Efficient Disparity Map Generation for Moving Multi-camera System using TOF Depth Sensor

Yun-Suk Kang and Yo-Sung Ho
School of Information and Communications
Gwangju Institute of Science and Technology (GIST)
Gwangju, Republic of Korea
{yunsuk, hoyo}@gist.ac.kr

***Abstract*— In this paper, we propose an efficient disparity map generation method for multi-view video sequences captured by a forward moving multi-camera system. Forward camera moving increases the minimum and maximum disparity value of each frame with time. Thus, computational complexity can be increased and the quality of generated disparity can be degraded since the stereo matching operation has to deal with a wide range of the disparity candidates. In order to solve this problem, we employ a time-of-flight (TOF) depth sensor as a guide to find the minimum and maximum disparity values for each frame. Then, the stereo matching process for video sequences captured by the moving multi-camera system becomes simple without quality degradation.**

Keywords— moving multi-camera, depth generation, stereo matching, time-of-flight depth sensor

I. INTRODUCTION

Three-dimensional (3D) video is one of the most powerful media which provides immersive and realistic sense to users [1]. Unlike the conventional two-dimensional (2D) video that has a single viewpoint, the 3D video is composed of more than two viewpoints including the parallax between adjacent views. Therefore, by watching the video captured by these multiple viewpoints through stereoscopic or auto-stereoscopic displays, we can enjoy the 3D feeling. However, one problem of the multiple viewpoints is the huge amount of image data. The amount of image data is multiplied by the number of cameras. It affects the encoding and transmission processes. Also, because of the camera size, it is difficult to construct a dense camera arrangement, which is highly related to generate more natural 3D view.

In order to solve the mentioned problems, depth information of the scene is required. The depth of the scene represents the range to each object from the camera in the scene. Using the color image and the corresponding depth, we can synthesize an intermediate view image between two real view images by using the depth image-based rendering (DIBR) technique [2]. Thus, we can obtain a densely captured multi-view image without a relatively small number of cameras which are not densely arranged. Also, the efficiency of encoding and transmission can be increased since if we encode the color and depth image at the encoder side instead of all the multiple color images, then the decoder decodes the transmitted data and generate the multi-view using the color and depth images.

Stereo matching is the most famous depth estimation method that takes two input images and generates the disparity maps [3]. For a pixel and the surrounded region in the target image, stereo matching finds the best matched pixel with respect to the color difference in the reference image. Then, the disparity is pixel coordinate difference in y-direction from the current pixel to the best matched pixel. However, stereo matching has critical weaknesses for such as the occlusion and textureless regions, since it operates based on color difference. It also has very high computational complexity. In general, users assign the minimum and maximum disparity values of the scene to decrease the processing time, then, the stereo matching algorithm calculates the proper disparity value for each pixel in the input disparity range.

However, in the case of the video sequence, usually the maximum disparity value varies with time. Especially for the scenes captured by the moving forward and backward multi-cameras or scenes including zoom-in and zoom-out, both the minimum and maximum disparity values can be widely changed with time. Therefore, the input minimum and maximum disparity values to the stereo matching have to cover wide range. It leads to not only increase the complexity but also decrease the quality of the disparity map.

In this paper, we propose an efficient disparity map generation method for forward moving multi-camera system using a time-of-flight (TOF) depth sensor as a guide. Since the TOF depth sensor captures the depth of the scene in real-time, it helps to calculate the appropriate minimum and maximum disparity values as the input to the stereo matching for each frame. It also decreases the complexity of the stereo matching, and also maintains the quality of the disparity maps. Moreover, the proposed method does not require the registration process between the color image and the TOF depth image. In some fusion methods for depth acquisition, which means the combined system of the color and TOF depth sensor [4] [5], the registration between the color and TOF depth is one of the difficulties due to the differences in the image resolution and field of view. Therefore, the proposed method can be utilized various camera systems that have various characteristics.

978-1-4799-0944-5/13/$31.00 ©2013 IEEE

II. Data Acquisition from Forward Moving Multi-camera

Figure 1 shows the moving multi-camera arrangement on the rail which makes the camera arrangement move forward and backward manually. Therefore, the captured video sequences by this moving multi-camera have the similar effect to multi-view zoom-in or zoom-out.

