Accurate Range Image Generation Using Sensor Fusion of TOF and Stereo-based Measurement

Tomoaki Fukuda, Yonghoon Ji, and Kazunori Umeda
Department of Precision Mechanics, Chuo University,
1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan
Email: Fukuda@sensor.mech.chuo-u.ac.jp

***Abstract*— This paper proposes a system for measuring high precision range images under various conditions by combining range image sensors with different measurement principles. We utilize a stereo camera and a TOF (Time-of-Flight) range image sensor for our measuring system. Stereo cameras can measure relatively long range; however, measurement errors occur frequently. On the other hand, TOF sensors can provide relatively accurate distance measurements, but it cannot deal with under sunlight condition. In this respect, we develop a novel system that is able to generating a range image under any environment by sensor fusion. Experiment results demonstrate that the proposed system can generate the range image accurately.**

I. INTRODUCTION

In recent years, various range image sensors have been developed. These are used for an autonomous mobile robot and an UGV (Unmanned Ground Vehicle) in order to perform person tracking or obstacle recognition. Therefore, accurate range measurement is requited under various conditions. The method of range image generation can be divided into two types. First, TOF (Time-of-Flight) method calculates a distance based on the time when the light reflected on the object returns. Second method uses the shifted optical path depending on the measurement position. For example, stereo cameras are the most popular system that use two or more monocular cameras in order to obtain distances based on the principle of triangulation from parallax occurring between the images from each camera.

Depending on the measurement method, measurement may be difficult due to illumination condition or the reflection characteristic of the object [2]. Thus, a sensor fusion technology have attracted the attention of a lot of researchers given that the advantages of each sensor can be fully exploited. Mori et al. proposed a system that uses a monocular camera and ultrasonic sensors [3]. This system compensated the disadvantage of the ultrasonic sensor that cannot specify the position of the measured object in the lateral direction by using the monocular camera. However, since it uses a monochrome camera, there is a problem that it is affected by the target color for measurement. Gofuku et al. proposed a system that combines a stereo camera based on a middle focus lens and a monocular camera based on a wide angle lens with a swinging mechanism to enable a wide range measurement [4]. However, accurate measurement is difficult in case of an environment with a complex background. In this respect, we propose the highly accurate range image measurement system which is robust for change of the lighting environment and the measurement object by combining a stereo camera and a TOF range image sensor.

The reminder of this paper is as follows. Section II explains the characteristics of the sensors used and the structure of the sensor fusion system. Section III describe a method of generating range images with two sensors. Section IV describes measurement experiments with sensor fusion system. Section V describes conclusions and future prospects.

II. SENSOR CONFIGURATION

In this study, we use Microsoft Kinect v2 as the TOF range image sensor and STEREOLABS ZED as the stereo camera. ZED has relatively a long measurable distance and can be used under sunlight. However, it is unable to measure in dark environments and on uniform surfaces. On the other hand, Kinect v2 can measure even in dark environments and also on uniform surface. Furthermore, it provides relatively accurate distance information than the stereo camera. However, it cannot measure the surface that does not reflect infrared light. Moreover, the measurable distance becomes shorter under sunlight. In order to compensate for the abovementioned shortcomings of two sensors, we develop a novel sensor fusion system as shown in Fig. 1. We set fused camera coordinate system on the left lens of ZED and lens of Kinect v2 IR camera on the *Y* axis given that each range image is generated based on coordinates of a left color image of ZED and IR camera image of Kinect v2, respectively.

III. SENSOR FUSION

Figure 2 shows the flowchart of our system to capture accurate range images. After range image acquisition from each sensor, the overall process is divided into three steps and each detailed process is presented in the next subsections.

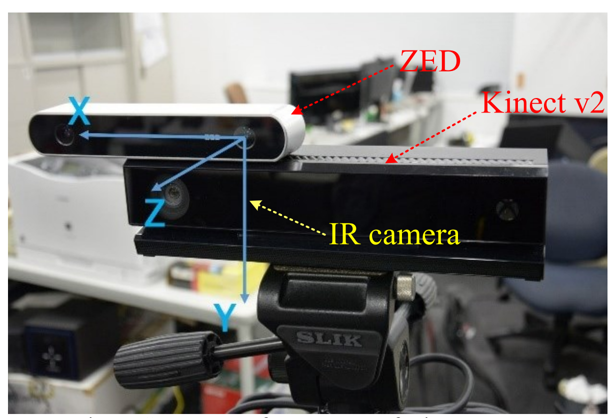


Fig. 1 Structure of our sensor fusion system.

