3DTV-CON 2010

THE TRUE VISION

Capture, Transmission and Display of 3D Video

The 3DTV Conference 2010 will be held on June 7-9 2010, in hotel Scandic Rosendahl, Tampere, Finland. Tampere has been named the most desirable place to work, live and study in Finland and hotel Rosendahl offers a peaceful experience of Finnish forests and lakes yet walking distance away from the effervescent city center.

The conference is organized by the Department of Signal Processing, Tampere University of Technology and Tampere International Center for Signal Processing (TICSP) and sponsored by the IEEE SP/CAS Chapter of Finland and Nokia.

This is the fourth edition of the conference, held for the first time in Kos Island, Greece, and then held in Istanbul, Turkey, and in Potsdam, Germany. The 3DTV-CON series of conferences originated from the FP6 3DTV Network of Excellence and has gotten continuing support from FP7 projects working in the area of 3D Media.

Preliminary call for papers is available here: 3dtv-con2010-cfp.pdf

Program

News:

- **18.05.2010** - Instructions for presenters are now available in the Author's kit section.
- **18.05.2010** - Conference program has been updated.
- **12.04.2010** - The accepted papers have now been announced.
- **30.03.2010** - The decision of paper acceptance will be announced by April 9, 2010. Correspondingly, camera ready submission dealine has beed extended to April 30, 2010.

Important dates:

Tutorial proposals deadline	December 18, 2009
Special session proposal deadline	January 29, 2010
Regular paper submission deadline	Extended to **February 26, 2010**
Notification of paper acceptance	Extended to **April 9, 2010**
Camera-ready paper submission deadline	Extended to **April 30, 2010**
Conference	**June 7-9, 2010**

Sponsors:

TAMPEREEN TEKNILLINEN YLIOPISTO

TICSP Tampere International Center for Signal Processing

IEEE

NOKIA Connecting People

3D MEDIA CLUSTER

Partners:

3D MEDIA

Veritas et Visus

Related events:

EuroITV 2010

CGIV 2010

3D MEDIA SUMMER SCHOOL

3DTV-Conference 2010

OpenConf Peer Review & Conference Management System

OpenConf Home | Email Chair

Full Program

Monday, 7 June 2010

8:30-17:00 — Registration

9:30-10:30 — Tutorial I: Vision Field Capturing
Chair: Karsten Müller
Qionghai Dai, Yebin Liu, Xun Cao
Department of Automation, Tsinghua University, Beijing, China

10:30-10:45 — Coffee break

10:45-12:00 — Tutorial I: Vision Field Capturing
Chair: Karsten Müller
Qionghai Dai, Yebin Liu, Xun Cao
Department of Automation, Tsinghua University, Beijing, China

12:00-13:00 — Lunch

13:00-13:10 — Welcome speech by Prof. Ulla Ruotsalainen, Dean of the Faculty of Computing and Electrical Engineering, Tampere University of Technology
Chair: Atanas Gotchev

13:10-14:10 — Keynote: Recent R&D Activities on 3DTV Broadcasting in Korea
Chair: Atanas Gotchev
Jinwoong Kim
Electronics and Telecommunications Research Institute (ETRI), Korea

14:10-14:30 — Coffee break

14:30-15:50

3D Capturing Technology
Chair: Aydin Alatan

Calibration Of A Synchronized Multi-camera Setup For 3d Videoconferencing
Wolfgang Waizenegger; Ingo Feldmann

3-d Video Generation Using Foreground Separation And Disocclusion Detection
Eun-Kyung Lee; Young-Ki Jung; Yo-Sung Ho

3d Free-viewpoint Video Caprturing Interface By Using Bimanual Operation
Tetsuya Watanabe; Itaru Kitahara; Yoshinari Kameda; Yuichi Ohta

Dynamic Adaptation Of Multi View Camera Structure
Mohamed ALI-BEY; Noureddine MANAMANNI; Saïd MOUGHAMIR

Special Session: Real-time high definition depth estimation
Chair: Oliver Schreer

3d-tv Ldv Content Generation With A Hybrid Tof-multicamera Rig
Anatol Frick; Bogumil Bartczak; Reinhard Koch

Adaptive Cross-trilateral Depth Map Filtering
Marcus Mueller; Frederik Zilly; Peter Kauff