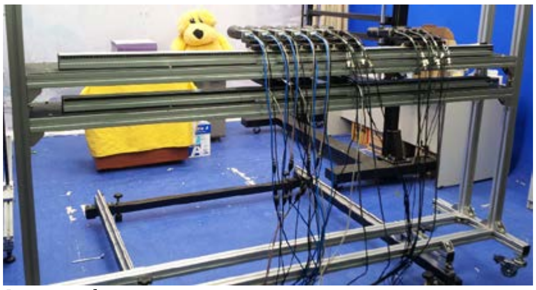


Fig. 1. Moving multi-camera arrangement

(a) (b)

Fig. 2. Multi-camera and TOF depth sensor (a) three color video cameras (b) SR4000

We used three color video cameras, shown in Fig 2(a), and one TOF depth sensor, shown in Fig. 2(b). The TOF depth sensor model is SR4000. The color cameras are parallel arranged with the interval of about 7cm, and the TOF depth sensor

Fig. 3. Captured images

We captured 100 frames of a scene using the moving multi-camera with forward moving, as shown in Fig. 3. The moving distance was approximately 60cm, and the moving speed was not steady since we manually move the camera frame. As the frame number increases, the minimum and maximum disparity values increases, too. Also all the pixel values of the TOF depth image becomes larger.

For the captured color images, the camera parameters are required for the multi-view image rectification [4]. The rectified three view images have parallel epipolar lines. After that, multi-view color correction is performed to reduce the color inconsistency among viewpoints.

III. Proposed Method

Figure 4 shows the process of the proposed method. In this section, we assume that the multi-view color image is calibrated, rectified, and color corrected. Then, we explain each step of the proposed method.

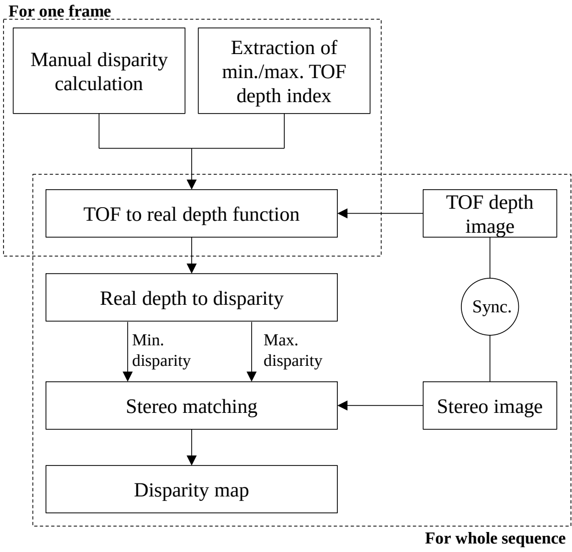


Fig. 4. Flow chart of the proposed method

A. Mapping function from TOF to real depth

The first part of the proposed method is to design the mapping function from the depth index D of the TOF depth image to the real depth Z. These two variables are proportional to each other. We consider the first frame of whole sequence to design the mapping function. However, we do not have the real depth information of the scene. Therefore, we manually measure the minimum and maximum disparity values of the first frame. Then, Eq. (1) calculates *Z* values using the disparity *d*, focal length *f*, and baseline *B*.

$$Z = \frac{f \cdot B}{d} \tag{1}$$

Let the nearest and farthest real depth values be Z_{near} and Z_{far}, respectively. Then, Z_{near} is inverse proportional to the minimum disparity d_{min}, and Z_{far} is also inverse proportional to the maximum disparity d_{max}. Z_{near} and Z_{far} correspond to D_{max}

and D_{min}, respectively, which are the maximum and minimum pixel values of the TOF depth image at the first frame. Finally, we can design the mapping function from D to Z as a linear function **M** shown in Fig. 5.

