3DTV: Technical Challenges for Realistic Experiences

Yo-Sung Ho and Eun-Kyung Lee

Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Korea
E-mails: {hoyo, eklee78}@gist.ac.kr

ABSTRACT

In recent years, various multimedia services have become available and the demand for three-dimensional television (3DTV) is growing rapidly. Since 3DTV is considered as the next generation broadcasting service that can deliver realistic and immersive experiences, a number of advanced 3D video technologies have been studied. In this paper, we are going to explain the fundamental principles of 3DTV. After reviewing the basic techniques for 3D image capturing and 3D video display systems, we are going to cover several technically challenging issues of 3D video processing, such as camera calibration, image rectification, illumination compensation and color correction.

Keywords: 3DTV; Realistic broadcasting system; Multi-view video-plus-depth data; Time-of-flight camera; Multi-view camera system; Free viewpoint TV; View synthesis.

1. INTRODUCTION

Owing to the rapid growth of various digital technologies, broadcasting services [1] has been changed from unidirectional services to bidirectional services or interactive services, such as stereoscopic TV [2], three-dimensional (3D) TV [3], and realistic broadcasting [4]. As shown in Fig. 1, the next-generation broadcasting system is supposed to provide a variety of user-friendly interactive information, as well as high-quality audio-visual broadcasting contents.

Especially, 3DTV is considered as a main theme for the future broadcasting system supporting natural viewing experience in the true three dimension. In general, 3D natural views are usually created from two 3D video representations: multi-view video [5] and video-plus-depth [6]. A multi-view video represents the 3D scene with the collection of multiple videos generated by capturing the scene at different camera locations. Since the multi-view video produces natural 3D views with a number of images at the viewing position, we can be easily immersed in the 3D content. However, we need to put more efforts to control a huge number of cameras at the same time. Moreover, since the multi-view camera system usually requires complicated coding and transmission schemes [7] in proportional to the number of cameras, it is hard to send its data to the receiver side within limited bandwidth channel environments.

As an alternative for the 3D video representation, it is widely accepted for a monoscopic color video enriched with per-pixel depth information, which is often referred as video-plus-depth. Since the video-plus-depth representation includes depth information as geometry data of the scene, we can generate free-viewpoint images using depth image-based rendering (DIBR) techniques [8] for the 3D video contents service. Although the video-plus-depth approach can support narrow-viewing angle views in comparison to the multi-view video, it is considered as a suitable 3D video representation for 3DTV because it can support both backwards compatibility to the current 2D digital systems and easy adaptability to a wide range of different 2D and 3D displays. Recently, the ISO/ICE JTC/SC29/WG11 Moving Picture Experts Group (MPEG) has also been interested in multi-view video with depth (MVD), which is the combination of the multi-view video and the video-plus-depth approaches, for free-viewpoint TV (FTV) and 3DTV[9] [10].

With respect to the current 3DTV and FTV research activities, it is very important for us to estimate accurate depth information from the natural scene. In the field of computer vision and image processing, a number of depth estimation algorithms have been proposed to generate accurate depth maps [11][12]. However, accurate measurement of depth information from the natural scene still remains problematic.

In general, there are two approaches to acquire depth information: depth from active sensor depth camera system and depth estimation from stereo matching. The latter takes a longer time and is more complex. In spite of its complexity, it does not guarantee accuracy of the estimated depth. On the other hand, as sensor technologies for obtaining depth information are developed rapidly, we can capture more accurate per-pixel depth information from the real scene directly using a depth camera system. However, the depth camera system has disadvantages: high cost and limited viewing range. Therefore, we need to develop a multi-view camera system to solve these problems.

