Proceedings of the 2016 IEEE/SICE International Symposium on System Integration, Sapporo Convention Center, Sapporo, Japan, December 13-15, 2016

Three-dimensional mapping by direct estimation of a small motion using range images and optical flow of color images*

Masashi Kimura[1], Gakuto Masuyama[2] and Kazunori Umeda[2]

***Abstract*— In this paper, a method that directly estimates motion parameters of a range image sensor using range images and optical flow of color images is proposed. Linear equations for motion parameters are introduced by utilizing optical flow. A three-dimensional (3D) map is constructed by registration of range images between frames using the estimated motion parameters. Experiments to construct a 3D map using the proposed method are given. In the experiments, the accuracy of the constructed map and the time of estimating the motion parameters are evaluated. It is shown that the proposed method can construct an accurate 3D map for several scenes and estimation time of motion parameters is less than 30 ms.**

I. INTRODUCTION

Autonomous mobile robots are expected to be used in many applications, such as assistance of care for elderly people and working in disaster sites. In such applications, obstacles are expected to exist. Robots need to localize self-position, avoid the obstacles and move efficiently. It is important to construct a three-dimensional (3D) map of the surroundings for the self-localization of robots.

Generally, a sensor that can capture range images and color images at the same time is used to construct a 3D map for mobile robots. A 3D map is often constructed by registration of range images using color images. Some studies use the methods of tracking corner point [1] and edge points of color images [2]. There are also some studies using both color images and range images [3], [4], [5], [6].

We constructed a high-speed RGB-D sensor that can capture range images and color images at as fast as 200 fps [7], and proposed a 3D mapping method [8]. The method associates time-series data of both color and range images captured by the sensor, and aligns the data. The issue of the method is that it cannot construct a 3D map on-line because the processing time of the 3D mapping is 200∼300 ms. To improve the processing speed, we have proposed a method of registration of range images after motion parameters of the sensor are estimated by using an equation about range images and the parameters [9]. The equation is called *range-motion equation*. By using this method, it is possible to capture the images and construct a 3D map in a short time by estimating the parameters in only 5 ms. However, the scenes in which the method is used are limited since some planes with different normals are necessary. In this paper, we propose a new 3D mapping method that eases the limits and is applicable to more variety of scenes. The method directly estimates the motion parameters using range images and optical flow of color images.

*This work was supported by JSPS KAKENHI Grant Number 26330205.

[1]Masashi Kimura are with the School of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan kimura@sensor.mech.chuo-u.ac.jp

[2]Gakuto Masuyama and Kasunori Umeda are with Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan {masuyama, umeda}@mech.chuo-u.ac.jp

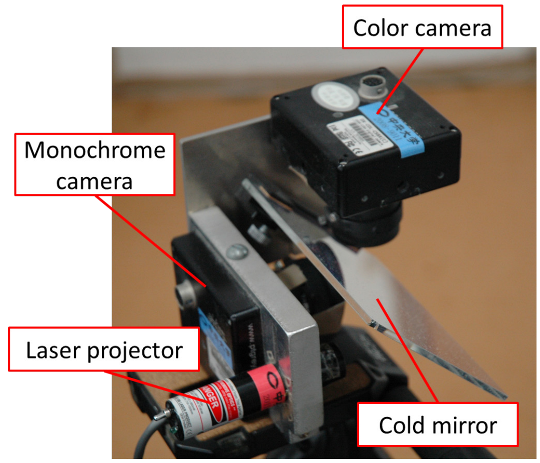


Fig. 1. RGB-D sensor for experiments

II. RGB-D SENSOR

In this section, we show the RGB-D sensor that we constructed and use for experiments. The appearance of the RGB-D sensor is shown in Fig.1.

The sensor can capture range images with 19×19 measurement points and color images with the resolution of 640×480 pixels at 200 fps. The sensor uses an active stereo method. The sensor captures range images using the monochrome CCD camera and a laser projector that projects 19×19 multi-spots infrared lights. The sensor captures color images using the color CCD camera. The sensor coaxially captures range images and color images by a cold mirror. The cold mirror passes infrared light and reflects visible light. In Fig.1, infrared light is observed by the monochrome CCD camera and visible light is observed by the color CCD camera. In addition, both cameras can observe the same scene without disparity. Consequently, aligned range images and color images are obtained simultaneously The measurement range of the sensor is 900∼2500 mm.

