MECATRONICS - REM 2016
June 15-17, 2016

High-speed Three-dimensional Mapping by Direct Estimation of a Small Motion Using Range Images

Shinta Nozaki and Masashi Kimura
School of Science and Engineering
Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan
Email: {nozaki, kimura}@sensor.mech.chuo-u.ac.jp

Gakuto Masuyama and Kazunori Umeda
Faculty of Science and Engineering
Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan
Email: {masuyama, umeda}@mech.chuo-u.ac.jp

***Abstract*—In this paper, a high-speed three-dimensional map generation method using direct estimation of the motion parameters of the sensor is proposed. Range images are aligned between frames after estimating the motion parameters of the sensor using the relationship of the range image and the motion parameters of a small motion. A 3D map can be created quickly because the method does not need to obtain the correspondences of features. Both the processing time required for the motion parameter estimation and the accuracy of the map are verified by experiments using a developed RGB-D sensor.**

I. INTRODUCTION

It is important for an autonomous robot to use a three-dimensional (3D) map to move efficiently. A robot understands its position by using a 3D map.

A 3D map is often created by aligning the features between multiple range images. Recently, many methods have been proposed for the alignment of range images.

Several studies [1], [2], [3], [4] are based on a Iterative Closest Point algorithm[5], [6]. Some studies use color images[7], [8]. And there are some studies using both range and color images[9], [10], [11]. Range images and color images used for the 3D mapping are captured by a range image sensor or an RGB-D sensor. In general, the measurement speed of these sensors is 30∼60 frames per second (fps). When the displacement between frames is large, the alignment of the range images is difficult. Therefore, if the measurement speed of the sensor is low, the movement speed of the sensor is limited. We have proposed a 3D mapping method[12] using a high-speed RGB-D sensor[13]. The RGB-D sensor can capture a range image and a color image at 200 fps. This method aligns range images by obtaining correspondences of the time series data using the features of the range image and the color image. As a result, it is possible to generate 3D maps of a variety of environments without limiting the movement speed of the sensor. However, there was a problem that the calculation cost for searching the corresponding points is large. The processing time is 200∼300 ms per frame. This makes the method difficult to use for creating a 3D map online. Therefore, we propose a 3D mapping system that will align the range images without searching for the corresponding points. The system is based on the method called a direct method[14], [15]. The direct method can estimate the motion of a sensor by using the relation of the range image and the motion parameters of a small motion. The proposed method aligns range images by using the motion parameters of a sensor, which will shorten the processing time significantly.

In Section II, we will give an overview of this method; in Section III, we will describe the RGB-D sensor; and in Section IV, we will explain the direct method. In addition, we will describe the experiments designed to verify the usefulness of the proposed method in Section V. Finally, in Section VI, we will conclude this paper.

II. OVERVIEW OF THE 3D MAP GENERATION

A flow chart of the proposed method is presented in Fig.1. In this method, a color image and a range image are obtained by using a high-speed RGB-D sensor. Then, outliers are removed by fitting the least-squares plane and setting up equations about the relation of the range image and the motion parameters on the small motion based on 3D information. Hereafter, in accordance with the references, we call the equation a *range-motion equation*. A 3D map is created by aligning the range images between two frames using the motion parameters of the sensor. An explanation of the *range-motion equation* and the removal of outliers by fitting the least-squares plane is given in Section IV.

III. RGB-D SENSOR

The RGB-D sensor will be explained in this section. The appearance of the RGB-D sensor is shown in Fig.2. Using an

