Multi-Depth Estimation using Time-of-Flight Depth Camera

Yun-Suk Kang and Yo-Sung Ho
School of Information and Communications
Gwangju Institute of Science and Technology, South Korea

1 Introduction

Recently, three-dimensional TV (3DTV) and 3D video contents are one of the most attractive issues in the world. By watching 3D video contents, users feel more realistic impression from more than two viewpoints. 3D contents or 3D images are widely used for TV programs, games, education, advertisements, culture, and so on. Moreover, users can watch these various 3D video contents using their personal imaging device or cell phone. Therefore, there is an increasing demand for various and realistic 3D contents, and the technologies for 3D capturing and 3D content generation has been developed together.

Basically, capturing of the 3D image is to capture the same scene from two different viewpoints. Therefore, it is required two cameras or stereo camera. In the case of the multi-view image, which is capture by the multiple cameras, provides the wider field of view and more realistic 3D feeling. The multi-view image is captured by an arrangement of more than two cameras. Also, we can construct several types of camera arrangements with the multiple cameras.

These stereo or multi-view images require the scene's depth information to generate an intermediate viewpoint image. We can synthesize the intermediate viewpoints using color images, corresponding depth maps, and camera geometric information. Since the quality of the synthesized image is highly dependent on the depth quality, to obtain the accurate depth information of the scene is very important for 3D content generation. Generally, the depth of the scene is calculated as the disparity form by the stereo matching.

The depth sensors are frequently used for scene's depth acquisition, recently. The depth sensor emits and receives the light signal to the space, and then it measure the range information of the scene in real-time. The measured depth information is also used for 3D content generation.

In this chapter, we introduce various 3D capturing methods and camera systems with our proposed camera system composed of five video cameras and three time-of-flight (ToF) depth cameras. We describe the principle and characteristics of ToF depth camera. We also explain the 3D image processing techniques such as extracting geometrical and physical information of the cameras, reducing noise and distortion, estimating the scene's depth, and intermediate view synthesis or 3D scene reconstruction. This chapter is organized as follows. In section 2, we explain cameras, camera arrangements, and their characteristics. In section 3, the proposed camera system and 3D image processing techniques to generate the depth information of the scene. After showing the experimental results in section 4, we finally conclude the chapter in section 5.

2 3D Capturing Methods using Various Camera Systems

In this section, we introduce various 3D capturing methods using multi-camera systems. Firstly, we explain multiple camera arrangements. Then, we introduce the characteristics of the ToF depth camera. Finally, we describe fusion camera systems such as the multi-depth camera system, that is composed of multiple color and ToF depth cameras.

2.1 Multiple Camera Arrangements and Capturing

In order to capture 3D contents, it is required at least two cameras, and these cameras have to be arranged in a certain arrangement in the space. In general, the parallel camera arrangements, which arrange all the cameras on a line in the space, are frequently used. In the parallel camera arrangements, the orientations of all the cameras are the same. As shown in Fig. 1(a), the parallel arrangements provide not only the wider capturing field, but also disparity information that is proportional to the object range from the camera.

Figure 1(b) shows the convergent camera arrangement whose optical axes passes one convergent point in the captured space. Usually, it is more difficult to calculate disparity information of the convergent arrangements than the parallel arrangements. However, since it captures the detail of the scene, the convergent arrangements are used for applications such as 3D scene or object reconstruction. The 2D arrangements are sometimes used for the light field rendering.

(a) (b)

Figure 1: Multiple camera arrangements: (a) parallel (b) convergent types

When we use more than two cameras, the basic distance between adjacent cameras is 6.5cm which is the eye to eye distance of human. However, this basic distance can be changed due to camera aperture or the other conditions. The important thing is that the disparity is clearly appeared in the captured images. Moreover, since the camera misalignment in the camera arrangement can causes large mismatch in the captured image, careful and precise camera arrangement is required. Also, camera's physical characteristics have to be considered, such as white balance, focus, exposure, and so on.

2.2 Time-of-Flight Depth Camera

The ToF depth camera is the most famous one of depth sensors. The principle of the ToF depth camera is to use the phase difference of the emitted light from an illumination source of the camera to the object and back to the camera. The full-phase value means the maximum capturing range. Thus, the modulation frequency determines the maximum range of the camera.