To solve these problems, hybrid camera systems have been proposed to generate enhanced depth maps by applying a stereo matching algorithm to multi-view images with depth

ISBN: 978-602-8944-11-3

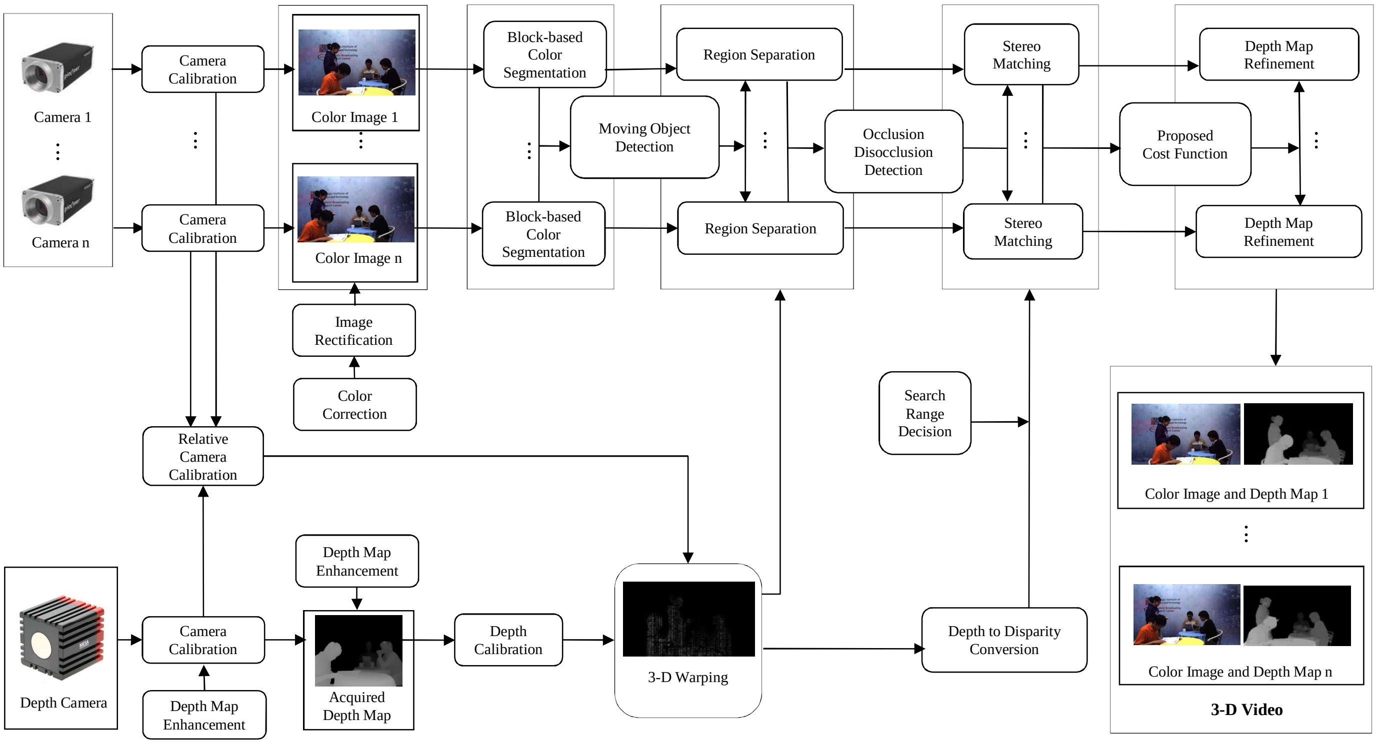


Fig. 1: Overall framework of multi-view image generation.

using the depth information captured by the depth camera [13]. However, these systems cannot produce high-resolution depth maps, because it completely depends on the low-resolution depth camera.

In this paper, we devise a hybrid camera system with one depth camera and multiple video cameras. The proposed system can produce multi-view images for dynamic 3D scenes by enhancing the low-resolution depth information measured by the depth camera. The main contribution of our work is to provide a practical 3D video generation solution for dynamic 3D scenes, which is applicable to 3D consumer devices.

2. HYBRID CAMERA SYSTEM

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

Figure 1 illustrates the overall framework to generate multi-view video sequences with their corresponding depth maps using the hybrid camera system. After calibrating each camera independently, we perform an image rectification to adjust vertical mismatches in multi-view images. Then, we apply a color correction operation to maintain color consistency among multi-view images. Figure 2 shows the proposed hybrid camera system and configuration.