Reliability-aware Cross Multilateral Filtering For Robust Disparity Map Refinement
Jörn Jachalsky; Markus Schlosser; Dirk Gandolph

Real-time Depth Estimation For Immersive 3d Videoconferencing
Ingo Feldmann; Wolfgang Waizenegger; Nicole Atzpadin; Oliver Schreer

15:50-16:10 — Coffee break

16:10-17:30

3D Display Technology
Chair: Mårten Sjöström

Multi-user Glasses Free 3d Display Using An Optical Array
RAJWINDER SINGH BRAR; PHIL SURMAN; IAN SEXTON; RICHARD BATES; FRANK NEUMANN

Development Of An Autostereoscopic Display System Using Projectors Array
Wu-Li Chen; Chang-Shuo Wu; Chang-Ying Chen; Shu-Chuan Cheng; Chao-Hsu Tsai

Analysis For Reproduced Light Field Of 3d Displays
Takafumi Koike; Kei Utsugi; Michio Oikawa

Multi-slm Holographic Display System With Planar Configuration
Fahri Yaraş; Hoonjong Kang; Levent Onural

Special Session: Enabling Technologies for Depth-Enhanced 3D Video
Chair: Karsten Müller

Temporally Consistent Stereo Matching Using Coherence Function
Dongbo Min; Sehoon Yea; Anthony Vetro

View Consistent Multiview Depth Estimation For Three-dimensional Video Generation
Sang-Beom Lee; Yo-Sung Ho

Error Supression In View Synthesis Using Reliability Reasoning For Ftv
Lu Yang; Tomohiro Yendo; Mehrdad Panahpour Tehrani; Toshiaki Fujii; Masayuki Tanimoto

Depth-based Inter-view Prediction Of Motion Vectors For Improved Multiview Video Coding
Jacek Konieczny; Marek Domański

18:30-21:00 — Reception

THREE-DIMENSIONAL VIDEO GENERATION USING FOREGROUND SEPARATION AND DISOCCLUSION DETECTION

Eun-Kyung Lee[1]*, Yun-Suk Kang*[1]*, Young-Kee Jung*[2] *and Yo-Sung Ho*[1]

Gwangju Institute of Science and Technology (GIST)[1], Honam University[2]
E-mail: {eklee78, yunsuk, hoyo}@gist.ac.kr, ykjung@honam.ac.kr

ABSTRACT

In this paper, we present a multi-depth camera system for 3-D video contents generation. We combine five video cameras and three TOF depth cameras to capture a 3-D scene in real time. By taking advantages of both active and passive sensor based depth acquisition methods, we can obtain accurate multi-view depth maps. After several preprocessing operations, each depth map is warped to the video camera position and used as initial depth values for refinement. To reduce mismatched depth values along object boundaries, we separate the foreground and background regions using moving objects detection and extract occlusion and disocclusion areas with the initial depth information. By applying the stereo matching method, we can improve multi-view depth maps. Intermediate view image sequences are then synthesized using both the multi-view color and depth information.

Index Terms **—** depth map generation, depth camera, depth estimation, multi-view camera

1. INTRODUCTION

As the three-dimensional (3-D) video becomes attractive in various 3-D multimedia applications, it is essential to obtain multi-view video sequences with corresponding depth maps, which are often called as multi-view video-plus-depth data [1]. In near future, consumers will be able to not only experience 3-D depth impression, but also choose their own viewpoints in the immersive 3-D visual scene created from the multi-view video. Recently, ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) has recognized the importance of the multi-view video-plus-depth data for 3DTV, and has investigated the needs for standardization of 3-D video coding [2].

In order to reproduce a 3-D scene at an arbitrary viewpoint from the multi-view video sequences, we need to obtain accurate depth information. There are several methods to obtain depth information of the 3-D scene. Generally, they can be divided into two major categories. One is based on passive methods and the other is based on active methods. Among the passive methods, stereo matching is the most popular one which is widely used. Although it is widely used in the fields, some difficult problems, such as occlusion, have been remained [3].

One of the most popular active methods is the time-of-flight (TOF) range sensor to obtain the depth information of the scene in real time. TOF depth acquisition is based on measuring the arriving time of the emitted infrared signal from the sensor. However, TOF has several problems to overcome such as low spatial resolution and noisy acquisition depending on the capturing environment [4].