There have been several ToF depth cameras. The first real-time ToF depth camera model is 3DV's ZCam. ZCam is composed of the combination of color camera and depth sensor using infra-red signal. It captures color and depth images at the same viewpoint. However, ZCam has some problem such as an optical noise and object-dependent capturing, and large camera aperture.

After ZCam, various types of ToF depth cameras have been developed by Swiss Ranger, PMD, Prime Sense, and so on. SR4000, developed by Swiss Ranger, is shown in Fig. 2. This model is emits the light signal from the LEDs which is covered with the illumination cover, and the optical filter receives the arriving signal. SR4000 is very small (65x65x68mm) and inexpensive. The minimum and maximum capturing ranges are 0.3m and 5.0m, respectively. It can capture the scene and provide two types of images as shown in Fig. 2, depth and intensity images.

Figure 2: Time-of-Flight depth camera – SR4000 and its output depth and intensity images.

However, output images of SR4000 shown in Fig. 2 have a large amount of lens radial distortion. There exists boundary noise which means the median depth value between the foreground and background. Also, the other limitation is that the resolution is too small (176x144). In order to use these output images, enhancing the image quality is required.

2.3 Fusion Camera Systems

As explained in previous sections, there are two types of obtaining scene's depth information. There are the passive range sensor based methods which are using the captured color images. On one hand, there are the active range sensor based methods which measure the range of the scene using the special equipment. Stereo matching is the most popular one of the passive range sensor based method. These methods are efficient since we only require the color images of the scene. However, for the textureless regions or occluded regions in the images, it is hard to obtain the accurate depth. Also, shape from silhouette, shape from focus, and shape from motion belong to this category.

The active range sensors based method directly measures the range from the camera to the object or background of the scene. The ToF depth camera is the representative one of this category. In this category, there are also 3D scanners, structured light pattern, and so on. Although the ToF depth camera measure the depth of the scene in real-time, the output image resolution is too small and noisy. Moreover, since it operates and provides images based on the receiving light signal, it captures the depth within the capturing range in indoor studio.

In order to complement those two methods each other, some approaches combine two types of depth acquisition methods. In order words, fusion camera systems try to obtain the advantages and discard the disadvantages of each method. Usually, stereo or multi-view camera with a ToF depth camera are used together. S.A. Gudmundsson et al. combined a stereo camera and a ToF depth camera. The ToF depth image is transferred to the color image position to initialize the disparity for stereo matching. B. Bartczak et al. captured multiple color images and one low resolution ToF depth image. The ToF depth image warped to the color views for depth generation. Kuhnert et al. and Zhu et al. also mentioned the integration of the color and ToF depth cameras to estimate the depth information of the scene.

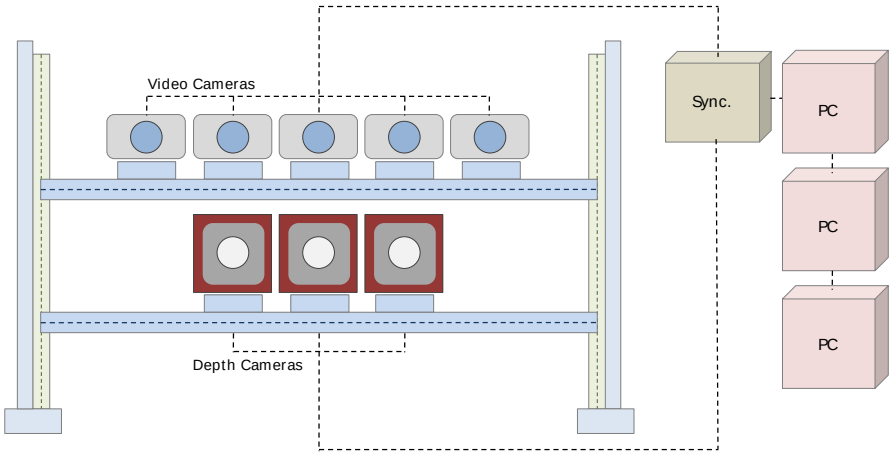


Figure 3: Multi-depth camera system

Figure 3 shows the proposed multi-depth camera system that is composed of five color cameras and three ToF depth cameras. Five color video cameras are arranged in parallel, and three ToF depth cameras are arranged in parallel below the color cameras. They are synchronized, and the control PCs capture and save the scene with the two different types of cameras.

