure as depth information by Z-Cam is from 1.75 m to 2.15 m. The baseline distance between HD left and right cameras is 20 cm. Table I shows camera specifications for the experiment.

TABLE I
The Specifications of Hybrid Camera System

Devices	Specifications	Details
Stereo cameras (Canon XL-H1)	Output format	NTSC or PAL (16:9 ratio, HD)
Depth camera (Z-Cam)	Measured depth range	1.75m to 2.15m
	Field of view	40 degrees
	Output format	NTSC or PAL (4:3 ratio, SD)
Sync. generator	Output signal	SD/HD sync. signals

With the hybrid camera set, we have captured BEAR and ACTOR images as test data. The image resolution of left and right images of the test data captured by two HD cameras was 1920 × 1080, while the image resolution of color images and their depth maps captured by Z-Cam was 720 × 486. Fig. 8 shows the test images. Especially, since a big bear doll hugged a small one in BEAR images, we would keep observation on their estimated depths generated by various stereo matching algorithms and the proposed method. ACTOR images are composed of the 160 frames totally.

Once we have test images, we have estimated ROI depths at the HD left camera by applying the state-of-the-arts stereo matching methods, which are SAD [24], belief propagation [21], graph cuts [25], dynamic programming [26], scan-line optimization [6], and Poznan algorithm [27], on the HD left and right images. For background depths, we only used SAD based on color segmentation in our experiment. The estimated ROI depths and background depths were merged to generated final depth maps. On one hand, we also made a ground truth depth map for the left image of BEAR images by projecting the depth data acquired by a 3-D scanning device [28] at the HD left camera. Fig. 9(a) shows the ground truth depth map.

The generated depth maps on BEAR images using various stereo matching algorithms and the proposed method are shown in from Figs. 9(b)–9(h). In order to measure the performance of our scheme objectively, a quantitative analysis based on the ground truth comparison is used.

As the objective evaluation methodology, we used two quality measures based on known ground truth data [6]: root-mean squared (RMS) error R_E and the percentage of bad matching pixels B_A. Here, bad matching means that the depth value is different from the corresponding ground truth depth value by more than one pixel value.

Table II shows the result of RMS error R_E, the R_E difference between the stereo matching algorithms and the proposed method R_{Diff}, the percentage of bad matching pixels B_A, and the B_A difference between the stereo matching algorithms and the proposed method B_{Diff}.

As shown in Table II, when we compared the accuracy of ROI depths generated by the proposed method with belief propagation, which was the best one among stereo matching algorithms, the depth map produced by our hybrid camera system was more accurate by approximately 2.1 for R_E and 11.2% for B_A than belief propagation for BEAR images.

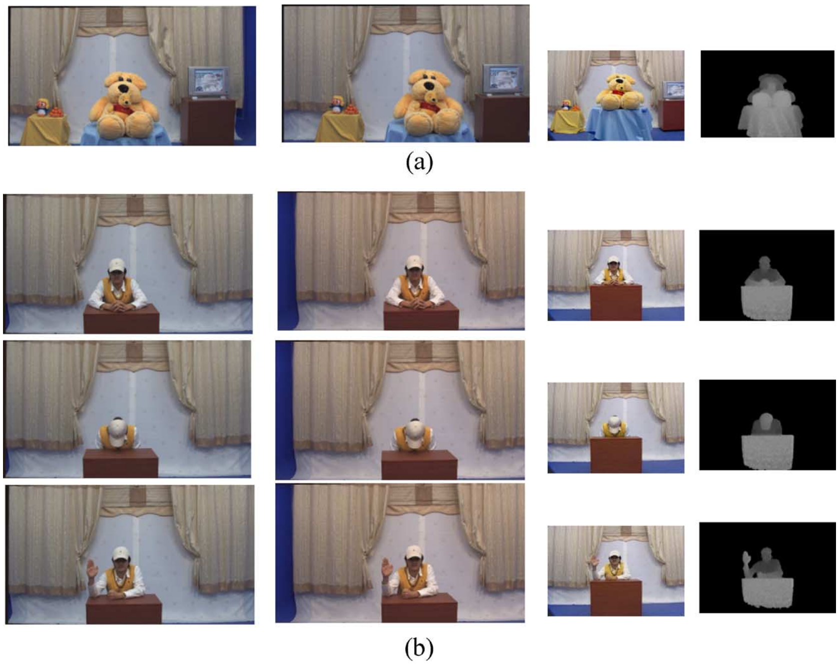


Fig. 8. Images captured by the hybrid camera system. (a) BEAR images; (b) ACTOR images.

