Proceedings of the

International Conference on 3D Systems and Applications

3DSA 2010

www.3dsa.org/

International Conference on 3D Systems and Applications
General Academy Center, Tokyo, Japan, May 19-21, 2010

3DSA 2010

Table of Contents

Multi-View Depth Generation using Multi-Depth Camera System

Yo-Sung Ho and Yun-Suk Kang
Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong Buk-gu, Gwangju, 500-712, Korea
{hoyo, yunsuk}@gist.ac.kr

Abstract— **In this paper, we propose a multi-view depth generation method using a multi-depth camera system that is composed of five video cameras and five TOF depth cameras. After performing a few steps of preprocessing on the captured color and depth images, the depth information from the TOF cameras is warped into the color image positions and used as the initial disparity values. By applying the stereo matching method using belief propagation with this initial disparity information, we obtained more accurate multi-view disparity maps than the result without the initial disparity information.**

Keywords-component; Depth map generation, multi-depth camera system, TOF camera, 3DTV

I. Introduction

In recent years, three-dimensional TV (3DTV) and free viewpoint TV (FTV) are investigated as next generation broadcasting systems that can satisfy the demand for more realistic multimedia services [1]. By watching 3D video contents on the auto-stereoscopic 3D display system, users feel more immersive and realistic sense at multiple viewpoints. In order to generate 3D video contents, multi-view image and depth information are required. The multi-view image is a set of images of the same scene captured by multiple cameras. We can generate depth maps representing distance information of the scene using the multi-view image. Based on depth maps, we can reconstruct intermediate images of the scene at arbitrary virtual viewpoints. It provides us a wider and more natural 3D view. Therefore, generation of high quality depth maps is essential since their quality heavily influences the quality of 3D video contents.

In general, there are two main categories for acquiring depth information of the scene. One is based on passive range sensors and the other is based on active range sensors. Stereo matching is the most popular one of the methods using passive range sensors [2]. It does not require any special equipment except cameras. However, the result of stereo matching relies on the texture of input images and object placement of the scene in many times. Thus, we have opportunities of having wrong depth value in textureless and occluded regions.

In the category of active range sensor, methods use equipments that measure the range of scene such as well-known time-of-flight (TOF) depth cameras. The TOF depth camera emits infra-red signals itself and then measuring the arriving back time of the signals. Then it translates the measured time to depth information of the scene. However, in spite of its high price, it merely yields small spatial resolution images with noise. Furthermore, we cannot use it for such as broad indoor studio or outdoor capturing since it limits its maximum capturing range.

Recently, there have been several attempts to fuse the active and passive range sensor based methods to obtain the advantages and discard disadvantages of each method. Many of these attempts are consisted of stereo or multi-view camera with one TOF depth camera. S. A. Gudmundsson *et al.* used stereo camera and a TOF depth camera. They obtained depth information from the depth camera and transferred it to the color image position for initializing the disparity for stereo matching [3]. B. Bartczak *et al.* captured HD multi-view color images and one low resolution depth image. They also warped a low resolution depth image to the reference views for depth map generation [4]. More approaches mention the combination of multiple cameras and a TOF depth camera to generate high quality depth maps [5] [6].

In this paper, we introduce our multi-depth camera system that has multiple video cameras and multiple depth cameras and explain the multi-view depth generation process using the multi-depth camera system. At first, we capture five color images and five depth images of the scene. After several preprocessing steps, depth images are warped to the corresponding video camera positions and used as initial disparity values for stereo matching. With this initial value, we calculate the data cost and then refine using belief propagation to generate multi-view disparity maps.

II. Multi-Depth Camera System

A. Camera Setup

In the proposed system, we combine five video cameras and five depth cameras [7]. The video camera model we use is Basler Pylon GigE [8] and the TOF depth camera model is Swiss Ranger SR4000 [9]. Five video cameras are mounted on the first row, and five depth cameras are also mounted on the second row, right below each video camera. They are all arranged in one-dimensional (1D) parallel setup. The width of the camera frame is 120cm and we can regulate the height of two parallel bars for mounting cameras from 50cm to 160cm. The minimum distances between two cameras are approximately 6.5cm in the horizontal direction and 4cm in the vertical direction.

Figure 1 shows our multi-depth camera system. There is a synchronizer that enables simultaneous capturing of video cameras [10]. However, in the case of TOF depth cameras, we have no synchronizer. Instead of it, we modified the capturing software for simultaneous capturing. Therefore, there is no synchronization between video cameras and depth cameras. Although there is no problem when we capture a static scene, we have different starting points of two types of cameras for dynamic scene capturing.

