The 2008 International Conference on Embedded Systems and Intelligent Technology
February 27-29, 2008, Grand Mercure Fortune Hotel, Bangkok, Thailand

HYBRID CAMERA SYSTEM FOR THREE-DIMENSIONAL REALISTIC BROADCASTING SERVICES

Yo-Sung HO, Sung-Yeol KIM, and Eun-Kyung LEE

Gwangju Institute of Science and Technology (GIST)
E-mail: {hoyo, sykim75, eklee78}@gist.ac.kr

In this paper, we present a new system to generate multi-view video sequences with depth information (MVD) by integrating multiple high-definition (HD) camera arrays and one standard-definition (SD) depth camera. In the proposed hybrid camera system, we first create the initial disparity for each HD color image by applying a three-dimensional (3-D) warping operation on the depth map acquired by the depth camera. Then, the final disparity for each HD image is obtained by a stereo matching algorithm with the initial disparity. Finally, we generate the HD depth map for each HD image by converting the final disparity into the depth information. Experimental results show that the depth map generated by the hybrid camera system provides reliable depth information for 3-D realistic broadcasting services. Besides, the proposed system minimizes the inherent problems of conventional depth cameras, such as limitation of measuring distance for depth information and generation of low-resolution depth maps.

HYBRID CAMERA SYSTEM FOR THREE-DIMENSIONAL REALISTIC BROADCASTING SERVICES

Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
Email: hoyo@gist.ac.kr

ABSTRACT

In this paper, we present a new system to generate multi-view video sequences with depth information (MVD) by integrating multiple high-definition (HD) camera arrays and one standard-definition (SD) depth camera. In the proposed hybrid camera system, we first create the initial disparity for each HD color image by applying a three-dimensional (3-D) warping operation on the depth map acquired by the depth camera. Then, the final disparity for each HD image is obtained by a stereo matching algorithm with the initial disparity. Finally, we generate the HD depth map for each HD image by converting the final disparity into the depth information. Experimental results show that the depth map generated by the hybrid camera system provides reliable depth information for 3-D realistic broadcasting services. Besides, the proposed system minimizes the inherent problems of conventional depth cameras, such as limitation of measuring distance for depth information and generation of low-resolution depth maps.

1. INTRODUCTION

As immersive multimedia services are expected to be available anytime and anywhere through high-speed optical networks in the near future, the importance of three-dimensional (3-D) video is increasingly recognized as the essential part of high-quality visual media. Recently, the ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) has been interested in the multi-view video with depth information (MVD) related to free-viewpoint TV (FTV). In addition, 3-D TV is widely accepted as the next generation television broadcasting service, because it can present more natural viewing experience in the three dimensions [1].

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

We can obtain a depth map from the natural scenes in real time using active range depth cameras, such as the Z-Cam developed by 3DV Systems, Ltd. [2] or the NHK Axi-vision HDTV camera [3]. These kinds of depth cameras can capture color images and their associated per-pixel depth map simultaneously by integrating a high-speed pulsed infrared (IR) light source with a typical broadcast TV camera.

Few years ago, the European ATTEST project showed a possibility of realizing a 3-D TV system using a depth camera [4]. Recently, Realistic Broadcasting Research Center (RBRC) in Korea has demonstrated full 3-D contents for future home-shopping channel broadcasting using a depth camera [5]. However, although the depth cameras can produce useful depth information in real time, there are some in-built problems in the currently available depth camera systems.

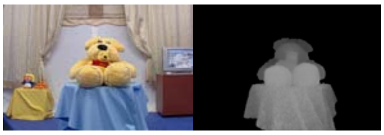
1) The depth map generated by Z-Cam usually includes optical noises. Optical noises occur due to differences in the reflectivity of IR sensors from color variations of objects in the scene.
2) The depth camera has limitations in measuring the distance of objects. In practice, the range of the depth distance by Z-Cam is from 1m to 5m.
3) The current depth camera can generate only low-resolution depth maps. Typically, the resolution of the depth map acquired by Z-Cam is 720×486.

Recently, a depth map fusion method that combines multi-view cameras and a depth camera has been introduced [6]. This hybrid camera system generates enhanced depth maps by applying a stereo matching algorithm to multi-view images with depth information captured by the depth camera. However, this system cannot produce high-resolution depth maps, because it completely depends on the low-resolution depth camera.