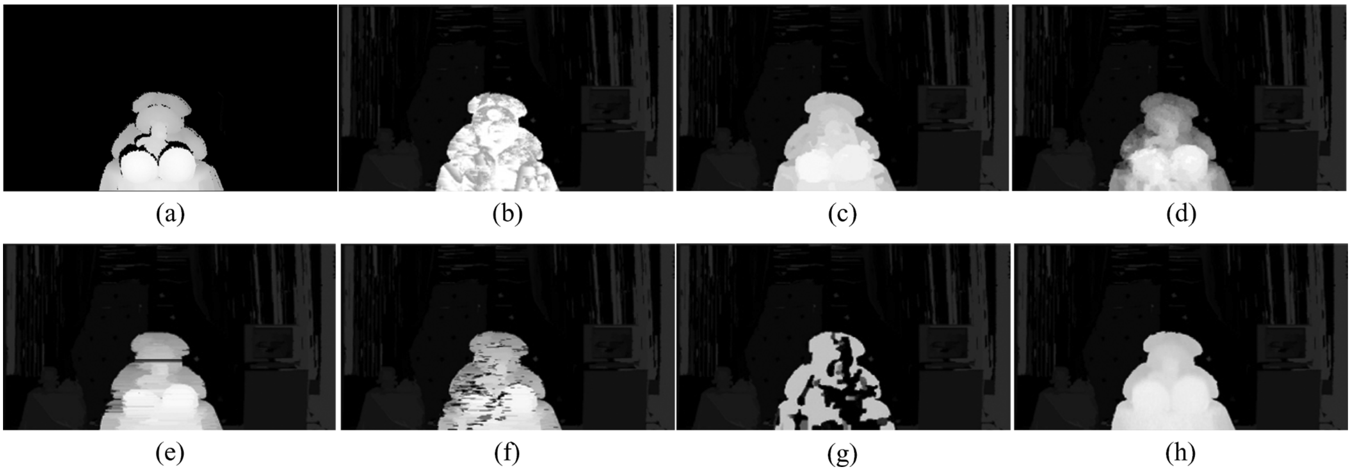


Fig. 9. Comparative results on BEAR images. (a) Ground truth; (b) SAD; (c) belief propagation; (d) graph cuts; (e) dynamic programming; (f) scanline optimization; (g) Poznan; (h) proposed method.

TABLE II
ROI DEPTH QUALITY EVALUATION

Methods	R_E	R_{Diff}	B_A	B_{Diff}
SAD	60.5	+36.1	85.3%	+46.4%
Belief gropagation	26.5	+2.1	50.1%	+11.2%
Graph cuts	62.1	+37.7	83.3%	+44.4%
Dyanaimc programimg	46.1	+21.7	76.7%	+37.8%
Scanline optimization	67.7	+43.3	79.5%	+40.6%
Poznan algorithm	132.4	+108.0	93.5%	+54.6%
Proposed method	24.4	-	38.9%	-

Fig. 10 shows the results of 3-D scene reconstruction on ROI of BEAR images. After extracting ROI depths from each depth map, we have made the 3-D scene on ROI using hierarchical decomposition [29]. As shown in Fig. 10, when we subjectively compare the scenes with the one generated with the ground truth for BEAR images, the 3-D scene generated by the proposed method more closely resembled the original scene than other methods. Especially, the regions of the big bear doll's leg and the small bear doll designated by circles in the original scene were much similar with ours. Hence, we subjectively notice that the depth map obtained by the proposed scheme has more reliable depth data than the full stereo matching methods.