Note that the proposed method in this paper is not specialized in the sensor.

III. ESTIMATION OF MOTION PARAMETERS OF THE SENSOR

In this section, we propose a method to estimate motion parameters of the sensor. When motion parameters are estimated, we can align multiple range images with color images and construct a 3D map.

A. Estimation of optical flow of color images

We utilize optical flow for estimation of the motion parameters. Optical flow is obtained as movement of feature points between frames of color images. We use KLT Tracker [10] to extract feature points in the images. KLT Tracker is a standard method to track feature points of images, and can track feature points fast. On the other hand, it has a weak point that feature points with huge changes in color images are often failed to track. The sensor shown in Fig.1 can capture images fast, at 200 fps. Therefore, KLT Tracker can be adopted.

B. Estimation of motion parameters using optical flow of color images

In this section, we introduce linear equations for motion parameters by utilizing optical flow.

First, we define some variables as follows. $\mathbf{X} = (X, Y, Z)^T$ indicates the 3D coordinate of a measurement point. $\mathbf{V} = (V_x, V_y, V_z)$ indicates the motion velocity vector at the measurement point $\mathbf{X}$. $\mathbf{v_0} = (v_{0x}, v_{0y}, v_{0z})$ indicates the translation velocity vector. $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ indicates the rotational velocity vector, whose origin is assumed to be located at the position of sensor. We assume that the motion of the sensor is small. Then, the following equation holds:

$$\mathbf{V} = \mathbf{v_0} + \boldsymbol{\omega} \times \mathbf{X} \tag{1}$$

Suppose $\mathbf{X}$ is projected to $(u, v)^T$ in the color image. The relation between $(u, v)^T$ and $\mathbf{X}$ is given as follows.

$$u = \frac{X}{Z}\alpha_u + c_u \tag{2}$$

$$v = \frac{Y}{Z}\alpha_v + c_v \tag{3}$$

$$\alpha_u = \frac{f}{\delta_x} \tag{4}$$

$$\alpha_v = \frac{f}{\delta_y} \tag{5}$$

f indicates the focal length of the sensor, δ_x and δ_y indicate the pixel size of x-axis and y-axis direction, and $(c_u, c_v)^T$ indicate the image center.

We assume that a point moved from $(X, Y, Z)^T$ to $(X + V_x\Delta t, Y + V_y\Delta t, Z + V_z\Delta t)^T$ in time Δt, and the corresponding point in the color images moved from $(u_1, v_1)^T$ to $(u_2, v_2)^T$ as shown in Fig.2.

We define the optical flow between the two frames as $(\Delta u, \Delta v)^T$. Using eq.(2) and (3), the components of the optical flow are given as follows:

$$\Delta u = u_2 - u_1 = \frac{X + V_x\Delta t}{Z + V_z\Delta t}\alpha_u - \frac{X}{Z}\alpha_u \tag{6}$$

$$\Delta v = v_2 - v_1 = \frac{Y + V_y\Delta t}{Z + V_z\Delta t}\alpha_v - \frac{Y}{Z}\alpha_v \tag{7}$$

By substituting eq.(1) for eq.(6) and (7), the following equations are obtained:

$$\begin{aligned} Zv_{0x} - (X + Z\Delta u')v_{0z} - (X + Z\Delta u')Y\omega_x \\ +(X^2 + Z^2 + XZ\Delta u')\omega_y - YZ\omega_z = Z^2\Delta u' \end{aligned} \tag{8}$$

$$\begin{aligned} Zv_{0y} - (Y + Z\Delta v')v_{0z} - (Y^2 + Z^2 + YZ\Delta v')\omega_x \\ +(Y + Z\Delta v')X\omega_y + XZ\omega_z = Z^2\Delta v' \end{aligned} \tag{9}$$

$$\Delta u' = \frac{\Delta u}{\alpha_u} \tag{10}$$

$$\Delta v' = \frac{\Delta v}{\alpha_v} \tag{11}$$

Eq.(8) and (9) are linear equations for unknown motion parameters: translation velocity vector $\mathbf{v_0}$ and rotational velocity vector $\boldsymbol{\omega}$. The motion parameters of the sensor can be estimated by solving simultaneous equations that are formed by obtaining eq.(8) and (9) at three or more points.