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

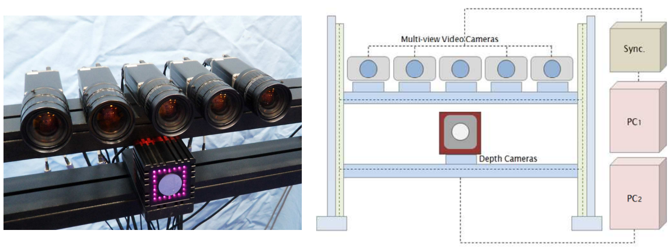


Fig. 2: Proposed hybrid camera system..

2.1 Relative Camera Calibration

Since the proposed fusion camera system consists of two different types of cameras, a depth camera and stereo video cameras, it is essential to find out relative camera information through camera calibration. For that, we apply a camera calibration algorithm to each camera in our camera system and obtain projection matrices for the depth camera and each video camera.

$$P_s = K_s[R_s \mid t_s] \quad (1)$$

$$P_k = K_k[R_k \mid t_k] \quad (2)$$

where P_s is the projection matrix of the depth camera represented by its camera intrinsic matrix K_s, rotation matrix R_s, and translation vector t_s. P_k means the projection matrices of the k^{th} video camera which consisted of its camera intrinsic matrix K_k, rotation matrix R_k, and translation vector t_k.

We then employ a rectification operation. The cameras have geometric errors because they are set manually by hand. In order to minimize the geometric errors, we find the common baseline, and then apply the rectifying transformation to the stereo image. Consequently, the projection matrix of video cameras are changed as

$$P'_k = K'_k[R'_k \mid t'_k] \quad (3)$$

where $K_k{}'$ and $R_k{}'$ are the modified camera intrinsic matrix and rotation matrix of the k^{th} video camera, respectively. 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}. The translation vector t_s of the depth camera is also changed into the zero matrix O by subtracting the translation vector t_s. Hence, we can define new relative projection matrices for the stereo camera on the basis of the depth camera as

$$P'_s = K_s[I \mid O] \quad (4)$$

$$\tilde{P}'_k = K'_k[R'_k R_s^{-1} \mid t_k - t_s] \quad (5)$$

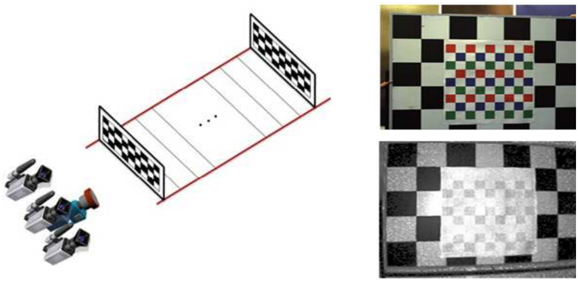
where $P_s{}'$ and $\tilde{P}'_k$ are final projection matrices of the depth camera and the k^{th} video camera, respectively. After relative camera calibration, we resolve the color mismatch problem of stereo images using a color calibration method. The color characteristics of captured images are usually inconsistent due to different camera properties and lighting conditions even the hardware type and specification of the multiple cameras are the same. Thereafter, we perform bilateral filtering to reduce optical noises included in the depth map acquired from the depth camera.

2.2 Depth Calibration

The depth values measured by the 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.

For depth calibration in indoor environments, we compute the depth of the planar checker pattern within the limited space by increasing the distance from the image pattern to the depth camera using our system as shown in Fig. 3. To extract the corresponding feature points in two different types of cameras efficiently, we use the color checker pattern. The pattern image is captured in every 5*cm* distance. The plane pattern is orthogonal to the image plane.

Since the proposed fusion camera system consists of two different types of cameras, a depth camera and stereo video cameras, it is essential to find out relative camera information through camera calibration. For that, we apply a camera calibration.



(a) Pattern acquisition (b) Pattern images from hybrid camera system

Fig. 3: Acquisition of the planar check pattern for depth calibration.