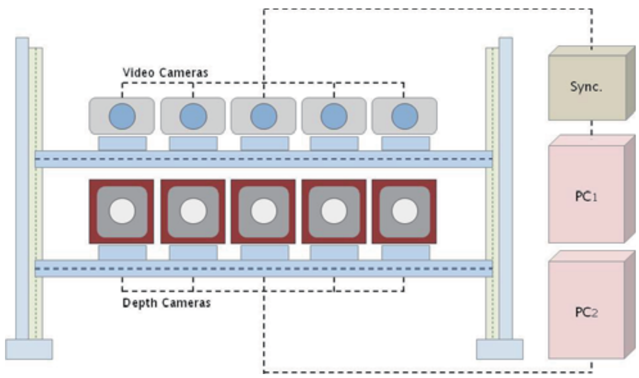
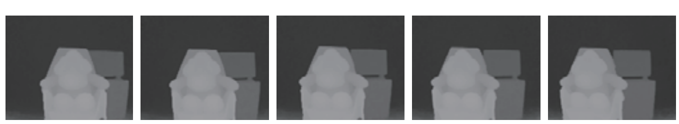


Figure 1. Multi-depth camera system

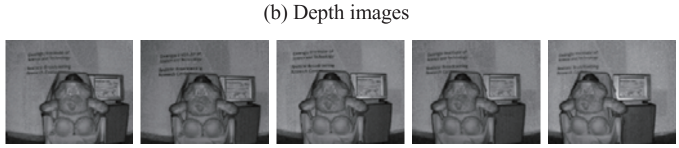
B. Data Acquisition

By using the multi-depth camera system, we obtain high resolution multi-view color image and low resolution multi-view depth image. Our video camera fundamentally provides HD resolution. The lower resolution image is generated by cropping the original image boundaries to the desired resolution. Figure 2(a) shows the multi-view color images of resolution of 800×600.

(a) Color images



(b) Depth images



(c) Intensity images

Figure 2. Captured image set

Figure 2(b) and Figure 2(c) are output images of five depth cameras. SR4000 gives us two types of output images which are depth images and intensity images. The depth image represents range information of the scene with 256 levels. The intensity image is similar to the grey image of the scene. We usually use the intensity image for point extraction such as camera calibration.

However, SR4000 has a few inherent problems. One problem is relatively low resolution of the output image. SR4000 provides 176×144 resolution. The other problem is that there is a lens distortion in the output image. In addition, we cannot use more than three depth cameras simultaneously due to the interference between infra-red signals. During the capturing with multiple cameras, the depth cameras modulate their infra-red signals with different modulation frequencies. Because SR4000 has three different frequencies for modulation, we captured one frame of depth images of the scene twice. We used three cameras with power off of the other two, and then vice versa.

III. Proposed Method

In this section, we introduce how to generate depth maps using the proposed multi-depth camera system. After capturing, we perform a few preprocessing steps to images. Then we warp the low resolution depth image to the color image position to initialize the search range of data cost calculation. We finally obtain high quality multi-view disparity maps as a result of stereo matching using belief propagation. The whole process is shown in Fig. 3.

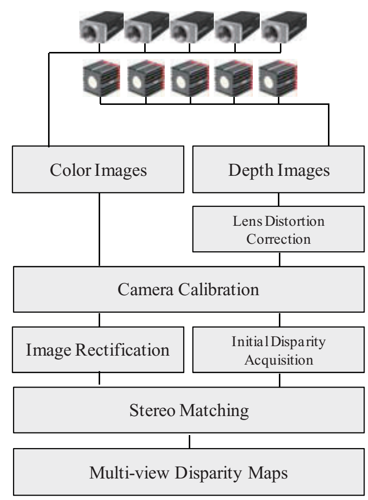


Figure 3. Procedure of the proposed method

A. Preprocessing

For captured color and depth images, it is essential to perform a few preprocessing processes. The purpose of the preprocessing is to obtain metric information and to correct the mechanical defects to increase the image correlation between views and between different cameras. In this subsection, we briefly explain each preprocessing step.

As shown in Fig. 2(b) and Fig. 2(c), depth images captured by SR4000 have a large amount of lens radial distortion. 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. Therefore, we corrected the lens distortion using the camera calibration toolbox for MATLAB after obtaining the camera parameters [11]. The result is shown in Fig. 4(a).

After minimizing the lens distortion of depth images, we estimate camera parameters of all the cameras. In this step we also used Camera Calibration Toolbox for MATLAB. In order to calibrate, we captured several images of grid-pattern in different poses. In the case of depth cameras, lens distortion in the images of grid-pattern is corrected and these images are used for camera calibration.

After camera calibration, we perform multi-view image rectification to the color images to minimize the geometric error [12]. As a result, we obtain the rectified multi-view image shown in Fig. 4(b) and rectified camera parameters.