The optical flow that is used in eq.(8) and (9) is obtained by KLT Tracker. KLT Tracker often produces outliers of optical flow, that should be excluded to estimate motion parameters accurately. Therefore, we apply RANSAC to the estimation of motion parameters to reduce the influence of the outliers.

C. Acquisition of 3D coordinates of a feature point by interpolation

Eq.(8) and (9) include 3D coordinates of a feature point. The tracked feature point of KLT Tracker does not contain 3D coordinates, so it is necessary to interpolate the neighboring points with known 3D coordinates.

Range images and color images are given in different coordinate systems. The measurement points are projected in color images after range images are applied the affine transform to associate the range images and color images. The triangle that includes the feature point inside is obtained by choosing three neighbor measurement points as shown in Fig.3.

$\mathbf{P_0}$, $\mathbf{P_1}$ and $\mathbf{P_2}$ indicate the measurement points used for the interpolation. In the image coordinates, $\mathbf{Q_1}$ indicates

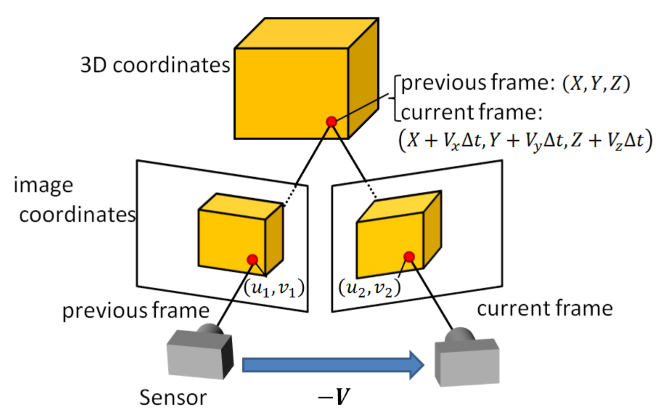


Fig. 2. Relation of 3D and 2D coordinates

intersection of side $\mathbf{P_0P_1}$ and a line parallel to side $\mathbf{P_0P_2}$ through feature point $\mathbf{X}$. $\mathbf{Q_1}$ divides $\mathbf{P_0P_1}$ internally at m to $1-m$. Similarly, $\mathbf{Q_2}$ indicates intersection of side $\mathbf{P_0P_2}$ and a line parallel to side $\mathbf{P_0P_1}$ through feature point $\mathbf{X}$, and $\mathbf{Q_2}$ divides $\mathbf{P_0P_2}$ internally at n to $1-n$. By using the internal ratios m and n, the 3D coordinate of the feature point $\mathbf{X}$ is given as follows by interpolation.

$$\mathbf{X} = (1-m-n)\mathbf{P_0} + m\mathbf{P_1} + n\mathbf{P_2} \tag{12}$$

D. Comparison of the proposed method and the previous method

In this section, we compare the proposed method and the method using *range-motion equation* [9].

Range-motion equation is given as follows.

$$\mathbf{n}^T\mathbf{v_0} + r(\mathbf{t}\times\mathbf{n})^T\boldsymbol{\omega} = \dot{r}(\mathbf{n}^T\mathbf{t}) \tag{13}$$

$\mathbf{n}$, $\mathbf{t}$ and r indicate the normal, measurement direction, and the measurement distance respectively, at the measurement point. $\dot{r}$ indicates the change rate of the distance r. Note that $\mathbf{n}$ and $\mathbf{t}$ are normalized, and $\mathbf{t}$ is fixed between frames.