In recent years, fusion methods have been proposed to take advantages of both passive and active methods by combining multiple video cameras with a depth camera [5-7]. The TOF depth camera produces an initial depth map by integrating a high-speed pulsed IR light source with a conventional broadcast TV camera. The hybrid camera system can enhance the initial depth map measured by the depth camera by applying a disparity estimation method on color images captured by multiple cameras. *Zhu* presented a calibration method to improve the depth quality using a TOF depth sensor [5]. He used the probability distribution function of the depth information measured by the TOF depth sensor and produced a more reliable depth map. However, his method produced only low-resolution depth maps and focused on generating depth maps for static 3-D scenes.

To solve the low resolution problem, hybrid methods have been proposed to obtain high-resolution and high-quality depth maps [6-7]. They used the initial depth information to avoid the mismatch problem. However, there exist mismatched regions along object boundaries. To reduce mismatched depth value in the boundary of moving objects, we propose foreground and background separation and occlusion and disocclusion extraction.

Since future 3-D applications are expected to use high-quality and high resolution 3-D videos, we need to create such multi-view video-plus-depth data in high quality. In this paper, we introduce a multi-depth camera system that has multiple video cameras and multiple depth cameras. They are arranged in two rows with parallel type. The captured color and depth image sequences with 30fps are preprocessed to increase the inter-view correlation and the computational accuracy. The preprocessed depth image sequences are then warped to the corresponding video cameras and used as initial depth values for stereo matching. After depth estimation, we obtain the 3-D video that is composed of the multi-view color image sequences and the corresponding multi-view depth sequences.

2. HYBARID 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. Figure 1 shows the proposed hybrid camera system and Table 1 is the specification of the hybrid camera system.

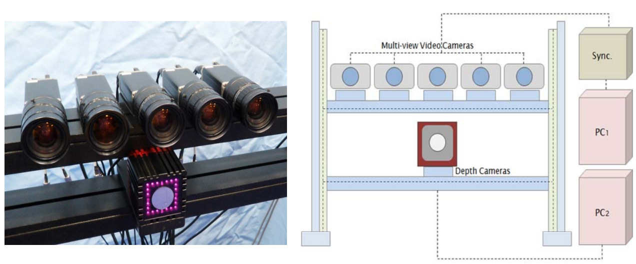


Figure 1. Hybrid camera system

978-1-4244-6378-7/10/$26.00 ©2010 IEEE

Table 1. Specification of hybrid camera system

Devices	Specifications	Details
Multi-view cameras (pcA1900-32gc)	Output format	NTSC or PAL (16:9 ratio)
	Pixel array size	HD (1920 (h) x 1080 (v))
Depth camera (SR400)	Measured 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

In order to capture multi-view video sequences simultaneously, we connect each video camera to a PC for storing through a synchronizer CA-1000 by National Instrument. For synchronized capturing of a depth map, we modify the software development kit (SDK) provided by *Swiss Ranger*. However, there is no instrument to synchronize two different types of cameras. Thus, video and depth capturing is controlled manually.

3. PREPROCESSING FOR 3-D VIDEO GENERATION

3.1 Relative Camera Calibration

Since the proposed hybrid camera system consists of two different types of cameras, a depth camera and multiple HD video cameras, it is essential to find out relative camera information through camera calibration. In our hybrid camera system, we apply a camera calibration algorithm to each camera and obtain projection matrices for the depth camera and each video camera independently.

$$P_d = K_d[R_d \mid t_d] \qquad (1)$$
$$P_k = K_k[R_k \mid t_k]$$

where P_d is the projection matrix of the depth camera represented by its intrinsic matrix K_d, rotation matrix R_d, and translation vector t_d. P_k indicates the projection matrix of the k^{th} video camera which consisted of its intrinsic matrix K_k, rotation matrices R_k, and translation vector t_k.