3 3D Content Generation using Multiple Color and Depth Cameras

This section describes the method to generate 3D contents using the multi-depth camera system. Figure 4 shows the whole process of the proposed method. After capturing the scene, we perform several steps of post-processing for the captured images to reduce image noise and distortion. Then, the ToF depth camera data is 3D warped to the color image position to be the initial values for the depth generation. If the scene's depth is generated, we can synthesize the intermediate view images that mean the 3D contents.

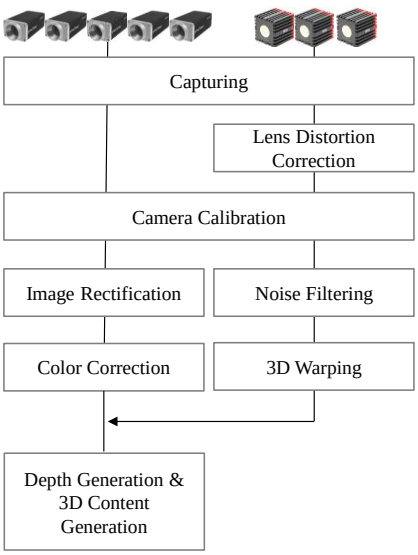


Figure 4: Procedure of the proposed method

3.1 Capturing and Image Characteristics

Figure 5 shows the captured color, ToF depth, and ToF intensity images by the multi-depth camera system shown in Fig. 3. The color images captured with 6.5cm of camera offset, and the distance between the color and ToF depth camera is 7cm.

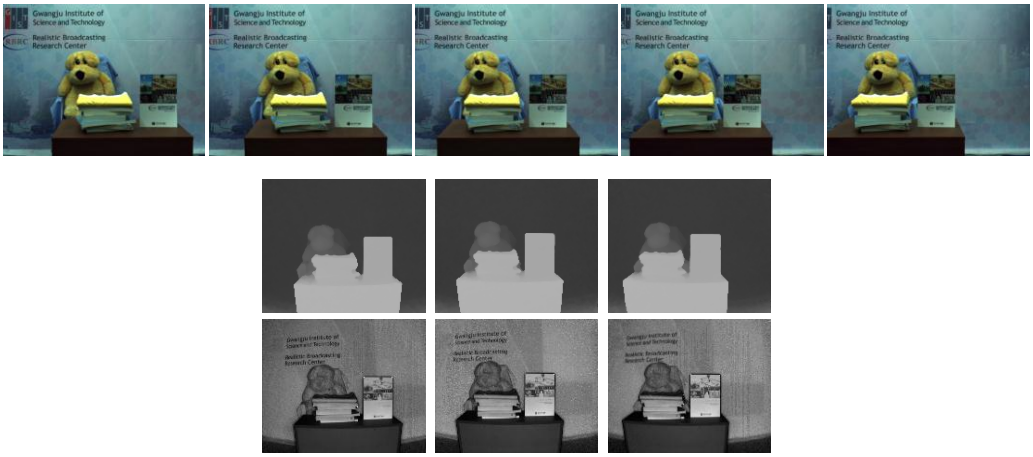


Figure 5: Captured images from the multi-depth camera system

Multi-depth camera system can simultaneously operate maximum three ToF depth cameras due to the modulation frequency of the depth camera. As mentioned before, the ToF depth camera emits and receives the