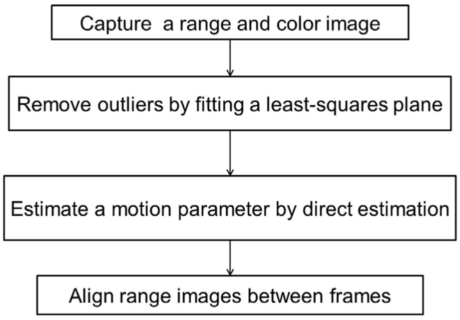


Fig. 1. 3D mapping method

978-1-5090-1787-4/16/$31.00 ©2016 IEEE

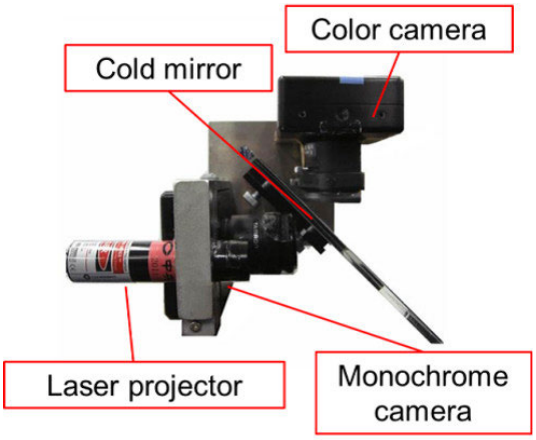


Fig. 2. RGB-D sensor

active stereo method, the sensor can capture a range image with 361 measurement points and color VGA (640×480) images at 200 fps. The sensor captures the range image using a monochrome CCD camera and a laser projector that projects 361 spots of Infrared (IR) light. The sensor captures the color image using a color CCD camera. The sensor coaxially captures a color image and a range image using a cold mirror that passes IR light and reflects visible light. Due to this characteristic, when the camera is positioned as shown, the visible light is observed by the color camera, and the IR light is observed by a monochrome camera. Therefore, the two cameras can observe the same scene. The measurement range of the sensor is 900 mm∼2500 mm. The sensor can also measure outdoors. This is because the laser spots are bright with single wavelength and disturbance lights are eliminated by using two optical filters (band-pass and low-pass) .

IV. MOTION PARAMETER ESTIMATION OF THE SENSOR

In this section, we will describe the method for removing outliers and the *range-motion equation*.

A. Range-motion equation[14]

As shown in Fig.3, r indicates the measurement distance, and $\mathbf{t}$ indicates the measurement direction. We assume that the measurement direction is unchanged and known, which is usually satisfied when an active range image sensor, such as an active stereo camera or a Time-of-flight (TOF) range image sensor, is adopted. In addition, $\mathbf{n}$ indicates the normal unit vector of the measurement point, and $\dot{r}$ indicates the rate of change of the distance. $\mathbf{v_0}$ indicates the translation velocity vector of the measurement point, and $\boldsymbol{\omega}$ is the rotational velocity vector whose origin is located in the sensor's position. When we assume that the motion of the sensor is small, the following equation holds:

$$\mathbf{n}^T\mathbf{v_0} + r(\mathbf{t}\times\mathbf{n})^T\boldsymbol{\omega} = \dot{r}(\mathbf{n}^T\mathbf{t}). \tag{1}$$

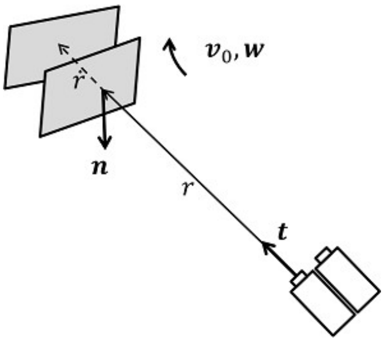


Fig. 3. Parameters of the range-motion equation

Eq.(1) has an unknown quantity that indicates a translation velocity vector and a rotational velocity vector. The equation is called a *range-motion equation*. A *range-motion equation* is obtained for each measurement point that satisfies the assumptions given in IV-B. The following equations hold simultaneously.

$$\mathbf{Ax} = \mathbf{y} \tag{2}$$

$$\boldsymbol{A} = \begin{bmatrix} \frac{\mathbf{n_1}^T}{\mathbf{n_1}^T\mathbf{t_1}} & \frac{r_1(\mathbf{t_1}\times\mathbf{n_1})^T}{\mathbf{n_1}^T\mathbf{t_1}} \\ \vdots & \vdots \\ \frac{\mathbf{n_m}^T}{\mathbf{n_m}^T\mathbf{t_m}} & \frac{r_m(\mathbf{t_m}\times\mathbf{n_m})^T}{\mathbf{n_m}^T\mathbf{t_m}} \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} \dot{r}_1 \\ \vdots \\ \dot{r}_m \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{v}_{\mathbf{0}}^T \\ \boldsymbol{\omega}^T \end{bmatrix}$$

where $\boldsymbol{x}$ is the motion parameter.

The maximum likelihood estimate of the motion parameters can be calculated by using the pseudo-inverse matrix as follows:

$$\hat{\mathbf{x}} = \mathbf{A}^{+}\mathbf{y} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} \tag{3}$$

Alignment of the range images between frames is accomplished using this motion parameter.

B. Outlier removal by the least-squares plane fitting

The direct method calculates the motion parameters of the sensor by assuming several conditions. They are indicated as follows:

1) Movement between the frames is small enough.
2) The measurement direction is known and unchanged.
3) The measurement points of a measurement direction exist on the same smooth surfaces before and after movement.