Similar to eq.(8) and (9), eq.(13) is a linear equation for the velocity vector $\mathbf{v_0}$ and rotational velocity vector $\boldsymbol{\omega}$. When eq.(13) are obtained at six or more points, the motion parameters of the sensor can be estimated by solving simultaneous equations.

The proposed method uses the features of color images and range images. On the other hand, range-motion equation uses only features of range images. However, range-motion equation needs multiple measurement points around the target point in order to make an equation, because it is necessary to obtain a normal. And it is sometimes difficult to obtain accurate normals from a range image, that affects the accuracy of estimation of motion parameters.

Another difference is that the proposed method has to track feature points between frames, which is not necessary for the range-motion equation.

In summary, it can be said that these two methods have complementary characteristics. Combination of the two methods, that is easily accomplished by just joining the simultaneous equations, may be effective.

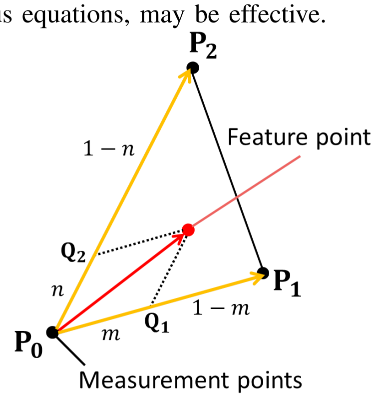


Fig. 3. Interpolation of 3D coordinates

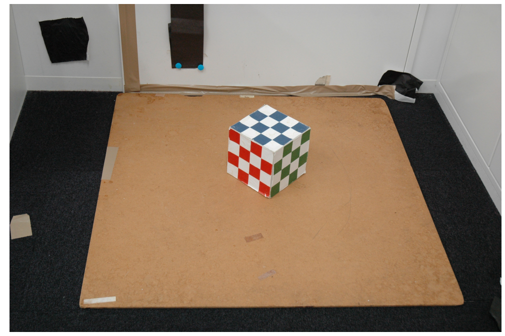


Fig. 4. The measurement object for evaluation of accuracy: a cubic box

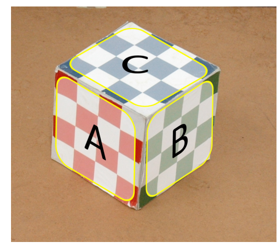


Fig. 5. The planes of a cubic box for evaluation of accuracy

IV. EXPERIMENTS

In this section, we verify the proposed method by experiments. We conduct two kinds of experiments. First, we evaluate accuracy of 3D mapping. Secondly, we show results of 3D mapping for two scenes: a scene containing several planes, and a scene with only small and complex objects.

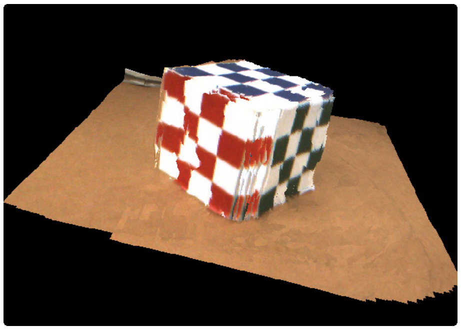
We compare three methods: the method using range-motion equation (13), the proposed method using eq.(8) and (9), and the combination of the two methods. We refer to the three methods as previous method, proposed method, and combined method.

A. Evaluation of accuracy of 3D mapping

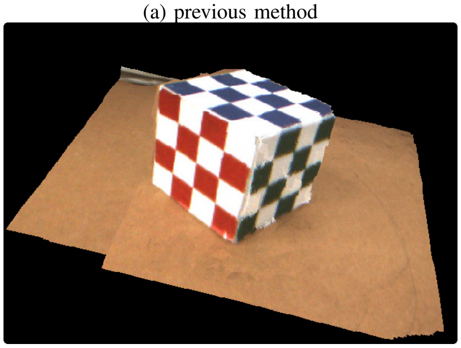
1) Experimental scene: A 3D map is constructed by aligning range images. When range images are aligned accurately, measurement points are expected to distribute on the real shape of the measured object with small dispersions. To evaluate the accuracy of 3D mapping, we measured a cubic box as shown in Fig.4 and constructed a 3D map. In this case, the measurement points are expected to distribute on planes.