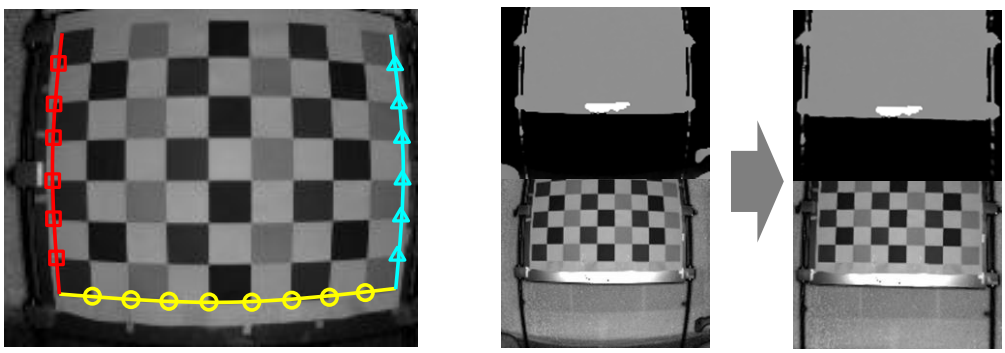
light signal and then generates the depth image. If there are multiple ToF depth cameras, they distinguish the receiving signals by their own modulation frequencies. SR4000 has three different modulation frequencies (29, 30, and 31MHz). Therefore, we operate three ToF depth cameras after the three cameras have different modulation frequency one another.

Captured color images have two problems. One is geometric error and the other is color consistency problem. The geometric error can be caused by camera misalignment. The color inconsistency occurs due to camera inherent characteristics, lights, and shadow. These problems make the 3D content generation process difficult, and also decrease the content quality.

The ToF depth camera also has built-in problems. As shown in Fig. 2, the output images have lens radial distortion and median-value error at the object boundaries. Furthermore, the ToF depth camera position is slightly different to the color camera position with respect to the optical axis direction. This causes the depth difference between the measured depth by the depth camera and the calculated depth by the color cameras.

3.2 Lens Distortion Correction

Lens radial distortion appeared in the ToF depth image causes a shape mismatch problem between the color and depth images. This distortion also affects the point-based processing such as camera calibration. In order to reconstruct the undistorted image, we perform lens distortion correction to the depth image using the intensity image. We can calculate the distortion center and the distortion parameters using at least three distorted lines in the intensity image. Figure 6(a) shows that three lines and their component points are extracted. By using these distortion center and parameters, we reconstruct the undistorted image as shown in Fig. 6(b). However, the depth image enlarges at the image borders during the reconstruction. This lens distortion correction algorithm is also applied to the color images in the same way.



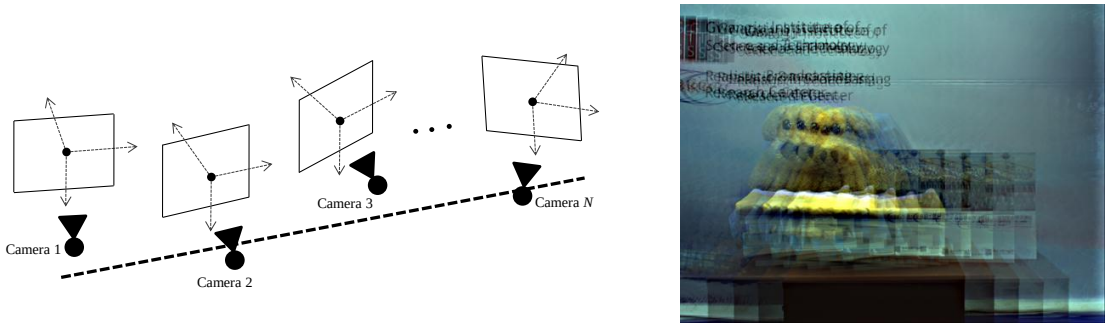
(a) (b)

Figure 6: Lens distortion of the ToF depth camera image: (a) Extracted three lines (b) Results

3.3 Camera Calibration

Camera parameters are the essential information for 3D image processing. They are composed of the intrinsic, extrinsic matrices, and translation vector. The 3x3 intrinsic matrix have camera's physical characteristics such as focal length and principal point. The 3x3 rotation matrix represents camera's orientation. The 3x1 translation vector is related to camera location.

Camera calibration is the process to estimate these camera parameters from the captured images. For camera calibration, it is required approximately ten checker board images for each camera. In the case of the ToF depth camera, camera calibration is performed after lens distortion correction.