Condition 1 and 2 are already satisfied because the sensor used in this study can capture at high speed (200 fps) using an active stereo method. However, in some cases, condition 3 will not be satisfied for all measurement points. Therefore, if the measurement point is not on a locally smooth surface, it is necessary to remove the outlier.

In this method, we create a least-squares plane by using a measurement point and its surrounding points as shown in

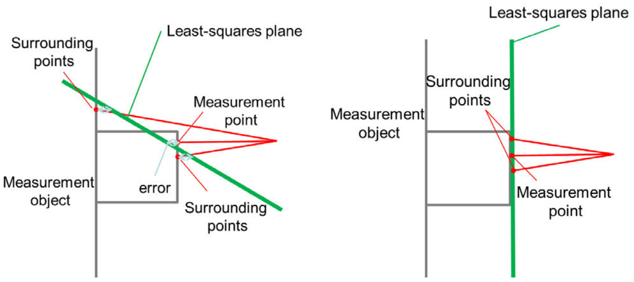


Fig. 4. Fitting a least-squares plane. Left: with outlier, right: without outlier

TABLE I
LENGTH COMPARISON

	L1 [mm]	L2 [mm]
Real object	161	161
3D map	155	157

Fig.4. If the average value of the error between the points and the plane is greater than the threshold, the measurement point is removed as an outlier. If the measurement point and its surrounding points exist on the same smooth surface as shown in Fig.4 right, and the average value of the error between the points and the plane is lower than the threshold, we use it to estimate the motion parameters of the sensor. Additionally, a normal vector of the least-squares plane is used to estimate the motion parameter.

Note that the smooth surface is not necessary to be a plane, as long as it is locally planar enough to measure a normal vector. In addition, the change of the normal vector, that occurs mainly by the relative rotation of the measurement object, is acceptable as long as the change is small, that can be satisfied by the high-speed measurement.

V. EXPERIMENT

A. *Evaluation of accuracy*

1) Experimental environment: We verified the accuracy of the 3D map that was created using the proposed method. The range images and the color images were captured using the RGB-D sensor held by a person. The 3D map was created offline. The object to be measured was a cube-shaped box as shown in Fig.5. Accuracy was verified by comparing the length of the cube's sides on the 3D map with that in the real environment. The comparison parts are labeled L1 and L2 in Fig.6. The number of range images is 230 frames.

2) Experimental result: Fig.7 shows the measurement results of Fig.5. Table I shows the length of each side of the box.

The error of the sides of the cube is less than 6 mm, and the results of the experiment show that the alignment error is kept low. Therefore, it can be said that the proposed method will be able to create an accurate 3D map in an environment that includes sufficiently smooth surfaces.

Fig. 5. The measurement object for evaluating accuracy: a cube-shaped box

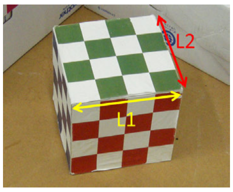


Fig. 6. Parts to compare length

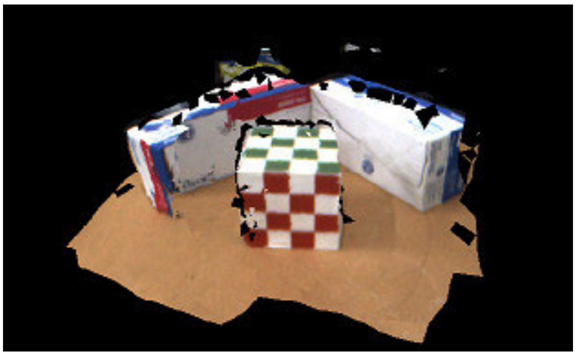


Fig. 7. Constructed 3D map containing a cube-shaped box

B. *3D map generation*

1) Experimental environment: We generated a 3D map in the two environments using the proposed method. The range images and the color images were captured by using the RGB-D sensor held by a person. The 3D map was created offline.

One of the measurement environments was a scene containing smooth planes as shown in Fig.8. Another measurement environment was a scene with only compact objects as shown in Fig.9. The compact objects were 50∼60 mm wide. This is approximately three times the interval of the measurement points. The number of range images for Fig.8 and 9 is 340 and 170 frames respectively.