Range images and color images were captured using the RGB-D sensor held by a person, and a 3D map was constructed off-line. The number of aligned images was 50. We fitted least-squares planes to the points on three planes in the constructed 3D map, as illustrated in Fig.5.

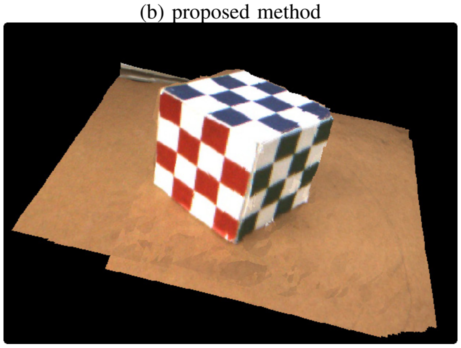
2) Evaluation method of accuracy: The average of absolute error of the points from the fitted plane was obtained, and accuracy of the alignment is evaluated by the error. In addition, angles between two fitted planes, that should be 90° ideally, were also calculated to evaluate the accuracy of the alignment.



(a) previous method



(b) proposed method



(c) combined method

Fig. 6. Constructed 3D map with texture for Fig.4

3) Experimental result: Fig.6 show the constructed 3D map with texture mapping for Fig.4. Table I shows the average of absolute error of the points from the fitted plane for each plane in Fig.5. Table II shows the measured angles between two planes.

4) Experimental discussion: It is shown that, when the proposed method is used, the fitting error for each plane is less than 4 mm, and the error of the angles between two planes is less than 2°. It can be said that the proposed method can construct an accurate 3D map for the scene shown in Fig.4. Additionally, differences between the accuracy of the three methods for plane fitting are not apparent in Table I and II. However, it is obvious that the quality of texture mapping of the proposed method and the combined method is much better than the previous method, that shows the effectiveness of the proposed method.

B. 3D map construction

1) Experimental scene: We constructed 3D maps using the three methods. We conducted experiments for two different scenes: a scene containing several planes as shown in Fig.7, and a scene with only small and complex objects as shown in Fig.8. The number of range images for Fig.7 and 8 is 150 and 200 frames, respectively.

TABLE I

AVERAGE OF ABSOLUTE ERROR OF THE POINTS FROM THE FITTED PLANE

plane	previous method	proposed method	combined method
plane A [mm]	2.89	3.12	3.13
plane B [mm]	1.98	1.14	1.34
plane C [mm]	1.11	1.34	1.36

TABLE II

ANGLE BETWEEN TWO PLANES

planes	previous method	proposed method	combined method
A and B [°]	89.00	88.26	88.14
B and C [°]	88.33	88.54	88.36
C and A [°]	89.01	89.86	89.97

2) Experimental result: Fig.9, 10 and 11 show the constructed 3D map for Fig.7 using the three methods respectively. And Fig.12 shows the constructed 3D map with texture mapping for the same scene.

Fig.13, 14 and 15 show the constructed 3D map for Fig.7 using the three methods respectively. And Fig.16 shows the constructed 3D map with texture mapping for the same scene. Table III shows the average estimation time of the motion parameters. The average number of linear equations was 159, 291 and 448 for the three methods respectively.

3) Experiment discussion: The experimental results of constructing a 3D map for Fig.7 have shown that each method can construct an accurate 3D map for the scene containing several planes. On the contrary, the experimental results for Fig.8 have shown that the proposed method and the combined method can construct an accurate 3D map even for the scene with only small and complex objects, but that the previous method cannot. The reason is thought that the previous method requires some planes in the scenes to obtain range-motion equations at sufficient number of points, that is not satisfied in Fig.8. On the contrary, the proposed method can estimate the motion parameters if only optical flow in color images are obtained (with range images), that is possible even for the small and complex objects in Fig.8.