Thereafter, we make a four dimensional look-up table (LUT) mapping 3D positions of the multiple video cameras and the depth value from the depth camera. 3D position is constructed by x, y position of the feature point and the real depth value calculated by the multi-view image. Depth accuracy test using the acquired depth map and calibrated depth map the real depth value z calculated from the multi-view image by pairwise stereo matching. Since we have already obtained camera parameters, the real depth value is calculated by

$$d_n(p_x, p_y) = \frac{K \cdot B}{D_n(p_x, p_y)} \quad (6)$$

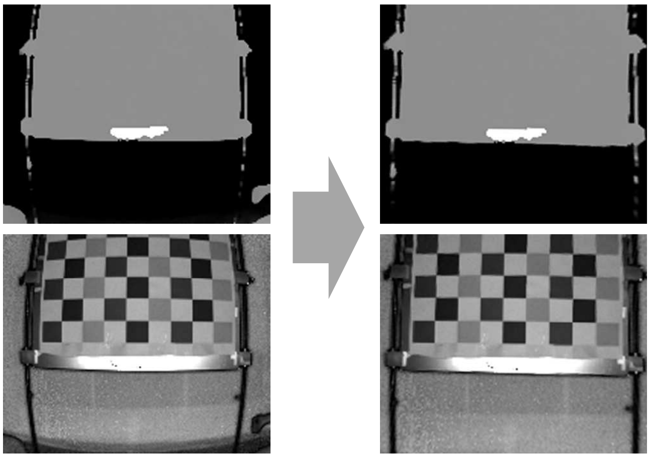
where K is the focal length of the left camera and B is the baseline distance between two video cameras. $D_n(p_x, p_y)$ is the real depth value corresponding to the measured disparity value $d_n(p_x, p_y)$ at the pixel position (p_x, p_y) in the checker pattern.

2.2 Radial Distortion Correction

Depth map from the depth camera have a large amount of lens radial distortion. There are two types lens distortion which are barrel distortion and pincushion distortion. In this case, the barrel distortion is occurred by the intrinsic problem of the depth camera. This distortion causes not only the shape mismatch between the color image and the corresponding depth image but also the errors in the results of some feature point based processing such as camera calibration.

In order to avoid that situation, we have to perform radial distortion correction to the obtained depth images. In general, there are two main categories of radial distortion correction. Methods in the first category use the point correspondences between two or more views. The second category also has lots of approaches which are based on the distorted straight line components in the image.

In the proposed fusion camera system, we use one of the second approaches to correct the radial distortion in the depth images. After finding the curved straight line component in the captured image, we estimate the distortion center and the distortion parameter. With the distortion information, we can reconstruct the image from the distorted image. Figure 4 shows the depth and intensity images before and after the correction.



Distorted Corrected

Fig. 4: Radial distortion correction.

3. MULTI-VIEW DEPTH MAP GENERATION

3.1 3D Warping of Depth Map

We generate initial depth of the multi-view image by performing 3D warping of the depth values obtained from the depth camera. First, we project pixels of the depth map into the 3D world coordinate using the depth values. We then reproject the 3D points into each view.

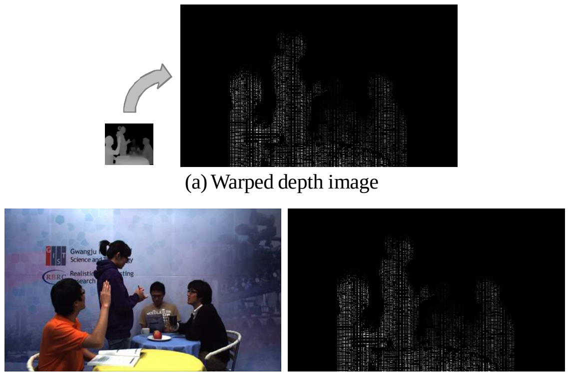
Let us assume that $D_s(p_{sx}, p_{sy})$ is the depth intensity at the pixel position (p_{sx}, p_{sy}) in the depth map. $P_s(x_{sx}, y_{sy}, z_{sz})$ is a 3D point corresponding to D_s. The backward projection for moving D_s to the world coordinate is carried out by

$$P_s = K_s^{-1} \cdot p_s \qquad (7)$$

where K_s^{-1} indicates the intrinsic matrix of the depth camera. In the backward 3D warping, since rotation and translation matrices of the depth camera are the identity matrix I and zero matrix O as Eq. 4, we have only to consider its intrinsic matrix. Thereafter, we project the 3D points P_s into the each view to get its corresponding pixel position $p_k'(u_k, v_k)$ of the k^{th}-view image by

$$p'_k = \tilde{P}'_k \cdot P_s \qquad (8)$$

where P_k' indicates the projection matrix of the k^{th}-view video camera. Figure 5 shows the result of 3D warping using the acquired depth maps.



