Iterative Closest Points Method based on Photometric Weight for 3D Object Reconstruction

Dong-Won Shin and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST),
123 Cheomdangwagi-ro, Buk-gu, Gwangju, 61005, Republic of Korea
E-mail: {dongwonshin, hoyo}@gist.ac.kr

***Abstract*—Interests on 3D object reconstruction digitizing the shape and color of an object from the real world are getting popular. 3D object reconstruction consists of various steps such as image acquisition, image refinement, point cloud generation, iterative closest points, bundle adjustment and model surface representation. Among them, iterative closest points method is critical to calculate the accurate initial value for the optimization in the following bundle adjustment step. There is the object drift problem in the existing iterative closest points method due to the accumulated trajectory error as time flows. In this paper, we performed a more accurate registration between point clouds by SIFT features and the weighting on them. We found the proposed method decreases the absolute trajectory error and reduces the object drift problem in the reconstructed 3D object model.**

I. INTRODUCTION

3D reconstruction is the technology to digitize the shape and color of the object in the real scene, which it can be widely exploited for various applications such as virtual reality (VR), augmented reality (AR), E-commerce, medical imaging and game. It helps giving a realistic feeling of scenes and objects for an immersive play.

Passive and active methods are the two main folds of 3D reconstruction. The passive method calculates the shape information with the disparity or illumination of color images from multiple views. It has a relatively low cost. On the other hand, the active method projects the infrared pattern to the scene and measures the variation. It can compute the more reliable shape compared to the passive method [1].

In this paper, we represent 3D reconstruction method using the information from both passive and active method jointly using RGB-D camera. Figure 1 shows the flowchart of the general 3D reconstruction system using a hand-held RGB-D camera [2].

First, it obtains a series of color and depth images from the RGB-D camera and refines the depth image to fill the hole region with the proper depth values. 3D point cloud is calculated with the refined depth image and the intrinsic camera parameter. The camera extrinsic parameters including the rotation matrix and translation vector are calculated with a pair of point clouds at time *t-1* and *t* in 3D model registration step. Lastly in 3D model representation step, it represents the surface model of 3D voxel from the point clouds, then finally we can get a complete 3D model of the object.

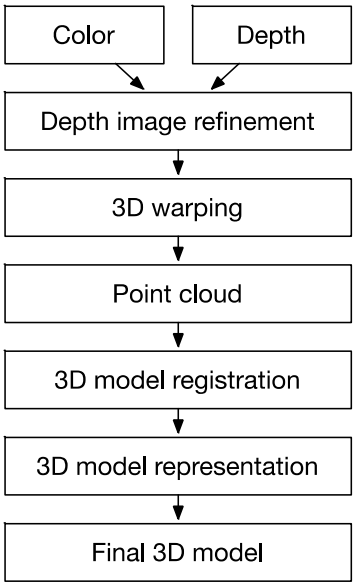

Fig. 1 Flowchart of the general 3D reconstruction system using a hand-held RGB-D camera

In the overall procedure, 3D model registration step is important to improve the accuracy of the reconstructed 3D object. If the tracking information of the camera is contaminated, it directly degrades the consistency of the object. Therefore, it is crucial to estimate the correct tracking information of the camera since the error would propagate as time goes on.

Usually, 3D model registration step is representative by the iterative closest points (ICP) method and we mainly focused on this procedure. There are two conventional ICP methods: point-to-point and point-to-plane. Arun et al. presented the point-to-point ICP method using the singular value decomposition (SVD) of the least-squares problems with decoupling the rotation and translation from the transformation [3]. Afterward, Low introduced the point-to-plane ICP method minimizing the distance between a point and the tangent plane at its correspondence point [4]. However, both methods have the drift problem on the reconstructed 3D model as accumulating the trajectory error due to the nature of ICP. Hence, in this paper we propose a ICP method using the color information from SIFT features to improve the registration ability and reduce the drift problem.

II. PROPOSED ICP METHOD

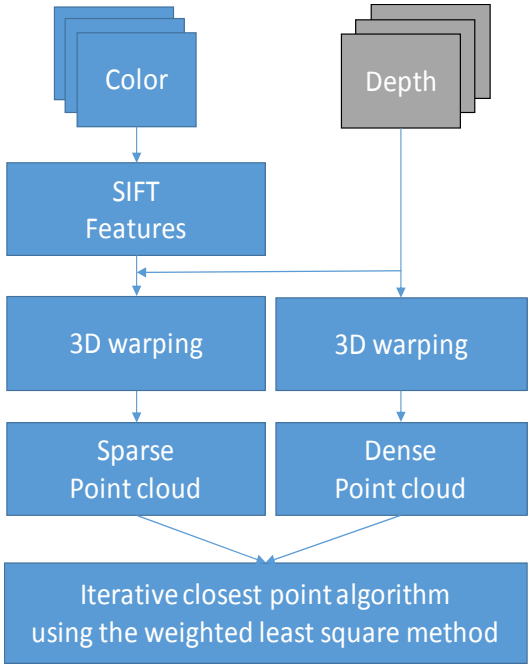

Fig. 2 Flowchart of the proposed ICP method

Figure 2 shows the flowchart of the proposed ICP method. First of all, we capture the color and depth images from the RGB-D camera. After detecting the SIFT features from the color image at *t* frame, we calculate the sparse point clouds using the corresponding depth image and intrinsic camera parameter [5]. Similarly, we compute the dense point clouds from all valid pixels having non-zero value in the depth image. Afterward, in the time *t+1* frame, we obtain the sparse point cloud from the color image and the dense point cloud from the depth image. At this point, in order to find the correct rotation and the translation of the camera between *t* and *t+1* frame, we give the additional weight on the correspondences from the sparse point clouds and solve the cost function using the weighted least square method.