With the development of camera sensor technologies, we can upgrade the current depth camera and obtain high-resolution depth maps for more natural and realistic 3-D multimedia applications. However, due to many challenges in real-time distance measuring systems, upgrading and improvements of depth cameras are very slow and expensive.

In this paper, we propose a new scheme to generate high-resolution depth maps using multiple HD cameras

(a) Multi-view images



(b) Foreground

(c) Background

Figure 1. Images from the hybrid camera system

and the standard-definition (SD) Z-Cam. The proposed scheme generates multi-view HD depth maps by refining the depth map initially acquired by the SD depth camera.

Our main contribution is that we have developed a production system for high-resolution depth maps using the features of passive depth sensing and depth camera technology. In addition, the proposed scheme provides a practical solution to the inherent problems of depth maps captured by the currently available depth camera system.

2. HYBRID CAMERA SYSTEM

2.1. Construction of Hybrid Camera System

The proposed hybrid camera system consists of five HD cameras and a depth camera. All cameras are connected to a personal computer equipped with a video capturing board. In addition, a clock generator is linked to the camera set to provide synchronization signals constantly.

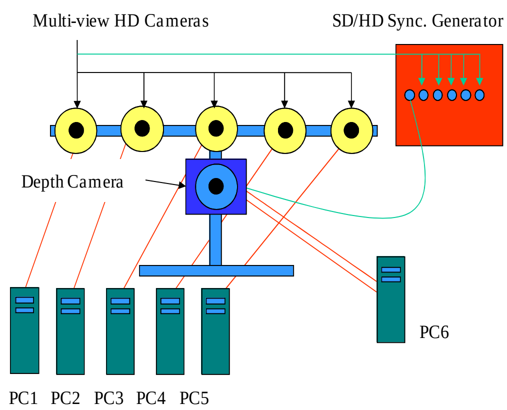


Figure 2. Hybrid camera system

With the hybrid camera system, we capture seven synchronized images: five HD images from the multi-view cameras, one SD color image and its corresponding depth map from the depth camera. Since the measureable distance for the depth map using Z-Cam is very limited in the practical environments, we capture a color image and its depth map for background in advance. Figure 1 shows the nine images captured by the hybrid camera system, and Figure 2 shows its main components.

2.2. Overall Framework

Figure 3 describes the overall framework of the proposed multi-view depth map generation system. Before we capture synchronized multi-view images with the hybrid camera system, we calibrate each camera independently and calculate their relative camera information.

At the side of the depth camera, we first calculate the camera parameters using a camera calibration method. Since the depth camera has limitations in the depth range, we obtain two depth maps for foreground and background separately. After applying a 3-D warping method on the depth maps for foreground objects and background and projecting the depth information into the world coordinate, we reproject the warped depth information into the positions of multi-view cameras and generate their initial depth map. Finally, the pixel intensity values of the initial depth maps are used as the initial disparity for each view during stereo matching.

At the side of multi-view cameras, each HD image is rectified and color-segmented. The initial disparities, which are generated by a depth camera, are assigned into the corresponding segments. In addition, in order to detect occluded regions, we segment each image into three different categories: background, foreground, and unknown regions. The disparity of each segment is independently estimated by a stereo matching technique.

3. MULTI-VIEW DEPTH MAP GENERATION

3.1. Preprocessing

Since we have to combine two different types of images with different intrinsic and extrinsic camera parameters, we need to obtain relative camera information for each camera using a camera calibration method. For the camera calibration, we employ Tsai's camera calibration method [7]. After the camera calibration is completed, five HD images are rectified and color-segmented.

For multi-view image rectification, we find a common baseline that minimizes the sum of squared distances to the camera centers, and transform multi-view images into virtual camera images that are located on the same baseline. Then, each rectified image is color-segmented.