(a) Warped depth image

(b) Matched to color image

Fig. 5: 3D warped depth map.

3.2 Region Separation

To estimate depth maps of multiple video cameras using the warped depth information, we segment the multi-view image by a mean-shift color segmentation algorithm. However, we cannot control the maximum segment size because there is no parameter to control the maximum segment size.

When we perform the segment-based stereo matching, one segment has one depth value. If the size of segment is too large, we cannot get a smooth depth map. The other way, if the size of segment is too small, it is hard to overcome textureless problem during the stereo matching. To solve this problem, we split one image into 16×16 block segments, so that we can limit the maximum segment size.

Figure 6 shows the procedure of the segment merging. A block can have two or more color segments. Before merging the segment, we split the segmented image into block-based segment again. If each segment is smaller than half size of the block, we merge it into one segment by searching adjoined blocks to find the same indexed segment. If the size of the merged block is larger than threshold, the merging procedure is finished; otherwise we repeat the same process until merging condition is satisfied.

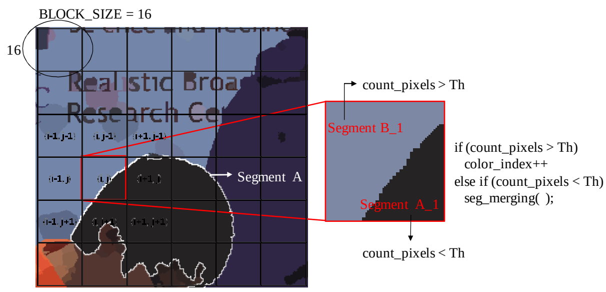


Fig. 6: Block-based segment merging.

The searching order of connected blocks is right, bottom, left, and top including the diagonal directions because left and top blocks are merged block and right and bottom block will be merged blocks. For example, *Segment A* divide into many block-based segments and *Block* (*i*, *j*) have two segments: *Segment A_1* and *Segment B_1*. Since the size of *Segment A_1* is smaller than the predefined threshold value in Fig. 8, the same indexed segment of *Segment A_1* is the blocks in (*i*+*1*, *j*), (*i*, *j*+*1*), and (*i*+*1*, *j*+*1*). We merge the current *Segment A_1* and the same indexed segment in (*i+1*, *j*) by the searching order.

3.3 Segment-based Depth Estimation

We define the initial depth of each segment as 3D warped depths in the segment; the assumption is that each segment has one depth value. However, there is one problem to set the initial depth using warped depth value. The 3D warping is performed from the small resolution depth map to the HD image in our system. Since there are many errors such as camera calibration error and depth error acquired from the depth camera, the warped result is not exactly matched with the HD image.

To obtain the accurate initial depth value, we use the warped results as multiple initial depth values for stereo matching. However, if the given initial depth is the error value, we could find wrong areas which has local minimum. Therefore, the assignment of the correct initial depth is crucial in using the depth camera. Because there are correct initial depths around the currently warped position, which are not exactly matched with the original image, we increase the candidates of the initial depth value to resolve this problem. By using the multiple initial depths, we can set initial depth for the depthless regions in the boundary of objects.

For determining the disparity of each segment, we calculate the mean of absolute difference (MAD) values between the segment in the current view image and its matched region in the left and right view images by

$$FG\_d_i(InitDisp) = \min(\min(\sum_{j=0}^{a} MAD(j)), \min(\sum_{k=0}^{b} MAD(k)) \qquad (9)$$

where *i* is the index of the segment, *j* and *k* means index of the multiple initial depth. *a* and *b* are the number of the initial depth in the horizontal and vertical regions, respectively. $FG\_d_i(InitDisp)$ is the refined initial depth value from pairwise stereo matching. Search range to estimate disparities of the current view image is from *InitDisp-5* to *InitDisp+5*. The disparity with the minimum MAD in the search range is chosen as the refined initial disparity of the segment in the current view image.