We then apply a multi-view rectification operation. The multiple cameras have geometric errors because they are set manually. In order to minimize the geometric errors, we find the common baseline, and then apply the rectifying transformation to the multi-view image. Consequently, the projection matrix of each video camera is changed as

$$\tilde{P}_k = K_k'[R_k' \mid t_k] \qquad (2)$$

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_d of the depth camera into the identity matrix I by multiplying the inverse rotation matrix R_d^{-1}. The translation vector t_d of the depth camera is also changed into the zero matrix O by subtracting the translation vector t_d. Hence, we can define new relative projection matrices for the multi-view cameras on the basis of the depth camera as

$$P_d' = K_d[I \mid O] \qquad (3)$$
$$\tilde{P}_k' = K_k'[R_k' R_d^{-1} \mid t_k - t_d]$$

where P_d' and $\tilde{P}_k'$ are final projection matrix of the depth camera and the k^{th} video camera, respectively.

After relative camera calibration, we solve the color mismatch problem of multi-view images using a color calibration method. The color characteristics of captured images are usually inconsistent due to different camera properties and lighting conditions even if the hardware type and specification of the multiple cameras are the same.

3.2 Depth Calibration

Depth information captured by the depth camera is very sensitive to color and motion. Even though the distance from the depth camera to the object is constant, depth information from the depth camera is different depending on the environments. Basically, the *SR4000* depth camera has its own depth calibration tool. However, it is very poorly calibrated.

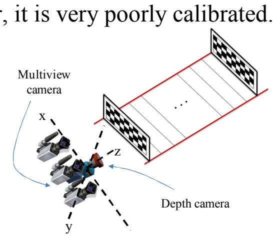


Figure 2. Depth calibration

To calibrate the measured depth value with its real depth, we find a mapping curve to compensate the difference between them. In this paper, we use the *Zhu*'s algorithm to analyze the characteristics of the depth camera. When objects move in the limited indoor environment, we check the depth of the planar image pattern within the limited space by increasing the distance from the image pattern to the depth camera as shown in Fig. 2. Since we already know the camera parameters of each camera, the real depth value is calculated by

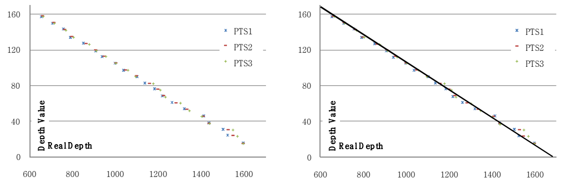
$$d_k(p_x, p_y) = \frac{K \cdot B}{D_k(p_x, p_y)} \qquad (4)$$

where K is the forcal length of the video camera, B is the baseline distance between neighboring two video cameras. $d_k(p_x, p_y)$ is the real depth value corresponding to the meaured depth value $D_k(p_x, p_y)$ at pixel position (p_x, p_y) in the depth map of image pattern.

Thereafter, we generate a mapping curve between real depths and measured depths from the depth camera. In order to generate the mapping curve, we find out the fitting curve using the cubic equation as

$$y = a + bx + cx^2 + dx^3 \qquad (5)$$

The cross small rectangular points on the *x*-*y* plane in Fig. 3 are formed by the measured depths x and real depths y that minimizes the sum of squared distances to these points. Figure 3 shows calibrated depth values using the mapping curve.



(a) Relationship between the depth value and real depth

(b) Function for depth value correction

Figure 3. Mapping curve for depth calibration

3.3 Moving Object Separation

To estimate depth maps of multiple video cameras using the warped depth, we segment the multi-view image by a mean-shift color segmentation algorithm. However, we cannot control the maximum segment size using the mean-shift algorithm. 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 9 shows the procedure of the segment merging. A block is able to 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.

The searching order of connected blocks is right, bottom, left, and top including the diagonal directions because left and top blocks are already merged block and right and bottom block will be merged blocks. For example, Segment A is divided 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. 4, 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)*.

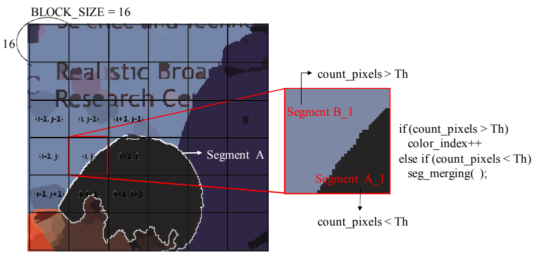


Figure 4. Block-based segment merging

Before we estimate depth maps, we separate foreground and background regions using moving object detection and the generated depth map in the previous frame. To extract the moving object in the current frame, we calculate color differences between the previous frame *n-1* and the current frame *n* by using the threshold which indicates the current position is foreground or not. We cannot directly use the segment-based moving object detection because shape of each segment can be varied in the temporal domain as shown in Fig. 5.