Figs. 11 and 12 show the result of depth map generation with the 1st and 70th frames of the ACTOR images. In addition, Fig. 13 shows the result of 3-D scene reconstruction on the 70th frame of the ACTOR images for belief propagation and the proposed method. As shown in Fig. 13, ROI depths of the ACTOR images were still smoother than ones generated by the other methods. Hence, we could see that the regions represented

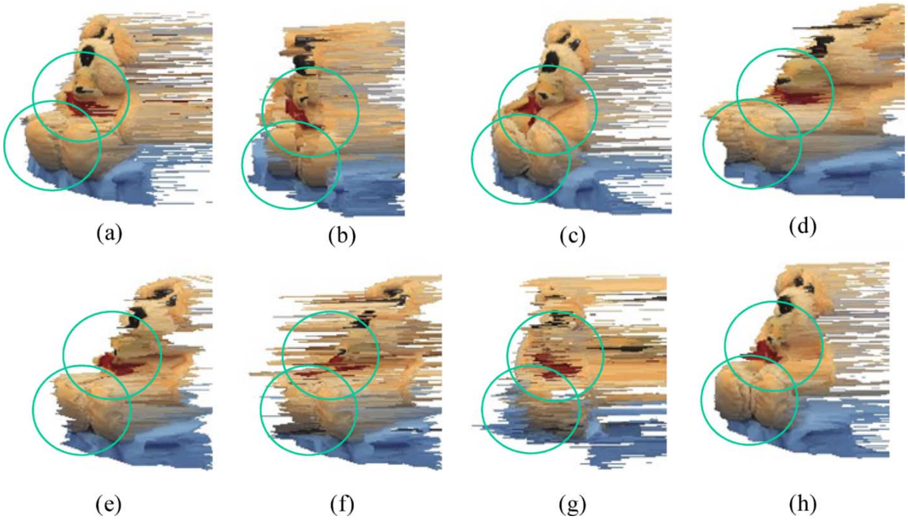


Fig. 10. 3-D scene reconstruction on ROI of BEAR images. (a) Ground truth; (b) SAD; (c) belief propagation; (d) graph cuts; (e) dynamic programming; (f) scanline optimization; (g) Poznan; (h) proposed method.

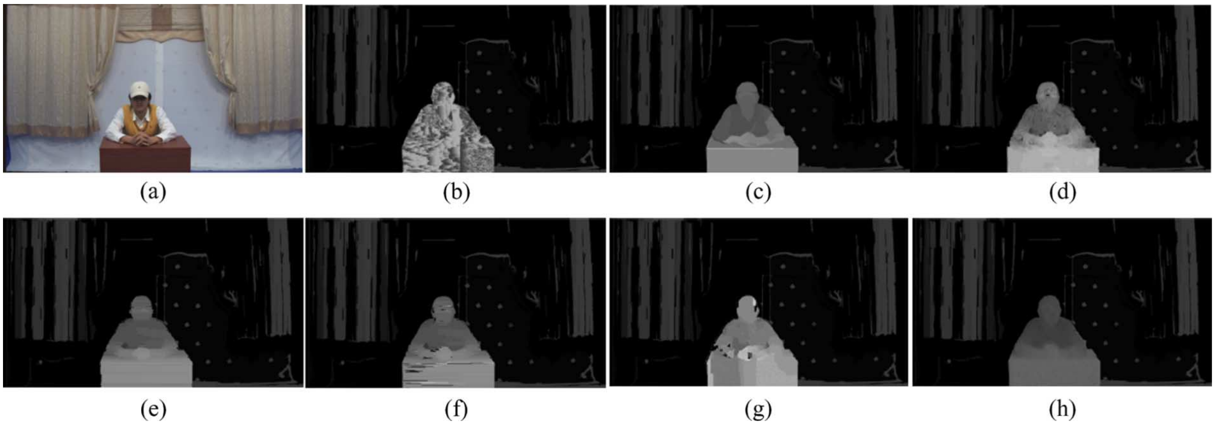


Fig. 11. Depth map generation on the 1st frame of ACTOR images. (a) Left image; (b) SAD; (c) belief propagation; (d) graph cuts; (e) dynamic programming; (f) scanline optimization; (g) Poznan; (h) proposed method.