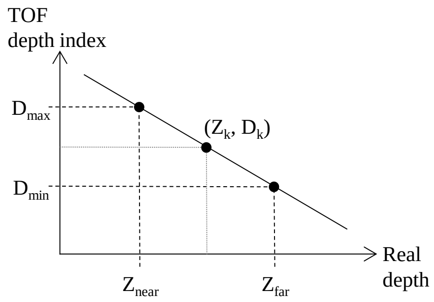


Fig. 5. Mapping function M

B. *Stereo matching with proper disparity values*

After defining the TOF to real depth mapping function **M**, we can calculate Z_k which is the real depth of an arbitrary TOF depth index D_k. It is equivalent to calculate the disparity d_k in the color image for D_k. Therefore, the minimum and maximum disparity values of n-th frame, $d_{min,n}$ and $d_{max,n}$, are calculated as Eq. (2) and Eq. (3).

$$d_{min,n} = \frac{f \cdot B}{Z_{far}} = \frac{f \cdot B}{\mathrm{M}(D_{min,n})} \tag{2}$$

$$d_{max,n} = \frac{f \cdot B}{Z_{near}} = \frac{f \cdot B}{\mathrm{M}(D_{max,n})} \tag{3}$$

Lastly, the input minimum and maximum disparity values for the stereo matching of n-th frame $d'_{min,n}$ and $d'_{max,n}$ are calculated as Eq. (4) and Eq. (5), respectively.

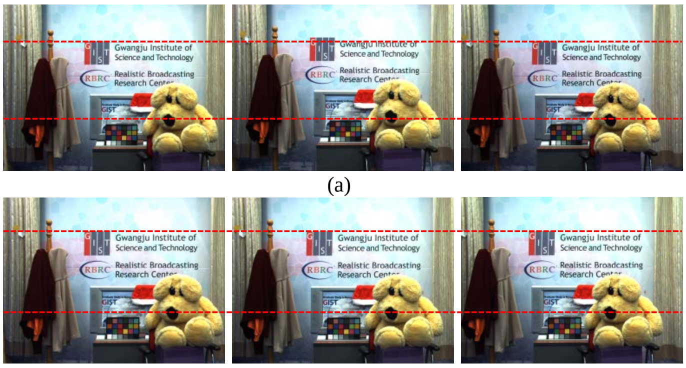
$$d'_{min,n} = \lfloor d_{min,n} \rfloor - c \tag{4}$$

$$d'_{max,n} = \lceil d_{max,n} \rceil + c \tag{5}$$

By considering the characteristics of the disparity, $d'_{min,n}$ is rounded down to nearest integer, while $d'_{max,n}$ is rounded up to nearest integer. The constant c is used as an offset value to avoid that the real disparity range of n-th frame exceeds $d'_{min,n}$ and $d'_{max,n}$.

IV. Experimental Results

The resolutions of the captured images were 1280x960 for color and 176x144 for TOF depth. In order to test the proposed method, three color viewpoints were calibrated [7], rectified [6], and color corrected [8], as shown in Fig. 6. In the original images, there exists a few pixels of vertical mismatch and the color of view 3 is darker than the other views. However, the multi-view image rectification reduced the vertical pixel mismatches and the color correction also decreases the color inconsistent among viewpoints.

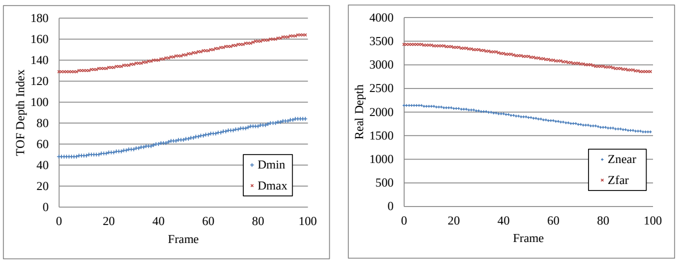


(a)

(b)

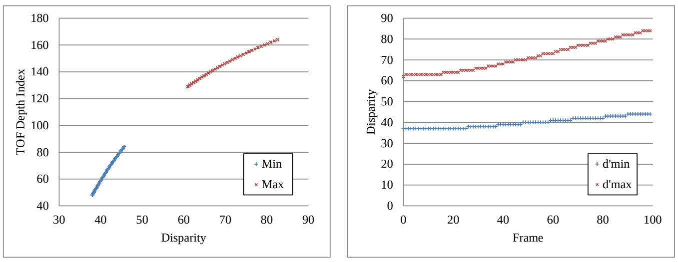
Fig. 6. Multi-view image rectification and color correction (a) original images (b) rectified and color corrected results

Figure 7(a) shows D_{min} and D_{max} of the TOF depth images. D_{min} and D_{max} are increased with time since the cameras only moved forward. The manually measured from the first frame are 38 and 61, respectively. Then, Z_{near} and Z_{far} calculated by Eq. (1) are shown in Fig. 7(b).