Figure 5. Segmentation results in the temporal domain

Since color segmentation is performed frame by frame, it is hard to find the same segment in the temporal domain. Therefore, we use the Euclidean distance between frames to extract the moving objects as

$$E(x,y)=\sqrt{(R_{n-1}(x,y)-R_n(x,y))^2+(G_{n-1}(x,y)-G_n(x,y))^2+(B_{n-1}(x,y)-B_n(x,y))^2} \quad (6)$$

where R, G, and B indicate the pixel values in RGB color domain. To find the moving object, we compute the *E(x, y)* at each pixel location for all pixels. If we subtract the RGB value between frames, camera noises can be mixed up. To remove them, we calculate the average RGB value for 3×3 block. If the average is larger than the threshold value, we set the center pixel of each 3×3 block as the foreground pixel. From our experiments, the threshold value of Euclidean distance, 10 is used. Figure 6 (a) and Figure 6 (b) present 78th frame and 79th frame images in *Camera 3* and Fig. 6(c) shows the result of moving objects.

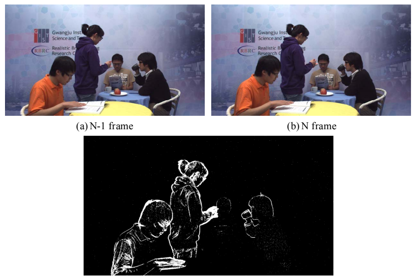

(a) N-1 frame (b) N frame (c) Extracted moving objects

Figure 6. Moving object detection using color difference between frames

4. DEPTH VIDEO GENERATION

To increase the depth accuracy for stereo matching, we utilize pairwise stereo matching method. When the current view is Camera 1, there is no left image. Therefore, we use the input images of *Camera 2* and *Camera 3* for current view. For input image of *Camera 3*, we use the value of the initial depth multiply by 2.

When we perform the stereo matching operation twice with the left and right images for one depth, we can find the occlusion and disocclusion regions. If the some regions are not existed in one view even though same regions are exist in the other view, this region is occluded and disoccluded regions. By using this theory, we can construct the reliable and unreliable regions. After getting corresponding value with the initial depth information in small search range, we also judge the occlusion and disocclusion regions using the calculated depth value.

In practice, since the depth quality of the initial disparity map generated by the depth estimation method is usually low, it is hard to use it as the multi-view disparity map for the multi-view image. In order to enhance the initial disparity map, we refine it according to regions: moving region and static region. We already define the moving regions using color difference between frames as shown in Fig. 6. If the there is no variance in the time domain, that is the static objects. Therefore we can refer the previous depth value. However there are some variances, we just use the refined disparity value pixel-based refinement by

$$E(x,y,d)=\begin{cases} w_s f_s(x,y,d_s(x,y)) \\ +w_d f_d(x,y,d_d(x,y)) \ \ if \ \ obj\_mov(x,y)=1 \\ w_s f_s(x,y,d_s(x,y))+w_d f_d(x,y,d_d(x,y)) \\ +w_t f_t(x,y,d_t(x,y)) \ \ if \ \ obj\_mov(x,y)=0 \end{cases} \quad (7)$$

where w_s, w_d, w_t are the weighting factors for depth refinement. $f_s(x, y, d_s(x,y))$ is the smooth term with gradient of the refined depth value in this refinement step and $f_d(x, y, d_d(x, y))$ means the refined initial depth value in the segment-based stereo matching step, and $f_t(x, y, d_t(x,y))$ is the depth value of the previous frame for the static objects. *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.

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

$$f_s(x,y,d_s(x,y))=med(d(x-1,y),d(x,y-1),d(x+1,y-1)) \quad (8)$$

where $d(x-1, y)$, $d(x, y-1)$, $d(x+1, y-1)$ are the refined depth value in this step.

5. EXPERIMENTAL RESULTS AND ANALYSIS

In order to evaluate the performance of the proposed method, we have constructed a hybrid camera system with five HD cameras and three depth cameras. The measuring distance for depth information of the depth camera was from 0.5m to 5.0m. The baseline distances among multi-view HD cameras were 6.5cm.