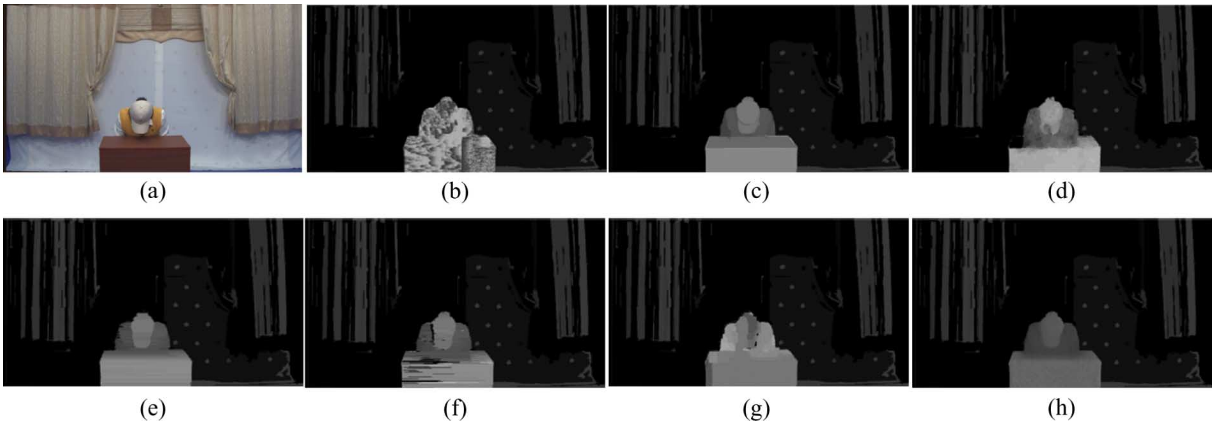


Fig. 12. Depth map generation on the 70th frame of ACTOR images. (a) Left image; (b) SAD; (c) belief propagation; (d) graph cuts; (e) dynamic programming; (f) scanline optimization; (g) Poznan; (h) proposed method.

by the mismatched depths on ROI were notably reduced by the proposed scheme.

Furthermore, although the image resolution of input depth maps captured by Z-Cam was 720 × 486, the image resolution of the output depth maps generated by proposed method was 1920 × 1080. Since we have projected the depth camera data into the high-resolution left camera, the image resolution of the ROI enhanced depth map was equal to the image reso-

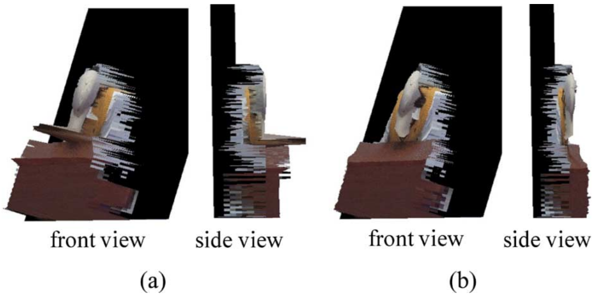


Fig. 13. 3-D scene reconstruction on ACTOR images. (a) Belief propagation; (b) proposed method.

lution of the high-resolution left camera. As a result, we could successfully generate high-resolution depth maps using the current low-resolution depth camera.

V. Conclusions

In this paper, we have proposed a new scheme to generate high-quality and high-resolution depth maps using a hybrid camera system. With the hybrid camera system, we could solve inherent technical problems in the currently available depth camera system. Especially, we have presented a 3-D image warping technique and a disparity map merging scheme to generate ROI enhanced depth maps. We have noted it by the objective and subjective evaluations that the proposed scheme could produce more accurate ROI depths than conventional stereo matching algorithms. The proposed scheme also generated higher resolution depth maps than what current depth cameras could produce. We hope that the proposed hybrid camera system can present new directions for further research related to depth estimation and will be used in future 3-D multimedia applications.

References

[1] C. Fehn, "A 3D-TV system based on video plus depth information," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, 2003, vol. 2, pp. 1529–1533.

[2] L. Zhang and W. Tam, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 191–199, 2005.

[3] *Preliminary FTV Model and Requirements*, ISO/IEC JTC1/SC29/WG11 N8944, 2007.

[4] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Processing Image Communication*, vol. 22, no. 2, pp. 217–234, 2007.

[5] C. Fehn, R. de la Barré, and S. Pastoor, "Interactive 3-DTV—concepts and key technologies," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 524–538, 2006.

[6] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Jour. of Computer Vision*, vol. 47, no. 1–3, pp. 7–42, 2002.

[7] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," in *Proc. of SIGGRAPH*, 2004, pp. 600–608.

[8] N. Atzpadin, P. Kauff, and O. Schreer, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 321–334, 2004.

[9] A. Hengel, A. Dick, T. Thormahlen, B. Ward, and P. H. S. Torr, "Video-trace: Rapid interactive scene modelling from video," in *Proc. of SIGGRAPH*, 2007, article 86.