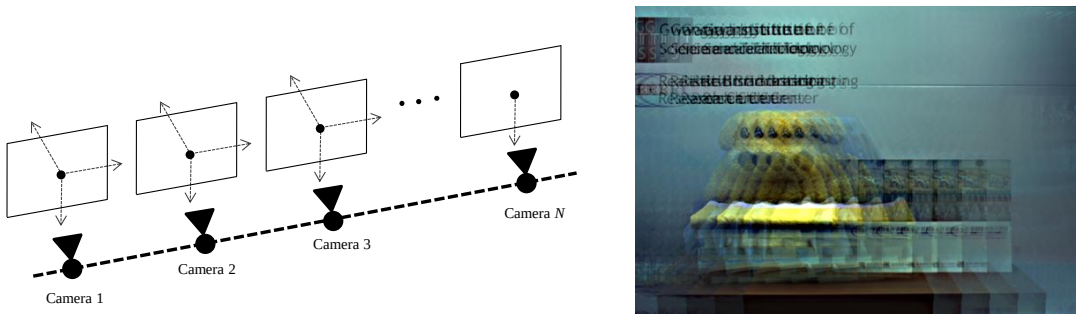
(a) (b)

Figure 7: Practical camera arrangement and the geometric error in the multi-view image

3.4 Multi-view Image Rectification

Multi-view image rectification reduces the geometric error in the multi-view image. As shown in Fig. 7(a), the multiple cameras are arranged and there are misalignments among cameras. The images captured by these multiple cameras have the geometric error that is appeared as an irregular disparity and vertical pixel mismatch among image as shown in Fig. 7(b). This geometric error decreases the efficiency of the stereo matching, since the matching region exceeds a 1D scanline region. In this case, the matching is performed on the epipolar line with the search window. Also, this error affects the visual quality of the 3D contents.

Therefore, the geometric error in the multi-view image has to be minimized using multi-view image rectification as shown in Fig. 8(a). At first, we estimate the camera parameters of the rectified multiple camera, and then calculate the rectifying transform using the original and estimated camera parameters. This transform verifies that there are the same intrinsic characteristics, the same rotation, and the translation vectors having only horizontal variations. Figure 8(b) shows the rectified multi-view image that has a uniform disparity range and few vertical pixel mismatches.



(a) (b)

Figure 8: Rectified camera arrangement and the reduced geometric error

3.5 Multi-view Color Correction

Multi-depth camera system uses the same camera model for color capturing. However, although we use the same camera model, there exists color inconsistency problem among multiple viewpoint images due to the different color response, light condition, and camera settings. This color inconsistency problem also makes the stereo matching difficult since the data cost term is calculated using color difference between the adjacent views.

The correction process of the color inconsistency problem in the multi-view image is performed using Macbeth color chart shown in Fig. 9. This chart has 24 colors, and each camera captures this chart and compares the captured value to the other views. In general, one viewpoint image is set to the reference, and the other viewpoints are adjusted to minimize the color inconsistence.

Figure 9: Color chart capturing for color correction

3.6 ToF Depth Warping

In order to use the ToF depth camera image at the color image position, each pixel data has to be transferred by using 3D warping. Before that, it is required to reduce the median-value noise at the boundary regions. The boundary regions mean the depth discontinuity regions in the ToF depth image. The median depth value does not exist in the color image, and it can be located in a wrong position in the color image position. Therefore, the median depth value has to be reduced.

In the proposed method, we use the shock filtering to the ToF depth image. As shown in Fig. 10(a), the shock filtering changes the smoothly increasing input signal to the step function signal. Also, it removes some noise at homogeneous regions. Figure 10(b) shows the shock filtering result.

(a)

(b)

Figure 10: (a) Input and output signals of shock filter (b) Shock filtering results(before and after)

For the next step before 3D warping of the ToF depth image, we calculate the converting function of the depth index value of the ToF depth image to the disparity value. In order to calculate this converting function, we capture the checker pattern images shown in Fig. 11 at the linearly defined position in the scene. From the background to in front of the camera, we obtain the pattern images by the multiple color and ToF depth cameras. After multi-view image rectification, we can obtain the disparity from the color image and the depth index value from the ToF depth image at each position. By using the least square method to the obtained data, we can acquire the converting function as shown in Fig. 12.