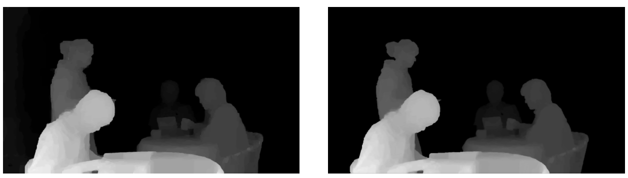
Figure 7 shows the final multi-view depth maps for the 1st frame of *Cafe*. As shown in Fig. 7, the depth quality of the yellow table was good, although the color of the table was monotonous. As a result, we could overcome the two main problems of passive depth sensing, depth estimation on the occluded and textureless regions, using the depth camera data as a supplement.

(a) Multi-view color sequence

(b) Multi-view depth sequence

Figure 7. Generated multi-view depth maps

To compare the depth quality of the proposed method with previous works, we have compared the disparity map generated by the DERS software [8] with the 3rd view image of the 1st frame in *Cafe*. The generated depth maps using DERS software and the proposed method are shown in Fig. 8(a) and Fig. 8(b). We could notice that some regions of the depth maps generated by the previous methods had mismeasured depths. On the other hand, the mismatched disparities were notably reduced by the proposed method. As a result, the proposed method outperformed the previous methods.



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

Figure. 8 Results comparison with DERS software

In order to measure the performance of our scheme objectively, we generated the synthesized images using the depth maps generated for *Camera 2* and *Camera 4*. As shown in Table 2, we compared the average peak signal-to-noise ratio (PSNR) of the synthesized images generated by DERS software and proposed method. The synthesized images produced by the proposed camera system were more accurate than the previous methods.

We have also generated intermediate views using the disparity maps and images of the *Camera 2* and *Camera 4* of the 1st frame in *Cafe*. In order to construct virtual views with the estimated multi-view video-plus-depth, we employed the VERS software. As shown in Fig. 9, we could generate intermediate views successfully and provide natural 3-D video service to potential consumers.

Table 2. PSNR comparison

	VERS(dB)	Proposed method
Average PSNR	34.89	35.18

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

Figure 9. Synthesized images

6. CONCLUSIONS

In this paper, we have presented a new approach to generate multi-view HD depth maps corresponding to HD color videos using a hybrid camera system. We have used the depth information acquired by a depth camera to generate the initial depth map through the 3-D warping operation and generated the final depth map using a segmentation-based stereo matching algorithm and a cost function. Experimental results have shown that our scheme produced more reliable depth maps compared to previous methods. With the proposed hybrid camera system, we could solve the two main problems in the current passive depth sensing and depth estimation in occluded and textureless regions.

7. ACKNOWLEDGEMENTS

This research was supported in part by MKE, Korea, under the ITRC support program supervised by NIPA (NIPA-2010-C1090-0902-0017) and in part by the KRF Grant (MOEHRD, Basic Research Promotion Fund) (R01-2007-000-20330-0).

8. REFERENCES

[1] C. Fehn, R. Barre, and S. Pastoor, "Interactive 3-DTV concepts and key technologies," Processing Image Communication, vol. 94, no. 3, pp. 524-538, March 2006.

[2] ISO/IEC JTC1/SC29/WG11 N8944, "Preliminary FTV model and requirements," April 2007.

[3] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Trans. on Graphics, pp. 600-608, August 2004.

[4] C.S. Swiss Ranger SR-2, Swiss Center for Electronics and Microtechnology. http://www.csem.ch/fs/imaging.htm.

[5] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of time-of-flight depth and stereo for high accuracy depth maps," IEEE Conference on Computer Vision and Pattern Recognition, pp. 231-236, June 2008.

[6] E. Lee, Y. Kang, Y. Jung, and Y. Ho, "3-D video generation using hybrid camera system," International Conf. on Immersive Telecommunications, pp. T5(1-6), May 2009.

[7] Y. Kang, E. Lee, and Y. Ho, "Multi-depth camera system for 3D video generation," International Workshop on Advanced Image Technology, pp. 44(1-6), January 2010.

[8] ISO/IEC JTC1/SC29/WG11 M15377, "Reference Softwares for Depth Estimation and View Synthesis," April 2008.