The combined method is thought to have the merits of both

Fig. 7. Measurement scene: objects containing several planes

Fig. 8. Measurement scene: small and complex objects

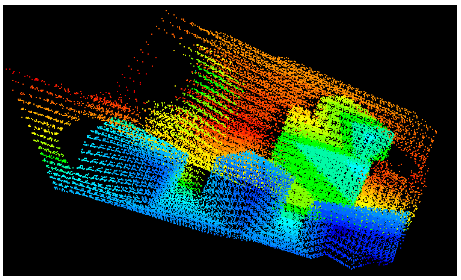


Fig. 9. Constructed 3D map for Fig.7: previous method

methods. For the scenes Fig.7 and 8, the combined method could construct an accurate 3D map thanks to the constraints given by the proposed method. Additionally, it is expected that the combined method can construct a 3D map for a scene with little texture, for which the proposed method is thought to be impossible to construct a 3D map accurately, thanks to the constraints given by the range-motion equation.

To summarize, it can be said that the proposed method can be used for more variety of scenes than the previous method and its measurement accuracy is superior to the previous method. In addition, the combination of the two methods can extend the applicable scenes.

Table III shows that estimation time of motion parameters using the proposed method is longer than 25 ms, and it is about 50 times as long as the previous method. This is because acquisition of optical flow takes long time. The proposed method cannot construct a 3D map on-line because of the long time, so it is required to reduce processing time to estimate the motion parameters.

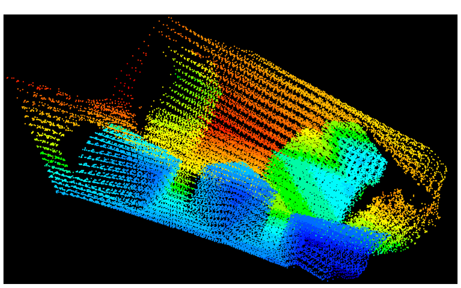


Fig. 10. Constructed 3D map for Fig.7: proposed method

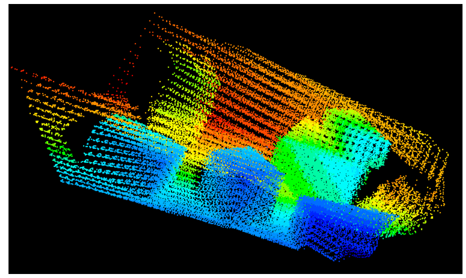
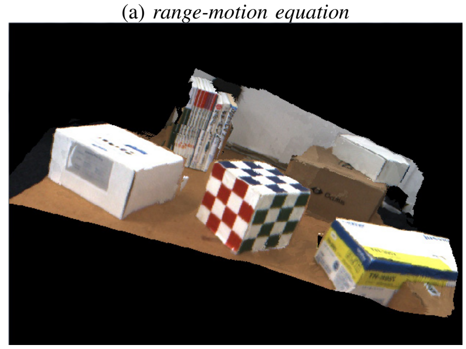
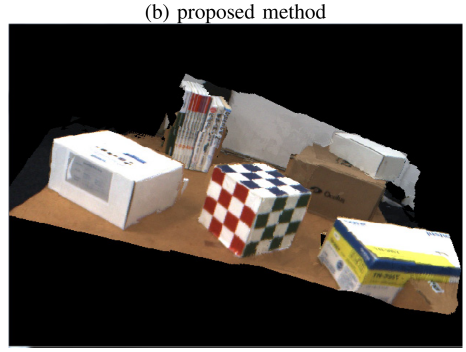