Since the acquired depth map is only for foreground regions, there is no depth information for background areas. In estimating depth of background, we set the minimum and maximum disparity value. We then find the minimum MAD as the initial disparity of the current segment in the background by

$$BG\_d_i(InitDisp) = \min(\sum_{i=\min Disp}^{\max Disp} MAD(i)) \qquad (10)$$

where $BG\_d_i(InitDisp)$ is the disparity for background, *minDisp* and *maxDisp* mean minimum and maximum disparity search range for background. The disparity with the minimum MAD is chosen as the initial disparity $d_i(Initdisp)$ of the segment *i* in the current view image *n* by

$$d_i(InitDisp) = \min(FG\_d_i(InitDisp), BG\_d_i(InitDisp)) \qquad (11)$$

3.4 Depth Refinement

In stereo matching, depth refinement usually enhances depth accuracy through iteration at the cost of long processing time, lots of memory requirement, and heavy computation. However, it has challenges when our target is to generate high-resolution 3D video based on multi-view depth maps. We therefore propose a simplified depth refinement approach using the proposed cost function for the depth map refinement, which has the following features: low memory consumption, fast processing time, and no iteration steps.

In order to enhance the multi-view depth map along the boundary of the objects, we refine it for two regions: moving region and static region. We have already defined the moving regions using color difference between frames as shown in Fig. 9. If there is no variance of a pixel in the time domain, we assume that pixel is static. In that case, we can refer the previous depth value for the static pixel. Otherwise, we just use the refined disparity value without referring the previous one.

$$E(x,y,d) = \begin{cases} w_s f_s(x,y,d_s(x,y)) + w_d f_d(x,y,d_d(x,y)) \;\; if \;\; obj\_mov(x,y) = 1 \\ w_s f_s(x,y,d_s(x,y)) + w_d f_d(x,y,d_d(x,y)) + w_t f_t(x,y,d_t(x,y)) \;\; if \;\; obj\_mov(x,y) = 0 \end{cases} \qquad (12)$$

where w_s, w_d, w_t are the weighting factors for depth refinement. $f_s(x, y, d_s(x,y))$ is the smoothness term with gradient of the refined depth value in this refinement step. $f_d(x, y, d_d(x, y))$ is the data term for the refined initial depth value in the segment-based stereo matching step and $f_t(x, y, d_t(x,y))$ is the temporal term for depth value of the previous frame for the static pixel. From our experimental, the weighting factors of the cost function w_s, w_d, w_t are 0.3, 0.5, and 0.2. $obj\_mov(x,y)$ indicates the result of the moving object detection. If $obj\_mov(x,y)$ is 0, this pixel is not moved. Then, we can refer the depth value of the previous frame.

$f_d(x, y, d_d(x, y))$ means the minimum MAD with the refined initial depth value in the search range from *InitDisp-5* to *InitDisp+5*. $f_s(x, y, d_s(x,y))$ is the depth difference with neighborhood depth in the same segment and calculated by

$$f_s(x,y,d_s(x,y)) = med(s_a(x,y), s_b(x,y), s_c(x,y)) \qquad (13)$$

We can calculate the smoothness value as shown in Fig. 13. $s_a(x,y)$ is the refined depth difference at positions between (*x -1*, *y-1*) and (x-1, y). $s_b(x, y)$ is the refined depth difference at positions between (*x-1*, *y-1*) and (*x*, *y-1*). $s_c(x, y)$ is the refined depth difference at positions between (*x*, *y-1*) and (*x+1*, *y-1*). The function *med*() takes the median value among arguments to avoid the wrong depth selection, so that it maintains depth continuity along the vertical and horizontal direction. If the selected smoothness gradient is a vertical direction, this depth difference is calculated from (*x*, *y-1*). Otherwise, the depth difference is computed from (*x-1*, *y*).