[10] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[11] M. Waschbüsch, S. Würmlin, D. Cotting, and M. Gross, "Point-sampled 3D video of real-world scenes," *Signal Processing Image Communication*, vol. 22, no. 2, pp. 203–216, 2007.

[12] S. M. Kim, J. Cha, J. Ryu, and K. H. Lee, "Depth video enhancement of haptic interaction using a smooth surface reconstruction," *IEICE Trans. Information and System*, vol. E89-D, pp. 37–44, 2006.

[13] G. J. Iddan and G. Yahav, "3D Imaging in the Studio and Elsewhere. . .," in *Proc. of SPUE Videometrics and Optical Methods for 3D Shape Measurements*, 2001, pp. 48–55.

[14] M. Kawakita, T. Kurita, H. Kikuchi, and S. Inoue, "HDTV axi-vision camera," in *Proc. of International Broadcasting Conference*, 2002, pp. 397–404.

[15] G. Um, K. Y. Kim, C. Ahn, and K. H. Lee, "Three-dimensional scene reconstruction using multi-view images and depth camera," in *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XII*, 2005, vol. 5664, pp. 271–280.

[16] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," *Proc. of Advances in Neural Information Processing systems*, pp. 291–298, 2005.

[17] B. Huhle, S. Fleck, and A. Schilling, "Integrating 3D time-of-flight camera and high resolution images for 3DTV applications," presented at the Proc. of 3DTV conference, 2007, paper no. 89, unpublished.

[18] "Camera Calibration Toolbox Program for Matlab Provided by Caltech," [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/

[19] Y. S. Kang, C. Lee, and Y. S. Ho, "An efficient rectification algorithm for multi-view images in parallel camera array," in *Proc. of 3DTV Conference*, 2008, pp. 61–64.

[20] J. H. Cho, I. Y. Chang, S. M. Kim, and K. H. Lee, "Depth image processing technique for representing human actors in 3DTV using single depth camera," presented at the Proc. of 3DTV Conference, 2007, paper no. 15, unpublished.

[21] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International Jour. of Computer Vision*, vol. 70, no. 1, pp. 41–54, 2006.

[22] S. U. Yoon and Y. S. Ho, "Multiple color and depth video coding using a hierarchical representation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1450–1460, 2007.

[23] *Description of Exploration Experiment in 3D video*, ISO/IEC JTC1/SC29/WG11 N9596, 2008.

[24] H. Hirschmuller, "Improvements in real-time correlation-based stereo vision," *Proc. of Stereo and Multi-Baseline Vision*, pp. 141–148, 2001.

[25] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. of International Conference on Computer Vision*, 2001, pp. 508–515.

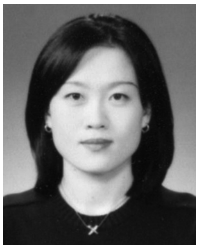
[26] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *International Jour. of Computer Vision*, vol. 33, no. 3, pp. 181–200, 1999.

[27] *Depth Map Estimation Software*, ISO/IEC JTC1/SC29/WG11 M15175, 2008.

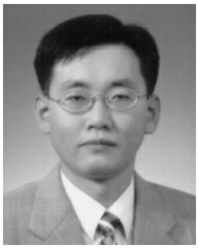
[28] "LMS-Z390i," [Online]. Available: http://www.riegl.com/

[29] S. Y. Kim, S. B. Lee, and Y. S. Ho, "Three-dimensional natural video system based on layered representation of depth maps," *IEEE Trans. Consumer Electronics*, vol. 52, no. 3, pp. 1035–1042, 2006.

Sung-Yeol Kim (S'06) received his B.S. degree in Information and Telecommunication engineering from Kangwon National University (KNU), Korea, in 2001 and M.S. degree in Information and Communication Engineering at the Gwangju Institute of Science and Technology (GIST), Korea, in 2003. 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 mesh representation, 3D mesh compression, 3D television, and realistic broadcasting.



Eun-Kyung Lee received both B.S. and M. S. degree in computer engineering from Honam University (HU), Korea, in 2002 and 2004, respectively. She is currently working towards her Ph.D. degree in the Information and Communications Department at the Gwangju Institute of Science and Technology (GIST), Korea. Her research interests include digital signal processing, multi-view video coding algorithms and systems, multi-view depth map generation, 3D television, and realistic broadcasting.



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 the 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.