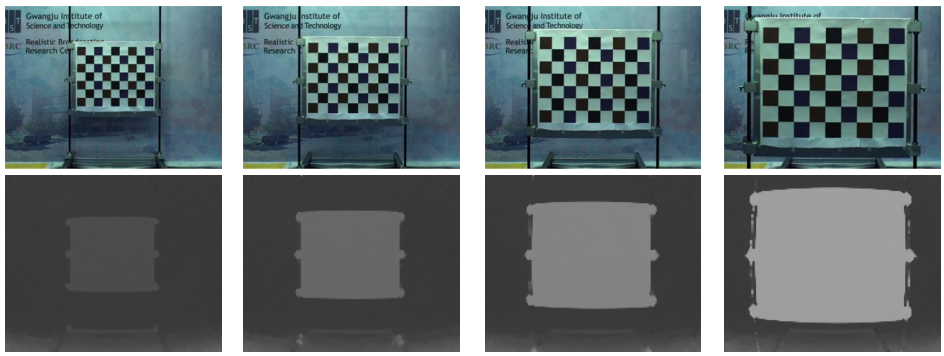


Figure 11: Checker images for depth-disparity mapping

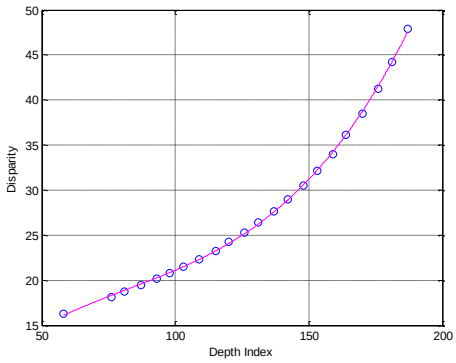


Figure 12: Converting function from the depth index to the disparity

3D warping of the ToF depth camera image is performed by using the depth index values, camera parameters, and minimum and maximum real depth values, as shown in Fig. 13. Each pixel of the ToF depth camera image is back-projected to the space, and then re-projected to the color image position. Each pixel value is converted to the disparity by using the converting function. However, due to the resolution difference, the warped ToF depth image has many holes which mean the pixels without disparity values as shown in Fig. 14.

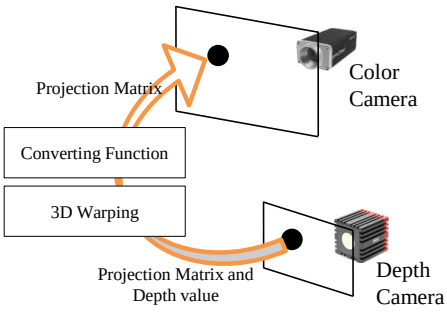


Figure 13: 3D warping of the ToF depth image

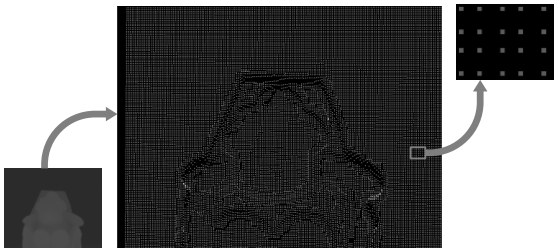


Figure 14: 3D warping result

3.7 Stereo Matching using ToF Depth Data

After all the processing explained before, we generate the 3D contents using the processed images. Actually, the captured stereo or multi-view images are also considered as 3D contents. However, to generate more viewpoints or to reconstruct 3D scenes provide more realistic 3D sense. Therefore, the depth information of the scene is essential. In order to estimate the depth of the scene, the stereo matching is used with the ToF depth information.

The warped ToF depth data is converted to the disparity, and then used as an initial disparity for the stereo matching. However, as shown in Fig. 14, there are many holes that mean no initial disparity pixels. For these pixels, we estimate the initial disparity values using the adjacent pixel values. There are three cases as shown in Fig. 15.

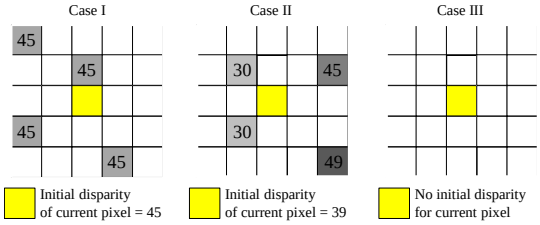


Figure 15: 3 cases for initial disparity acquisition

Case I: Homogeneous Regions

In this case, all the adjacent disparity values are the same. Therefore, the disparity of the current pixel is also the same, and we require the smaller search range for the stereo matching.