2) Experimental result: Fig.10 and 11 show the results of measuring the objects in Fig.8. Fig.10 (a) and (b) show the process of generating the map, Fig.10 (c) shows the completed map, and Fig.11 shows the results of applying texture mapping. Fig. 12 shows the results of measuring the objects of Fig.9. Fig.12 (a) shows the constructed map, and Fig.12 (b) shows the result of applying texture mapping.

Fig. 8. The measurement scene containing smooth planes

Fig. 9. The measurement scene with only small objects

TABLE II
ESTIMATION TIME OF THE MOTION PARAMETER

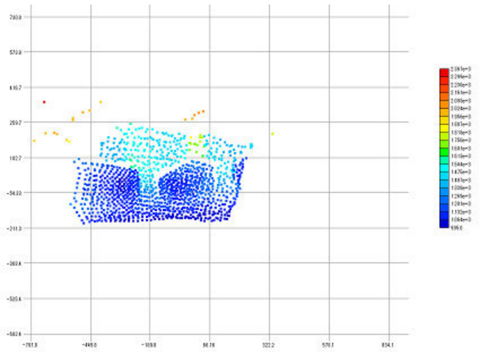
Scene	Fig.8	Fig.9
Time [ms]	0.55	0.76

In addition, Table II shows the average processing time for estimating the motion parameters.

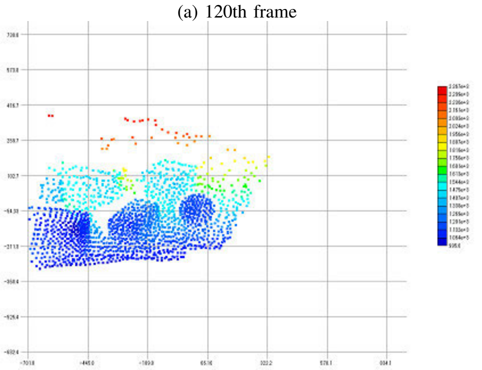
The experiments have shown that the proposed method can create an accurate 3D map in an environment that includes sufficiently smooth surfaces. However, the proposed method cannot create a 3D map in the scene of Fig.9. In this method, a measurement point and its surrounding points must exist on the same smooth surface. However, the condition is not satisfied when the small objects are measured. In this case, the motion parameter of the sensor cannot be estimated because there are few measurement points to estimate the motion parameters. Additionally, Table II shows that the processing time of the motion parameters estimation is within 0.8 ms. This processing time is much shorter than this sensor's sampling time (5ms). Therefore, it can be said that the proposed method can create a 3D map online.

C. Online measurement

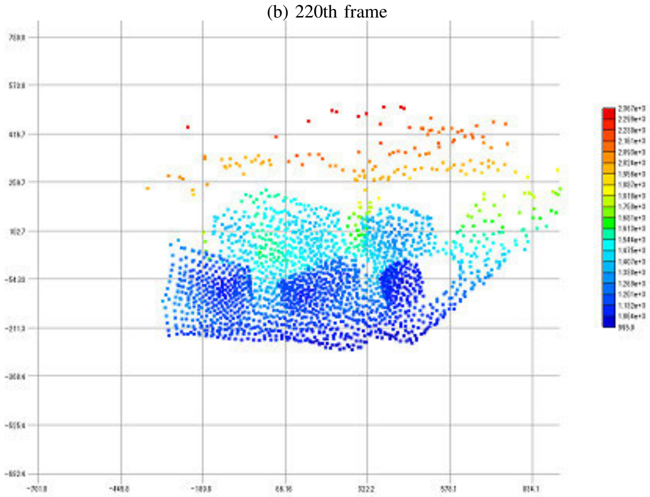
1) Experimental environment: In this section, we verify whether the proposed method enables the online mapping using the high-speed sensor. The object to be measured was cube-shaped boxes as shown in Fig.13.



(a) 120th frame



(b) 220th frame



(c) 340th frame

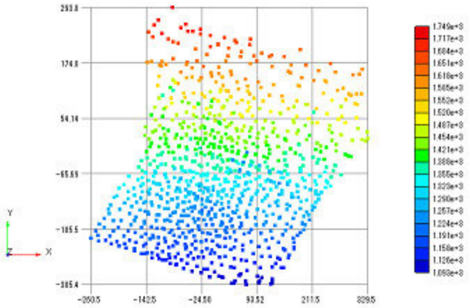
Fig. 10. Experimental results to construct 3D map for Fig.8

2) Experimental result: Fig.14 shows the measurement situation. Fig.15 shows the process of generating the map, and Fig.16 shows the completed map. The experiment has shown that the proposed method can generate the 3D map online, even if we use the sensor capturing range image at 200fps.