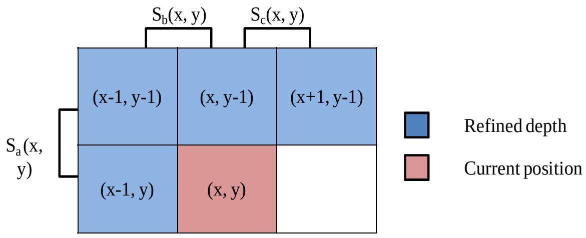


Fig. 7: Smoothness definition with gradient of the refined depth values.

4. EXPERIMENTAL RESULTS AND ANALYSIS

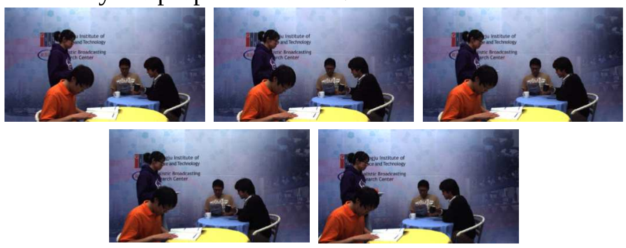
In order to generate the high-quality multi-view depth maps, we have constructed a hybrid camera system with five HD cameras and one depth camera. The measuring depth range of the depth camera is from 0.50m to 5.00m. The baseline distance among multi-view HD cameras are 6.5cm. Table 1 lists the specification of the hybrid camera system. Figure 8 shows the multi-view test sequences, Café, captured by the hybrid camera system. The resolution of the test multi-view images is 1920×1080, and that of the depth maps is 176×144.

Table 1: Specification of hybrid camera system.

Devices	Specifications	Details
Multi-view cameras (pcA1900-32gc)	Output format	NTSC or PAL (16:9 ratio, HD)
Depth camera (SR400)	Measured depth range	0.50m to 5.00m
	Field of View	43.6° (h) x 34.6° (v)
	Pixel Array Size	QCIF (176 (h) x 144 (v))
Sync. Generator (NI Trigger Box)	Output format	SD/HD Video Generation

Fig. 8: Test multi-view image and its depth map.

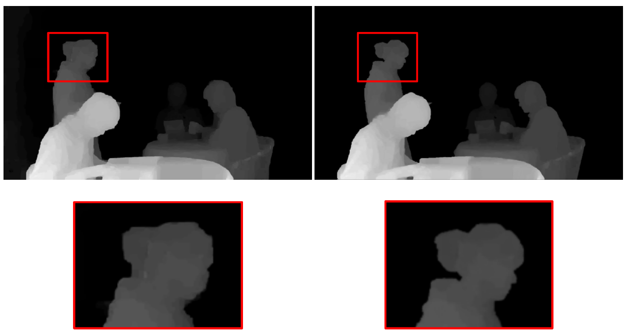
Figure 9 shows the final multi-view color images and their corresponding depth maps for the 1st frame of Café. To compare the depth quality of the proposed method with previous works, we have shown the disparity map generated by the DERS software for the 3rd view image of the 93rd frame of Café as shown in Fig. 10. We can observe that some regions of the depth maps generated by the previous method have noticeable errors in concave areas. Furthermore, the mismatched disparities in black hair were remarkably reduced by the proposed method.



(a) Multiview color sequence

(b) Multiview depth sequence

Fig. 9: Generated multi-view depth video.



(a) Depth map from DERS (b) Depth map from the proposed system

Fig. 10: Depth comparison with the previous work.

From Fig. 9 and Fig. 10, we notice that depths for the overlapped regions in foreground of *Café* were generated successfully, though the boundaries of the black hair were noisy. In addition, the yellow table expresses gradual depth difference despite the monotonous color of the table. As a result, we could overcome the two main problems of passive depth sensing efficiently, depth estimation on the occluded and textureless regions, using the depth camera data as the supplementary information.