Fig. 11. Constructed 3D map for Fig.7: combined method

(a) *range-motion equation*



(b) proposed method



(c) combined method

Fig. 12. Constructed 3D map with texture for Fig.7

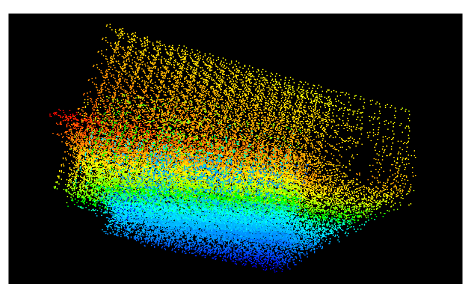


Fig. 13. Constructed 3D map for Fig.8: previous method

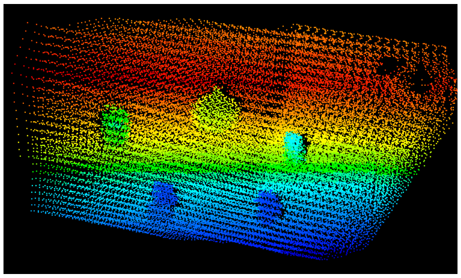


Fig. 14. Constructed 3D map for Fig.8: proposed method

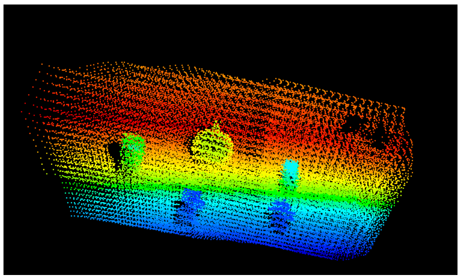
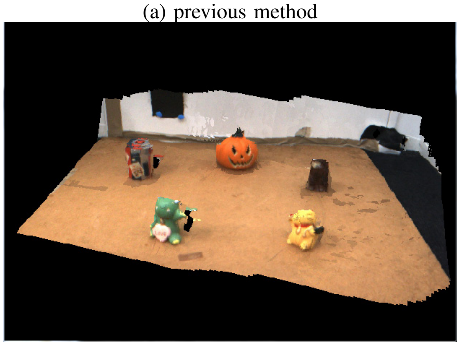
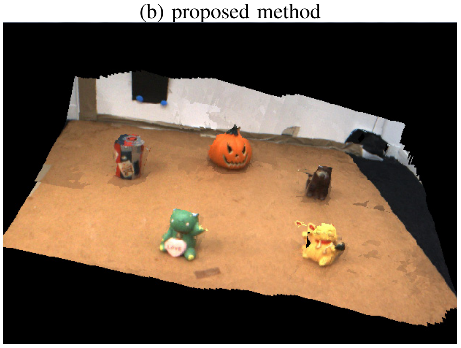


Fig. 15. Constructed 3D map for Fig.8: combined method

(a) previous method



(b) proposed method



(c) combined method

Fig. 16. Constructed 3D map with texture for Fig.8

TABLE III

ESTIMATION TIME OF THE MOTION PARAMETERS

Scene	Fig.7	Fig.8
Previous method [ms]	0.43	0.55
Proposed method [ms]	25.97	28.61
Combined method [ms]	28.14	28.86

V. CONCLUSIONS

We have proposed a method that directly estimate motion parameters of the sensor using range images and optical flow of color images. We have shown that the proposed method can construct an accurate 3D map. In addition, we have shown that the proposed method can be used for more variety of scenes than the previous method.

As a future work, we are planning to verify versatility of the proposed method by constructing a 3D map using different range image sensors.

REFERENCES

[1] S. Kagami, Y. Takaoka, Y. Kida, K. Nishiwaki, 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.

[2] M. Tomono, "Robust 3D SLAM with a Stereo Camera based on an Edge-Point ICP Algorithm," Proc. of 2009 IEEE International Conference on Robotics and Automation (ICRA), pp.4306-4311, 2009.

[3] G. Godin, D. Laurendeau, 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.

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

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

[6] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, W. Burgard, An Evaluation of the RGB-D SLAM System, Proc. of 2012 IEEE International Conference on Robotics and Automation (ICRA), pp.1691-1696, 2012.

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

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

[9] S. Nozaki, M. Kimura, G. Masuyama, K. Umeda, High-speed Three-dimensional Mapping by Direct Estimation of a Small Motion Using Range Images, Proc. of 11th France-Japan congress on Mechatronics (MECATRONICS - REM 2016). (submitted)

[10] C. Tomasi, and T. Kanade, Detection and tracking of point features, Proc. of Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.