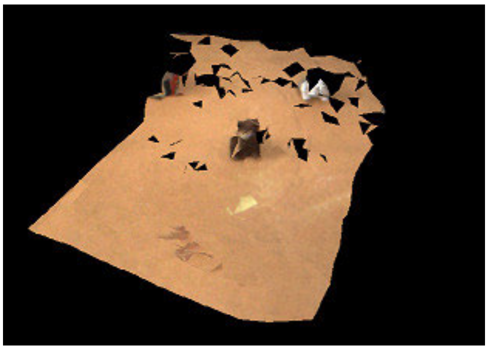
D. Application to Kinect for Windows v2

We apply the proposed mapping method to Microsoft Kinect for Windows v2, that has a wider field of view and can capture higher resolution range images than the constructed RGB-D sensor, to verify the versatility of the proposed method.

Fig. 11. Constructed 3D map for Fig.8 with texture mapping



(a) Constructed 3D map



(b) Constructed 3D map with texture mapping

Fig. 12. Experimental results to construct 3D map for Fig.9

1) Experimental environment: Fig.17 (a) shows the experimental scene. The Kinect v2 was held by a hand and rotated manually around the axis perpendicular to the floor about 65 degrees. The number of images was 85 frames. 3×3 points with the interval of every 3 point were used for the plane fitting to obtain the normal.

2) Experimental result: Fig.17 (b), (c) show the obtained map. It can be seen that a 3D map for a wide scene where several objects exist in a mess is generated accurately. One reason an accurate mapping is achieved is that the resolution of the Kinect v2 is very high, i.e., 512×424 pixels, and thus normals can be obtained accurately even for small shapes. Another reason is that the variety of the measurement direction $\mathbf{t}$ becomes larger by the sensor's wide field of view, that makes the condition to solve eq.(2) better.

Fig. 13. The measurement scene with cube-shaped boxes

Fig. 14. The measurement situation

The processing time of the motion parameters estimation was 332 ms, that is much larger than the results in Table II. This is caused by the large number of equations, i.e., more than 200,000, due to the high resolution of the Kinect v2. Appropriate sampling and parallel processing using GPU will be effective to shorten the processing time in the future.

VI. CONCLUSION

We have proposed a 3D map-generation method that uses the direct method. We have shown that the processing time for creating a 3D map can be greatly reduced using the proposed method, and online 3D map generation is possible. Motion parameter was estimated in less than 0.8 ms (using the developed RGB-D sensor), which is much faster than the existing registration studies. Although the accuracy of mapping (i.e., accuracy of motion parameter) is not good, we believe that the fast motion estimation is effective for several applications especially in robotics.

In the future, we are planning to construct a method that can create a map of the environment in which the constraints cannot be obtained correctly by adding color information.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 26330205.

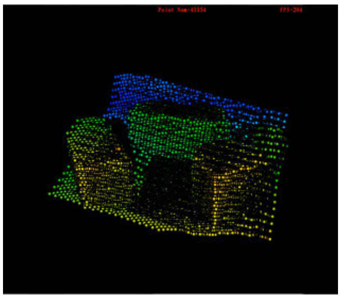


Fig. 15. The mapping process

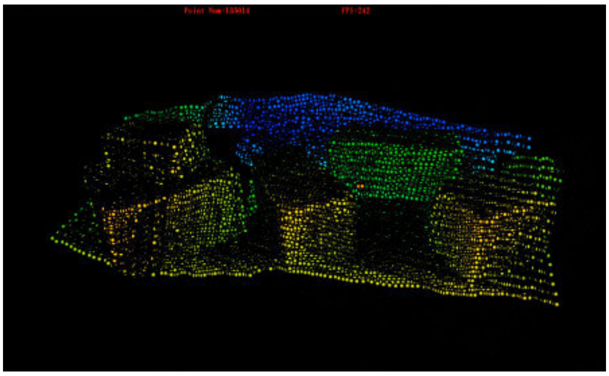
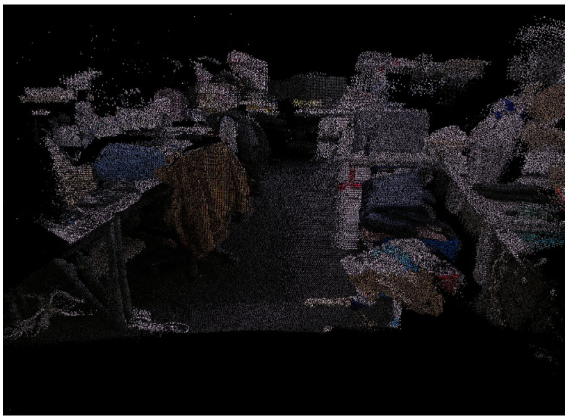
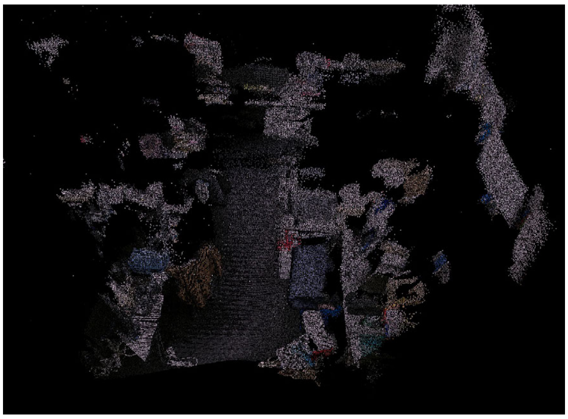