The 2008 International Conference on Embedded Systems and Intelligent Technology
February 27-29, 2008, Grand Mercure Fortune Hotel, Bangkok, Thailand

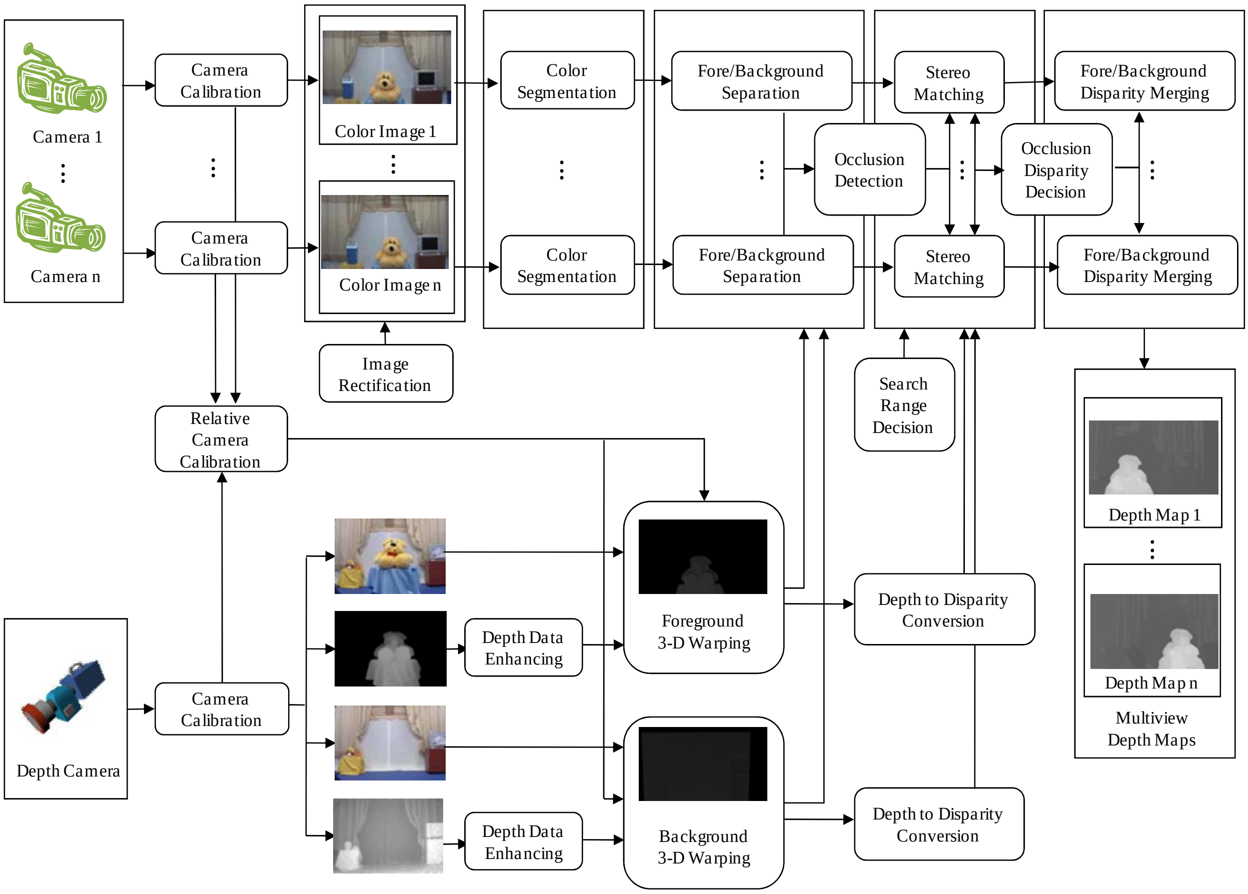


Figure 3. Overall framework of the generation of multi-view depth maps

In order to reduce optical noises in the depth map, we employ both down-sampling and linear interpolation operations. After we apply mean filtering on the depth map, the depth data enhancement algorithm executes down-sampling on the mean-filtered depth map. Then, we perform bilateral filtering on the down-sampled depth map. Finally, we recover the depth map using a simple linear interpolation.