Before After

(a) Result of lens distortion correction

Before After

(b) Results of multi-view image rectification

Figure 4. Preprocessing for captured images

B. Initial Disparity Acquisition

In this step, we regard depth information obtained by the depth cameras as the initial disparity of multi-view image for stereo matching. Depth values in the depth camera images are transferred to the color image positions using 3D warping. In order to calculate the real depth Z at each depth value, we use

$$Z(i,j) = \frac{1}{\frac{D(i,j)}{D_{max}} \cdot \left(\frac{1}{d_{min}} - \frac{1}{d_{max}}\right) + \frac{1}{d_{max}}} \tag{1}$$

where D_{max}, D_{min}, $D(i, j)$, d_{max}, and d_{min} represent the maximum and minimum depth values, depth value of pixel (i, j), and maximum and minimum disparity values of the scene, respectively.

Then, the transferred depth values to the color image positions are changed to disparity values d by using

$$d = \frac{f \cdot B}{Z} \tag{2}$$

where f, and B are focal length and distance between two adjacent cameras, respectively.

Figure 5 shows the warped results. Because we adjust each field of view of color camera and depth camera and there is displacement between two types of cameras only in the vertical direction, the object boundary regions of the initial disparity image is matched well to the color image.

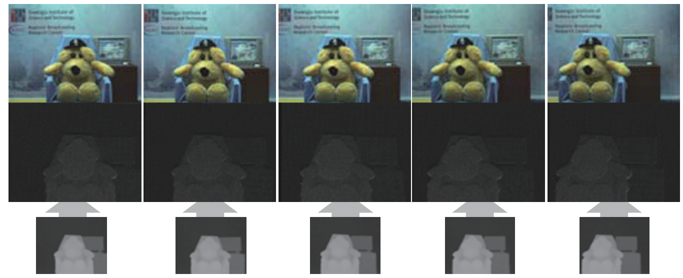


Figure 5. Initial disparity acquisition for stereo matching

C. Multi-view Depth Generation

In the final step, we obtain the disparity map of each view based on the initial disparity values from the depth cameras. However, as shown in Fig. 5, there are a great number of pixels having no initial disparity value due to the resolution difference between color images and depth camera images in 3D warping. Therefore, for these pixels, we estimate the initial values using the neighboring pixels.

In other to estimate the initial values, we consider three cases. The first case is when the estimated initial value at the pixel position having no initial value is the same as all the neighboring pixel values used. In this case, we trust these initial values and assign relatively short search range.

The second case is when the used neighboring pixels have different initial disparity values. We then take the average and assign the search range as the difference between the maximum and minimum neighboring pixel values.

The last case is when there is no pixel having its initial disparity in the search window around the current position. In this case, we just set the search range as the difference between the maximum and minimum disparities without the initial disparity. These three cases are shown in Fig. 6. By using the initial disparity values like this, we can obtain accurate disparity value for textureless and occluded regions.

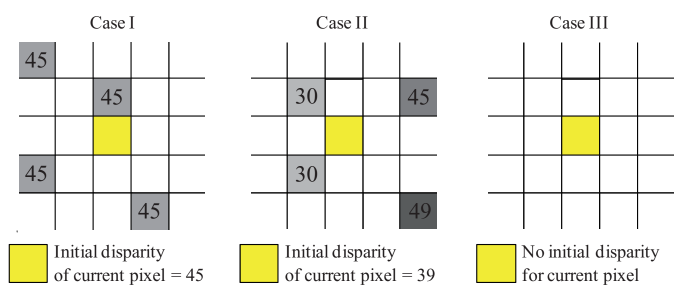


Figure 6. Example of initial disparity estimation for data cost calculation

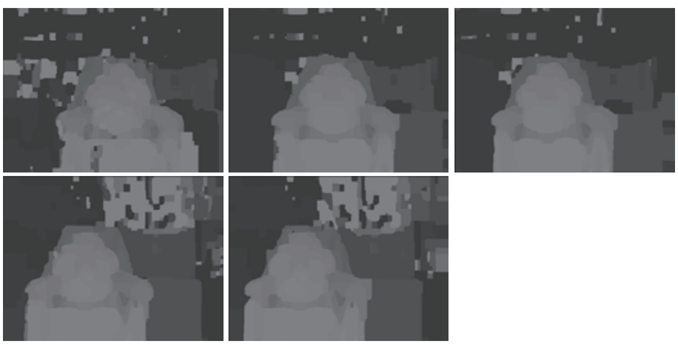
After that, the disparity map of each view is obtained that has disparity values which minimize the sum of absolute difference (SAD) between the current and reference images. Then, we refine this using belief propagation method [13].