Fig. 16. Constructed 3D map for Fig.13

(a) Experimental scene for mapping



(b) Constructed 3D map: front close-up view



(c) Constructed 3D map: top view

Fig. 17. Experimental results to construct 3D map using a Kinect v2

REFERENCES

[1] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fizgibbon, KinectFusion: Real-time Dense Surface Mapping and Tracking, Proc. of the 2011 10th IEEE Int. Symposium on Mixed and Augmented Reality, ISMAR'11, (Washington, DC, USA), pp. 127-136, 2011.

[2] S. Rusinkiewicz, and M. Levoy, Efficent variant of the ICP algorithm, Proc. of the 3rd International Conference on 3-D Digital Imaging and Modeling, pp.145-152, 2001.

[3] D. Hahnel, S. Thrun, and W. Burgard, An Extension of the ICP Algorithm for Modeling Nonrigid Objects with Mobile Robots, Proc. of the International Joint Conference on Artificial Intelligence (IJCAI), 2003.

[4] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments, Int. J. of Robotics Research, vol.31, no.5, pp.647-663, 2012.

[5] P. J. Besl, and N. D. Mckay, A Method for Registration of 3-D Shapes, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.14, no.2, pp.239-256, 1992.

[6] Y. Chen, and G. Medioni, Object modelling by registration of multiple range images, Image and Vision Computing, vol.10, no.3, pp.145-155, 1992.

[7] S. Kagami, Y. Takaoka, Y. Kida, K. Nishiwaki, and T. Kanade, Online dense local 3D world reconstruction from stereo image sequences, Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2005), pp. 3858-3863, 2005.

[8] M. Tomono, Robust 3D SLAM with a Srereo Camera based on an Edge-Point ICP Algorithm, Proc. of ICRA2009, pp.4306-4311, 2009.

[9] G. Godin, D. Laurendeau, and R. Bergevin, A method for the registration of attributed range images, Proc. of 2001 Third International Conference on 3-D Digital Imaging and Modeling (3DIM), pp.179-186, 2001.

[10] M. Paton, M., and J. Kosecka, Adaptive RGB-D Localization," Proc. of 2012 Ninth Conference on Computer and Robot Vision (CRV), pp.24-31, 2012.

[11] I. Dryanovski, C. Jaramillo, and J. Xiao, Incremental registration of RGB-D images, Proc. of 2012 IEEE International Conference on Robotics and Automation (ICRA), pp.1685-1690, 2012.

[12] Y. Uchida, S. Nozaki, G. Masuyama, and K. Umeda, 3D Mapping Using an RGB-D Sensor, Proc. of 20th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV2014), pp.27-32, 2014.

[13] Y. Uchida, S. Nozaki, K. Terabayashi, and K. Umeda,, Construction of a coaxial textured range image sensing system, Proc. of 11th Interational Conference on Quality Control by Artificial Vision (QCAV2013), pp.168-174, 2013.

[14] K. Umeda, and T. Arai, Measurement of 3D Motion Parameters from Range Images, Proc. of IEEE/RSJ Int. Workshop on Intellingent Robots and Systems '91, pp.383-388, 1991.

[15] M. Yamamoto, P. Boulenger, J. A. Beralin, and M. Rioux, Direct estimation of range flow on deformable shape from a video rate range camera, IEEE Trans. Patt. Anal. Machine Intell., vol. 15, no. 1, pp.82-89, 1993.