Case II: Depth Discontinuity Regions

Depth discontinuity regions have several initial disparity values. Therefore, for the current pixel, the average disparity value becomes the initial disparity, and the search range can cover from the minimum to the maximum disparities in the region.

Case III: No disparity Regions

This region could be an occluded region or image border regions. In this case, we do not have an initial disparity value, and use the full search range.

Then we design the energy function $E(f)$ for the stereo matching as Eq. 1. $D_p(f_p)$, $S(f_p, f_q)$, and $T_p(f_p)$ indicate the data cost, smoothness cost, and ToF cost, respectively. p, q, and f_p represent the current, and neighboring pixel positions, and possible disparity values in the given search range. The data term is calculated using sum of absolute difference (SAD) of the current block between the current view and the reference view. The smoothness term is calculated based on upper, lower, left, and right pixel values.

$$E(f) = \sum_{p} D_p(f_p) + \sum_{p,q} S(f_p, f_q) + \sum_{p} T_p(f_p) \tag{1}$$

The ToF term in Eq. 1 is defined as follows, where $d_{p,i}$ means the initial disparity at the pixel p. This energy function is then minimized by using belief propagation. Finally, the disparity candidate that has the minimum energy is decided to the disparity value for the current pixel.

$$T_p(f_p) = \begin{cases} |f_p - d_{p,i}| & (case\ I\ and\ II) \\ 0 & (case\ III) \end{cases} \tag{2}$$

4 Experimental Results

Figure 16 shows the multi-depth camera system. Using this camera system, we captured the color and ToF depth images for experiments. The color image resolutions are 800x600. The resolution of 800x600 reduces the difference of field of view between the color and ToF depth cameras. Figure 17 shows the second test images captured by the multi-depth camera system.

Figure 16: Camera arrangement for capturing

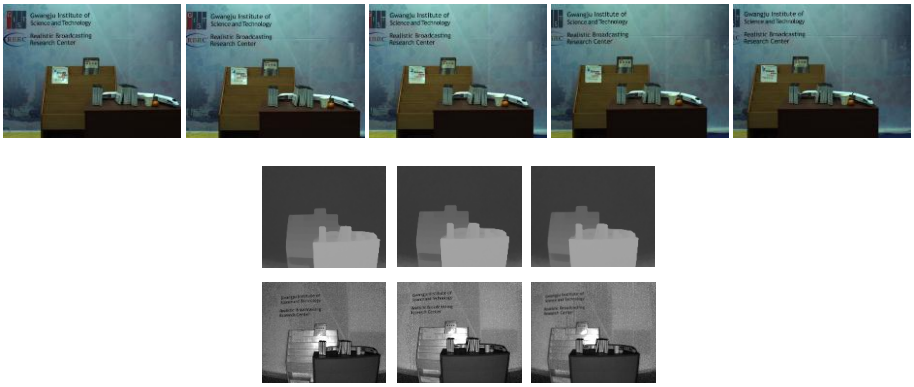
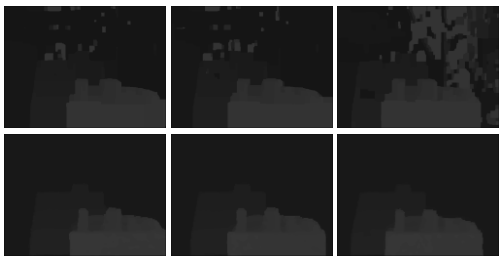


Figure 17: Captured images(the second image set)

Figure 18(a) and (b) show the depth estimation results. The upper row results are generated disparity maps by the stereo matching. The lower row results show the disparity maps by the proposed method. The results of the stereo matching show that there are inaccurate depth at background and the table since these regions are textureless. Some boundary regions also have inaccurate depth. However, the results of the proposed method provide the accurate and stable depth even at the textureless regions, since the disparity value for each pixel was calculated based on the initial disparity. Also, the adjustable disparity search range for the stereo matching according to the initial disparity case avoids that the wrong disparity value has the minimum energy for the current pixel.