To evaluate the subjective quality of the proposed method, we have synthesized views with the computed depth map. As shown in Fig. 11, the generated intermediate views using depth maps obtained by the proposed method are reasonable in the aspect of subjective quality. For objective comparison, we list the PSNR result of the synthesized view images using the previous method and the proposed one in Table 2.

(a) View 2 using depth map 1 and 3 (a) View 4 using depth map 3 and 5

Fig. 11: Synthesized images using generated depth maps.

Table 2: Average PSNR of synthesized images for CAMERA3

Sequence	Average PSNR	
	DERS	Proposed method
Café	34.89	35.18

Table 3 shows the comparison of the processing time in the depth refinement step. Since each algorithm have different processing step to generate the depth map, it is hard to measure the exact processing time in the same condition. Therefore, we compare the processing time for the depth map refinement step. As shown in Table 3, the proposed method is faster than others without the accuracy reduction for depth map generation. From the result, it is useful for the high-resolution multi-view depth map generation.

Table 3: Comparison of the processing time.

Sequence	Processing time (sec)	
	DERS	Proposed method
Café	836.26	337.21

5. CONCLUSIONS

In this paper, we have presented a new approach to generate depth maps corresponding to color images using the proposed hybrid camera system. We have used depth information acquired by a depth camera to generate the initial depth maps for stereo matching. We then have generated the final depth maps using segmentation-based stereo matching and the proposed cost functions. Experimental results have shown that our scheme produced more reliable depth maps and multi-view images compared with previous methods. With the proposed hybrid camera system, we could solve the two main problems in the current passive depth sensing, which is depth estimation on occluded and textureless regions. Therefore, our proposed system could be useful for various 3D multimedia applications and displays.

ACKNOWLEDGEMENTS

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2010-(C1090-1011-0003)).

REFERENCES

[1] Ministry of Science and Technology of Korea, National Technology Roadmap, 2003.

[2] K. Balasubramanian, “On the Realization of Constraint-free Stereo Television,” *IEEE Trans. on Consumer Electronics*, vol. 50, no. 3, pp. 895-902, 2004.

[3] C. Fehn, E. Barre, and S. Pastoor, “Interactive 3DTVConcepts and Key Technologies,” *Proceedings of the IEEE*, vol. 94, no. 3, pp. 524-538, 2006.

[4] J. Cha, S.M. Kim, S.Y. Kim, S. Kim, I. Oakley, J. Ryu, K.H. Lee, W. Woo, and Y.S. Ho, “Client System for Realistic Broadcasting: a First Prototype,” *Lecture Notes in Computer Science*, vol. 3768, pp. 176-186, 2005.

[5] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, “Efficient Prediction Structures for Multiview Video Coding,” *IEEE Trans. on Circuit and Systems for Video Technology*, vol. 17, no. 11, pp. 1461-1473, 2007.

[6] C. Fehn, “Depth-image-based Rendering (DIBR), Compression and Transmission for a New Approach on 3D TV,” *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 5291, pp. 93-104, 2004.

[7] K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, and Y. Yashima, “Multiview Video Coding Using New Interpolation and Color Correction,” *IEEE Trans. on Circuit and Systems for Video Technology*, vol. 17, no. 11, pp. 1436-1448, 2007.

[8] A. Ignatenko and A. Konushin, “A Framework for Depth Image-based Modeling and Rendering,” *Proc. of Graphicon*, pp. 169-172, 2003.

[9] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, 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.

[10] S.U. Yoon and Y.S. Ho, “Multiple Color and Depth Video Coding Using a Hierarchical Representation,” *IEEE Trans. on Circuit and Systems for Video Technology*, vol. 17, no. 11, pp. 1450-1460, 2007.

[11] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-frame,” *International Journal Computer Vision*, vol. 3768, pp. 164-175, 2005.

[12] C. Zitnick and T. Kanade, “A Cooperative Algorithm for Stereo Matching and Occlusion Detection,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, pp.675-684, 2000.

[13] E. Lee , Y. Kang, Y. Jung and Y. Ho, “Three-dimensional Video Generation using Foreground Separation and Disocclusion Detection,” 3DTV Conference 2010, pp. 69(1-4), 2010.