IV. Experimental Results

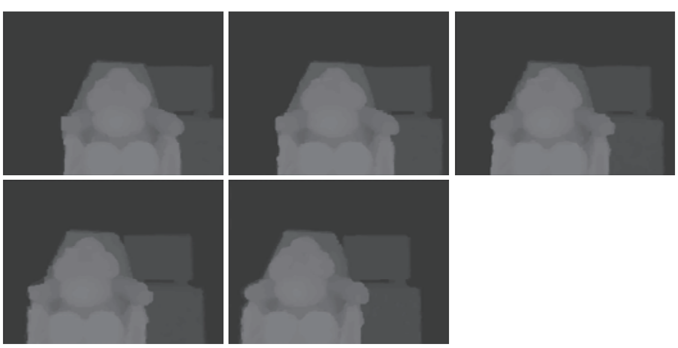
In order to test the proposed method, we used the multi-depth camera system shown in Fig. 7. As shown in Fig. 2, we captured one frame of 800 × 600 color images at five viewpoints and 176 × 144 depth images at another five viewpoints. The reason we decided the color image resolution like that is to match two types of camera field of views and to reduce the resolution difference. For the captured images, we performed aforementioned preprocessing as shown in Fig. 4.

Figure 7. Camera setup

In the proposed method, we used 11×11 search window for pixels having no initial disparity. We used 3×3 SAD window for the pixels having the initial disparity values. For pixels that have no initial disparity values, we adopted 7×7 SAD window. When we have a measured disparity value or reliably estimated initial value at a pixel, then we assign the short search range as ±2 pixels.



(a) Disparity maps by previous method



(b) Disparity maps by proposed method

Figure 8. Generated multi-view disparity maps

Figure 8(a) and Figure 8(b) show the generated depth maps based on stereo matching using belief propagation and the proposed method, respectively. For the background region, the previous method failed to estimate accurate disparity since there is very weak texture on the wall. However, the result of the proposed method shows that the disparity value of the background is accurate and stable. In addition, the overall quality of generated depth maps by the proposed method is better than the result by the previous method due to the correctly measured and estimated initial disparity.

V. CONCLUSION

In this paper, we proposed multi-view depth generation method using a multi-depth camera system. After capturing one frame of five color images and five depth images, we performed preprocessing to increase inter-view and inter-camera correlation. Then, depth images are warped to initial disparity values for stereo matching. For pixels having no initial disparity values, we estimated the values using neighboring pixels and then assigned search ranges. With the initial disparity values, we obtained multi-view disparity map as the result of the stereo matching. Since the proposed method processed the result based on the initial disparity values, we obtained more correct and stable disparity values in weak texture regions and occluded region than the result of the previous method.

ACKNOWLEDGMENT

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] A. Smolic and P. Kauff, "Interactive 3D Video Representation and Coding Technologies," Proceedings of the IEEE, Spatial Issue on Advances in Video Coding and Delivery, vol. 93, no. 1, pp. 99-110, 2005.

[2] J. Sun, N. N. Zheng, and H. Y. Shum, "Stereo Matching Using Belief Propagation," IEEE Transactions of Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pp. 787-800, 2003.

[3] S. A. Gudmundsson, H. Aanaes, and R. Larsen, "Fusion of Stereo Vision and Time-of-Filght Imaging for Improved 3D Estimation," International Journal of Intelligent Systems Technologies and Applications, vol. 5, no. 3, pp. 425-433, 2008.

[4] B. Bartczak and R. Koch, "Dense Depth Maps from Low Resolution Time-of-Flight Depth and High Resolution Color Views," Proc. of 5th International Symposium on Visual Computing, pp. 1-12, Nov. 2009.

[5] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of Time-of-Flight Depth and Stereo for High Accuracy Depth Maps," Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 231-236, 2008.

[6] K. D. Kuhnert and M. Stommel, "Fusion of Stereo-Camera and PMD-Camera Data for Real-time Suited Precise 3D Environment Reconstruction," Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4780-4785, Mar. 2006.

[7] Y. S. Kang, E. K. Lee, and Y. S. Ho, "Multi-Depth Camera System for 3D Video Generation," Proc. of International Workshow on Advanced Image Technology (IWAIT), pp. 44(1-6), Jan. 2010.

[8] http://www.baslerweb.com, Basler Pylon GigE Camera piA I1900-32gm/gc.

[9] http://www.mesa-imaging.ch, SwissRanger SR 4000.

[10] http://sine.ni.com, National Instruments CA-1000.

[11] http://www.vision.caltech.edu/bouguetj, Camera Calibration Toolbox for MATLAB.

[12] Y. S. Kang and Y. S. Ho, "Geometrical Compensation for Multi-view Video in Multiple Camera Array," Proc. of International Symposium on Electronics and Marine (ELMAR), pp. 83-86, Sept. 2008.

[13] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Belief Propagation for Early Vision," International Journal of Computer Vision, vol. 70, no. 1, pp. 41-54, 2006.