(a)



(b)

Figure 18: Generated disparity maps using stereo matching (upper row) and the proposed method (lower row)

5 Conclusion

In this chapter, we explained the multiple depth estimation method using the ToF depth cameras with the color cameras. The ToF depth camera complements the passive sensor based depth acquisition method using the color images. However, the ToF depth image has noise and distortion. Therefore, it is essential to reduce them to use with the color images for depth estimation and 3D content generation. The proposed method composed of several post-processing parts and depth estimation part. For the captured color images, we perform the multi-view image rectification and color correction to reduce the geometric and photometric errors, respectively. For the ToF depth images, lens distortion correction and noise reduction are requires. Then, the ToF depth image is warped to the color image positions as the form of the disparity. The warped pixel values are used as in initial disparity for the stereo matching that considers the ToF term. Since the proposed method estimates depth maps based on the initial disparity, we obtained more accurate and stable results than the stereo matching.

Acknowledgement

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

References

Bartczak, B. & Koch, R. (2009). 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).

Felzenszwalb, P.F. & Huttenlocher, D.P. (2006). Efficient belief propagation for early vision. International Journal of Computer Vision, 70(1), 41-54.

Frick, A., Bartczack, B. and Koch, R. (2010). 3D-TV LDV content generation with a hybrid ToF-multi camera rig. Proc. of 3DTV Conference (pp. 1-4).

Gilboa, G., Sochen, N., & Zeevi, Y.Y. (2002). Regularized shock filters and complex diffusion. Lecture Notes in Computer Science, 2350, 399-313.

Gudmundsson, S.A., Aanaes, H., & Larsen, R. (2008). Fusion of stereo vision and time-of-filght imaging for improved 3D estimation. International Journal of Intelligent Systems Technologies and Applications, 5(3), 425-433.

Ho, Y.S., & Kang, Y.S. (2010). Multi-view depth generation using multi-depth camera system. Proc. of International Conference on 3D Systems and Application (pp. 1-4).

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

Jiang, G., Shao, F., Yu, M., Chen, K., and Chen, X. (2006). New color correction approach to multi-view images with region correspondence. Lecture Notes in Computer Science 4113, 1224-1228.

Joshi, N., Wilburn, B., Vaish, V., Levoy, M., & Horowitz, M. (2005). Automatic color calibration for large camera arrays, UCSD CSE Technical Report, CS2005-0821.

Kang, Y.S. & Ho, Y.S. (2010). An efficient image rectification method for parallel multi-camera arrangement. IEEE Transactions on Consumer Electronics, 57(3), 1041-1048.

Kang, Y.S. and Ho, Y.S. (2011). Disparity map generation for color image using ToF depth camera. Proc. of 3DTV Conference (pp. 1-4).

Kang, Y.S., Lee, E.K., and Ho, Y.S. (2010). Multi-depth damera system for 3D video generation. Proc. of International Workshop on Advanced Image Technology (pp. 44(1-6)).

Kuhnert, K.D. & Stommel, M. (2006). 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 (pp. 4780-4785).

Salvi, J., Fernandez, S., Pribanic, T., and Llado, X. (2010). A state of the art in structured light patterns for surface profilometry. Pattern Recognition, 43(8), 2666-2680.

Smolic, A. & Kauff, P. (2005). Interactive 3D video representation and coding technologies, Proceedings of the IEEE, Spatial Issue on Advances in Video Coding and Delivery, 93(1), 99-110.

SR4000 User Manual Version 0.1.2.2, Mesa Imaging AG, 18-19.

Sun, J., Zheng, N.N., & Shum, H.Y. (2003). Stereo matching using belief propagation. IEEE Transactions of Pattern Analysis and Machine Intelligence, 25(5), 787-800.

Wang, A., Qiu, T., & Shao, L. (2009). A simple method of radial distortion correction with centre of distortion estimation. Journal of Mathematical Imaging and Vision, 35(3), 165-172.

Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 1330-1334.

Zhu, J., Wang, L., Yang, R., & Davis, J. (2008). Fusion of time-of-flight depth and stereo for high accuracy depth maps. Proc. of IEEE Conference on Computer Vision and Pattern Recognition (pp. 231-236).

ZCam Product Data Sheet, StudioGE.