(a) (b)

Fig. 7. TOF depth index and real depth values with respect to frames (a) depth index (b) real depth



(a) (b)

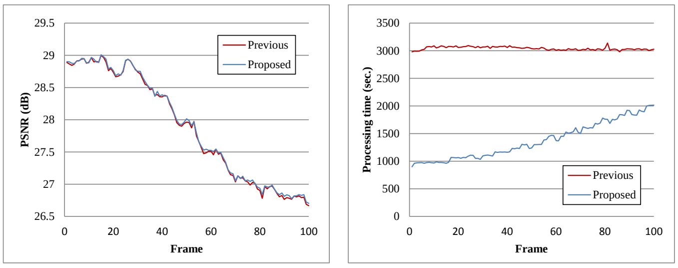
Fig. 8. TOF depth index and disparity (a) TOF and disparity (b) disparity with frame

Figure 8(a) shows the relationship between the TOF depth index and the disparity values for the nearest object and the farthest background of each frame. Finally, the input minimum and maximum disparity values for the stereo matching calculated by Eq. (4) and Eq. (5) are shown in Fig. 8(b). We used the offset constant c as 1.

Then, the rectified and color corrected multi-view images of view 1 and view 3 become the input for the stereo matching using belief propagation [9] with $d'_{min,n}$ and $d'_{max,n}$ for n-th frame. Figure 9 shows the calculated disparity maps of view 1 and view 3, and the synthesized color images of view 2.

Fig. 9. Generated disparity maps and synthesized view images

In order to evaluate the performance of the proposed method, we compared the PSNR values between the original and synthesized images of view 2, as shown in Fig. 10(a). Also, we compared the processing time with the conventional stereo matching algorithm. The fixed input minimum and maximum disparity values were 37 and 85 for all frames, respectively. They were measured manually and offset by *c*.



(a) (b)

Fig. 10. Performance evaluation (a) PSNR (b) processing time

The comparison of processing time to generate disparity maps for view 1 and view 3 for a frame is indicated in Fig. 10(b). Compared to the stereo matching without the proper minimum and maximum disparity values for each frame [9], the proposed method decreased the stereo matching complexity. As shown in the graphs, the proposed method remarkably reduced the processing time as maintaining or slightly outperforming the quality of the disparity.

Table 1 shows the average of PSNR and processing time. The average PSNR was increased by 0.2dB and the average processing time is reduced by about 55%.

TABLE I. AVERAGE PSNR AND PROCESSING TIME

	PSNR (dB)	**Time (sec.)**
Previous	27.88	3043.67
Proposed	27.90	1369.88

V. CONCLUSION

We proposed an efficient disparity map generation method for a moving multi-camera system using the TOF depth sensor as a guide. The proposed method calculates the proper minimum and maximum disparity values for each frame to increase the efficiency of the stereo matching. These input disparity values to the stereo matching are calculated using the TOF depth images. The experimental results of the proposed method decreases about 55% of processing time compared to the stereo matching without the proper disparity range for each frame. Also, the proposed method slightly increases the view synthesis quality that means the quality of disparity maps are maintained or slightly enhanced.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0009228).

REFERENCES

[1] A. Smolic, K. Muller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, "3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards," IEEE International Conference on Multimedia and Expo (ICME), pp. 2161-2164, July 2006.

[2] 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. 5921, pp. 93-104, May 2004.

[3] 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, Dec. 2006.

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

[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, June 2008.

[6] Y.S. Kang and Y.S. Ho, "An Efficient Image Rectification Method for Parallel Multi-Camera Arrangement," IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1041-1048, Aug. 2011.

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

[8] N. Joshi, B. Wilburn, V. Vaish, M. Levoy, and M. Horowitz, "Automatic Color Calibration for Large Camera Arrays," in UCSD CSE Tech. Rep. CS2005-0821, May 2005.

[9] Q. Yang, L. Wang, and N. Ahuja, "A Constant-space Belief Propagation Algorithm for Stereo Matching," Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1458-1465, June 2010.