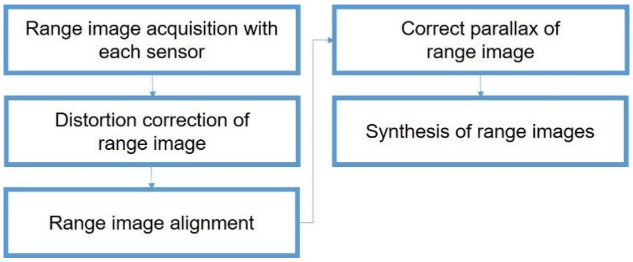
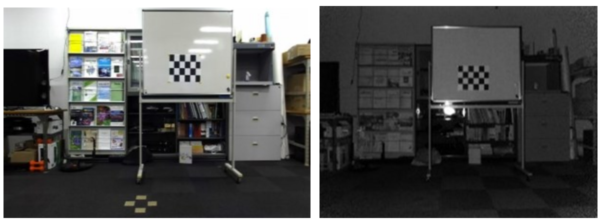


Fig. 2 Flowchart of overall process.



(a) ZED left image (b) Kinect v2 IR image

Fig. 3 Image example for alignment.

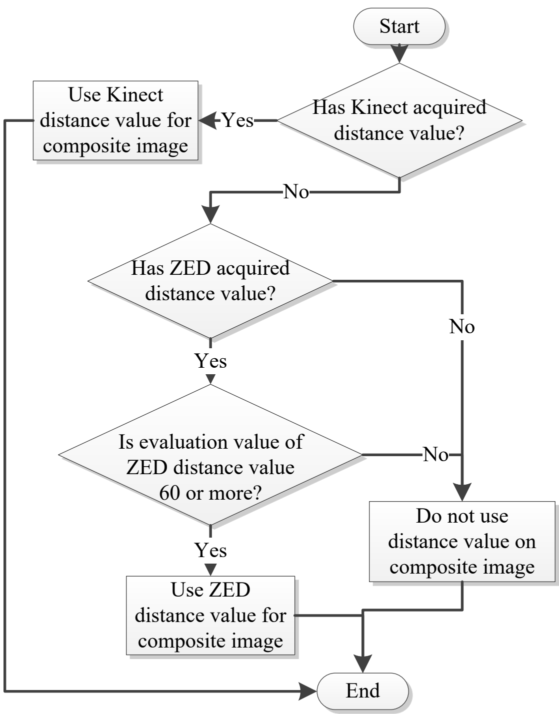


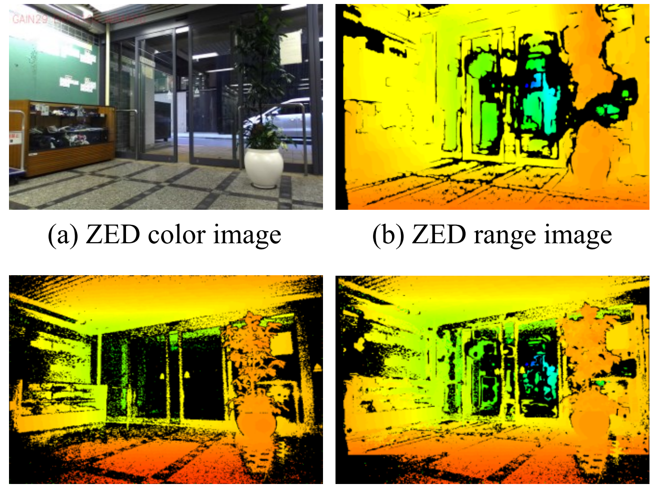
Fig. 4 Flowchart to generate synthesis range image.

A. Distortion correction

Firstly, we take multiple images including checkerboards using the left color camera of ZED and the IR camera of Kinect v2 to estimate internal parameters and distortion coefficients and use these for distortion correction of each range image.

B. Range image alignment

Alignment process of the two range images generated based on the left camera of ZED and the IR camera of Kinect v2 is performed. In this step, using two sensors, we capture many



(a) ZED color image (b) ZED range image

(c) Kinect v2 range image (d) Composite image

Fig. 5 example of the range images.

checkerboard images at the same distance according to *Z* axis from the origin of the fused camera coordinate system. An example of a set of checkerboard images is shown in Fig. 3.