3.2. Generation of Initial Disparity

In order to generate the initial disparity map, we apply a 3-D warping algorithm on the color and depth map from the depth camera. In the 3-D warping operation, we move the color and depth data to the world coordinate, and then project 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}) of the depth map, we can regard the pixel p_s $(p_{sx}, p_{sy}, D_s(p_{sx}, p_{sy}))$ as a point in the three-dimensional space. The corresponding point p_n of the n^{th} HD image is calculated by

$$p_n = \tilde{P}_n' \cdot P_s^{-1} \cdot p_s \qquad (1)$$

where p_n $(p_{nx}, p_{ny}, 1)$ includes the corresponding pixel position (p_{nx}, p_{ny}) of the pixel p_s in the n^{th} HD image.

In addition, the depth information $D_n(p_{nx}, p_{ny})$ of p_n is obtained by

$$D_n(p_{nx}, p_{ny}) = (\tilde{t}_{n_z} - t_{x_z}) + D_s(p_{sx}, p_{sy}) \qquad (2)$$

where $\tilde{t}_{nz}$ and t_{sz} indicate the third value of the transition matrix of the n^{th} camera $\tilde{t}_n$ and the transition matrix of the depth camera t_s, respectively.

3.3. Stereo Matching

For the stereo matching operation, we first calculate the average disparity value in each segment. Then, in order to refine its disparity more accurately, we examine the small neighboring area around the initial disparity. Since we have separate initial depth maps for foreground objects and background, we perform the stereo matching operation on each segmented region independently.

In general, the quality of the depth map is not good enough in the boundary of foreground objects due to occluded regions, as shown in Fig. 4(a). In order to solve this inherent problem, we extract occlusion regions from the multi-view images by merging neighboring segments between the foreground objects and background.

Figure 4(b) shows a segmented image of foreground, background, and occluded regions. In order to correct disparities of the occluded regions, we calculate the sum of squared difference (SSD) with the initial disparity of the segment in the foreground, and recalculate SSD with the initial disparity of the segment in the background.

(a) View 1 (b) View 2 (c) View 3 (d) View 4 (e) View 5

Figure 6. Results of multi-view depth maps

We regard one of two disparities as the disparity of the segment in the occlusion region by comparing their SSD values and choosing the smaller one. Figure 4(c) shows the refined disparity map after solving the occlusion problem using a region separation.

(a) Occlusion problem

(b) Region separation

(c) Refined disparity map

Figure 4. Solving occlusion problems

4. EXPERIMENTAL RESULTS

In order to evaluate the proposed multi-view depth map generation, we constructed a hybrid camera system with five HD cameras and one Z-Cam as the depth camera. Figure 5 shows the constructed hybrid camera system. Figure 6 show the results of multi-view depth map generation using the proposed hybrid camera system.

Figure 5. Construction of hybrid camera system

5. CONCLUSIONS

In this paper, we proposed a new scheme that generates multi-view color images and their corresponding depth maps using a hybrid camera system. Experimental results showed that we could generate high-resolution multi-view depth maps in real time. Since we refined the depth map initially acquired by a depth camera using several image processing techniques, the final depth data is quite reliable. The proposed hybrid camera system can be exploited for various 3-D multimedia applications.

ACKNOWLEDGEMENTS

This work was supported in part by ITRC through RBRC at GIST, and in part by MOE through the BK21 project.

REFERENCES

[1] 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.

[2] 3DV Systems, http://www.3dvsystems.com/, 2005.

[3] M. Kawakita, T. Kurita, H. Kikuchi, and S. Inoue, “HDTV Axi-vision Camera,” *Proc. of International Broadcasting Conference*, pp. 397-404, 2002.

[4] A. Redert, M. op de Beeck, C. Fehn, W. IJsselsteijn, M. Pollefeys, L. van Gool, E. Ofek, I. Sexton, and P. Surman, “ATTEST – Advanced Three-Dimensional Television System Technologies,” *Proc. of International Symposium on 3D Data Processing Visualization and Transmission*, pp. 313–319, 2002.

[5] J. Cha, S.M. Kim, S.Y. Kim. S.U. Yoon, 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.

[6] G. Um, K.Y. Kim, C. Ahn, and K.H. Lee, “Thee-dimensional Scene Reconstruction Using Multi-view Images and Depth Camera,” *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XII*, vol. 5664, pp. 271-280, 2005.

[7] R. Tsai, “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses,” *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323-344, 1987.