C. Parallax correction of range images

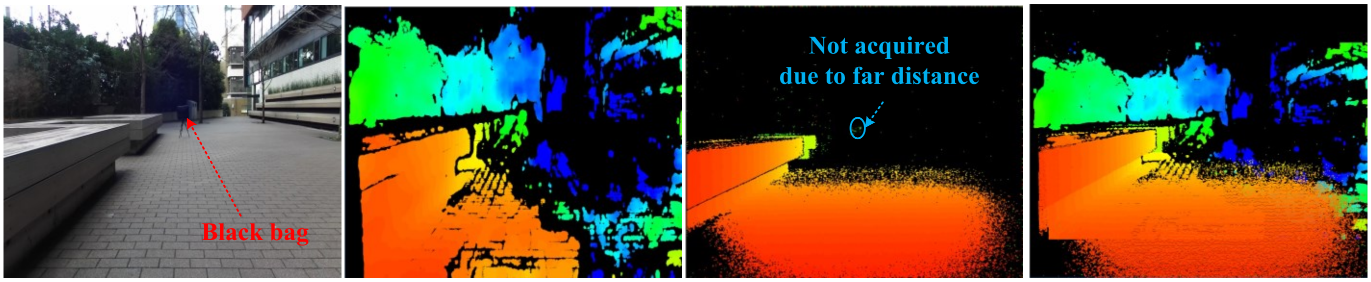
The left camera of ZED and the IR camera of Kinect v2 are located at different position on *Y* axis as mentioned in the previous section. Therefore, generally parallax occurs in range images from both sensors. After alignment, the parallax at the distance where checkerboard was taken becomes zero. The parallax Δy in the pixel units with respect to the range image of ZED is expressed as follows:

$$\Delta y = \alpha \frac{z_{cb} - z}{z_{cb} \cdot z} \qquad (1)$$

where z and z_{cb} respectively denotes a certain measured distance value from range image of Kinect v2 and the distance when capturing the checkerboard. a is a coefficient that can be obtained by calibration. In this study, the coefficient a was 49,350. Hence, we can correct parallax between two range images by applying Eq. (1) to all pixels of the range image of Kinect v2. Through the above processing, it is possible to match range image of Kinect v2 based on the ZED range image.

D. Synthesis of range images

An evaluation value expressed by zero to 100 can be obtained for each pixel on the range image from ZED. Here, higher evaluation values mean higher reliability of the distance values. Therefore, these evaluation value can be used when composing the range images. Based on our previous experiments, in general, a distance value obtained by Kinect v2 is more accurate than a distance value from ZED in case of that distance values from both sensors are obtainable for the same location. Therefore, a distance value from Kinect v2 take precedence at the location where the distance values obtained from both range images. On the other hand, when a distance value is acquired by only one side, we perform following process. If the distance is measured only by the Kinect v2, the



(a) ZED color image (b) ZED range image (c) Kinect v2 range image (d) Composite image

Fig. 6 Example of generated images in cloudy outdoor condition.

(a) ZED color image (b) ZED range image (c) Kinect v2 range image (d) Composite image

Fig. 7 Example of generated images in lightly dark indoor condition.

data is directly used for the synthesized range image. Among the measurement results from ZED, only the pixel data whose evaluation values exceed a threshold value are used for the synthesized range image and the rest are discarded as outlier. In this study, we applied the threshold value of 60. The flowchart of synthesis is shown in Fig. 4. This process is performed for all the pixels.

Examples of the range images from the above processing are shown in Fig. 4. A reliable range image (Fig. 4(d)) is generated by combining the advantages of the range images from ZED (Fig. 4(b)) and Kinect v2 (Fig. 4(c)) in an environment shown in Fig. 4(a).

IV. EXPERIMENT

We conducted experiments to verify the effectiveness of our combination method for two range image sensors under the conditions as follows. In an outdoor environment, a black bag as shown in Fig. 6(a) was used as an evaluation criterion for distance measurement. Experiments were performed under a forward light condition, a backlight condition, and a cloudy condition. Figures 6(b), (c), and (d) show examples of each range image including the black bag located eight meters from the sensor system in the cloudy condition. As shown in Fig. 6(c) Kinect v2 could not measure the distance to the black bag. ZED, meanwhile, cloud measure it correctly as shown in Fig. 6(b). We also used a checkerboard as another evaluation criterion for distance measurement under bright and lightly dark indoor conditions. Figures 7(b), (c), and (d) show examples of each range image including the checkerboard located four meters from the sensor system in the lightly dark condition. In this case, contrary to the above outdoor condition, ZED could not acquire the distance data to the checkerboard due to a matching failure while the Kinect v2 could measure it relatively accurately. Consequently, we were able to get the reliable range image from proposed sensor fusion system since if one sensor fails to obtain distance data, the other compensates the corresponding data well.

Table 1 shows experimental results in outdoor environment under forward light and backlight conditions. In case of the forward light condition, ZED was able to acquire distance data reliably. However, Kinect v2 could not measure the most distance data. On the other hand, opposite tendency occurred with respect to short measurement distance in case of the backlight condition. Finally, the results of our sensor fusion system show that two sensor data were appropriately merged depending on each of conditions.

Table 2 shows experimental results in indoor environment under bright and lightly dark conditions. In this case, sensor fusion results are almost same to the results from Kinect v2 because, in our proposed framework, we determined that data from Kinect v2 are more accurate if the data are taken from both sensors at the same location.

V. CONCLUSIONS

In this paper, we proposed a new sensor fusion method that combines two range image sensors with different characteristics in order to acquire reliable and high precision range images under most lighting conditions. The validity of the proposed scheme of range image generation was investigated through a substantial number of experiments, both in indoor and outdoor environments with target objects.

The future work related to this study is to improve the accuracy of range measurement by taking full advantage of the two sensors. For example, depending on illumination changes or measurement target, we can consider to calculate weighted average when data can be taken from both sensors.

Table 1 Experimental results in outdoor environment using black bag as measurement target

	ZED		Kinect v2		Sensor fusion system	
True value [mm]	Forward light outdoor	Backlight outdoor	Forward light outdoor	Backlight outdoor	Forward light outdoor	Backlight outdoor
1,000	1,031	N/A	1,066	1,038	1,066	1,038
2,000	1,962	N/A	N/A	2,029	1,962	2,029
3,000	2,965	3,007	N/A	3,024	2,965	3,024
4,000	3,956	4,120	N/A	4,005	3,956	4,005
5,000	4,973	5,329	N/A	5,031	4,973	5,031
6,000	6,018	N/A	N/A	N/A	6,018	N/A
7,000	7,178	N/A	N/A	N/A	7,178	N/A
8,000	8,203	N/A	N/A	N/A	8,203	N/A

Table 2 Experimental results in indoor environment using checkerboard as measurement target

	ZED		Kinect		Sensor fusion system	
True value [mm]	Bright indoor	Lightly dark indoor	Bright indoor	Lightly dark indoor	Bright indoor	Lightly dark indoor
1,000	994	987	1,008	1,013	1,008	1,013
2,000	1,964	1,945	2,015	2,016	2,015	2,016
3,000	2,881	2,883	3,007	3,012	3,007	3,012
4,000	3,761	N/A	4,011	4,017	4,011	4,017
5,000	4,767	4,724	5,017	5,026	5,017	5,026
6,000	5,533	5,573	6,021	6,022	6,021	6,022
7,000	6,311	6,433	7,030	7,045	7,030	7,045
8,000	7,134	7,183	7,985	N/A	7,985	N/A

REFERENCES

[1] F. Blais, "Review of 20 years of range sensor development," *Journal of Electronic Imaging*, Vol. 13, No.1, pp.231-240, 2004.

[2] Y. Isobe, G. Masuyama, and K. Umeda, "Target tracking for a mobile robot with a stereo camera considering illumination changes," *Proceedings of the 2015 IEEE/SICE International Symposium on System Integration*, pp. 704-707, 2015.

[3] T. Mori, H. Fukuda, K. Kamei, and K. Inoue, "An Environment Modeling Method by Sensor Fusion for an Indoor Mobile Robot," *IEEJ Transactions on Electronics, Information and Systems*, Vol. 114, No. 5, pp. 603-608, 1994 (in Japanese).

[4] A. Gofuku , Y. Kiyoi , I. Nagai, and Y. Tanaka, "Development of fusion module of the images from wide view field and stereo cameras for an autonomous running vehicle," *Transactions of the JSME*, Vol. 70, No. 693, pp. 1371-1379, 2004 (in Japanese).