A. Proposed point-to-point method

In this section, we derive the proposed point-to-point method weighting on the sparse point cloud from color images to find the rotation and translation of the camera. In the conventional point-to-point method, it minimizes the cost function at (1) to find the rotation matrix R.

$$\mathrm{E} = \sum_{i=1}^{N} w_i \| p_i^t - R p_i^{t+1} \|^2 \tag{1}$$

At this point, p_i^t means the *i*-th point from the point cloud at *t* frame including the sparse and dense point clouds. Equation (2) shows the expansion of (1).

$$\mathrm{E} = \sum_{i=1}^{N} w_i ((p_i^t)^T p_i^t + (p_i^{t+1})^T p_i^{t+1} - 2(p_i^t)^T R p_i^{t+1}) \tag{2}$$

In this case, Minimizing (2) is equal to maximizing (3) [3].

$$\begin{aligned} \mathrm{F} &= \sum_{i=1}^{N} w_i (2(p_i^t)^T R p_i^{t+1}) \\ &= \mathrm{Trace}(\sum_{i=1}^{N} w_i (p_i^t)^T R p_i^{t+1}) \\ &= \mathrm{Trace}(\mathrm{RH}) \\ \text{where } \mathrm{H} &= \sum_{i=1}^{N} w_i (p_i^t)^T p_i^{t+1} \end{aligned} \tag{3}$$

Matrix R maximizing (3) can be calculated by the result from the singular value decomposition (SVD) on matrix H as shown (4).

$$\begin{aligned} \mathrm{H} &= \mathrm{U}\mathrm{D}V^T \\ R^* &= VU^T \end{aligned} \tag{4}$$

The Translation vector *t* can be obtained by (5).

$$\begin{aligned} T^* &= p_c^t - R^* p_c^{t+1} \\ \text{where } p_c^t &= \frac{1}{N}\sum_{i=1}^{N} p_i^t \,, p_c^{t+1} = \frac{1}{N}\sum_{i=1}^{N} p_i^{t+1} \end{aligned} \tag{5}$$

B. Proposed point-to-plane method

In this section, we derive the proposed point-to-plane method weighting on the sparse point cloud. In the conventional point-to-plane method, it finds the vector x minimizing (6). Vector x includes the rotation and translation parameters.

$$\begin{aligned} E &= \sum_{i=1}^{N} w_i \| A_i x - b_i \|^2 \\ \text{where } A_i &= \begin{bmatrix} n_{iz} p_{iy}^t - n_{iy} p_{iz}^t, \\ n_{ix} p_{iz}^t - n_{iz} p_{ix}^t, \\ n_{iy} p_{ix}^t - n_{ix} p_{iy}^t, \\ n_{ix}, \\ n_{iy}, \\ n_{iz} \end{bmatrix}, \\ b_i &= \begin{bmatrix} n_{ix} p_{ix}^{t+1} + n_{iy} p_{iy}^{t+1} + n_{iz} p_{iz}^{t+1} \\ -n_{ix} p_{ix}^t - n_{iy} p_{iy}^t - n_{iz} p_{iz}^t \end{bmatrix}, \\ x &= [\alpha, \beta, \gamma, t_x, t_y, t_z] \end{aligned} \tag{6}$$

At this point, n represents the normal vector and α, β, γ mean the rotation angle for x, y, z axis respectively.

Equation (6) can be represented by the matrix multiplication as (7).

$$\mathrm{E} = \sum_{i=1}^{N} w_i r_i^2 = r^T W r = \left\| W^{1/2} r \right\|^2 \tag{7}$$

After taking the vector derivative on (7) and find the minimum, we can obtain the minimizer x^* as (8).

$$x^* = (A^T W A)^{-1} A^T W b \tag{8}$$

III. EXPERIMENT RESULTS

For the experiment, we exploited a Teddy sequence including color, depth images and ground truth camera trajectory from the computer vision group at Technische Universität München [6]. This data set was captured by Microsoft Kinect v1. Using the data set, we estimated the camera trajectory by the conventional and proposed ICP methods and calculated an absolute trajectory error (ATE) comparing the distance between the ground truth camera trajectory and the estimated camera trajectory from ICP methods.

First, we divide the sections for the whole trajectory (0-1900) by 100 frames each since it would be the bundle adjustment problem rather than the ICP problem. And then, ICP methods are performed for the sections.

In the proposed ICP methods, the amount of SIFT correspondences and the weighting value on them greatly affect the accuracy of the estimated trajectory. In order to find the minimum ATE for each section, we set a weight space and took all the feasible combinations in the space for the proposed ICP method. The minimum ATE for each section was selected and compared with the ATE from conventional methods. Table 1 shows the weight space used in this experiment.

The weights represent the dependency on the correspondences and the good matches thresholds mean the amount of the correspondences in the image.

Table 1. Weight space for the experiments

Types	Values
Weights	[0.01, 0,1, 10, 100]
Good matches thresholds (GMT)	[2,5,10]

Figure 3 shows the ATE results for each section from both conventional and proposed point-to-point ICP methods. The red and blue bars represent the results from the conventional and proposed method respectively. For the most part, the result from proposed point-to-point ICP method shows the reduced

ATE value and exhibits the definite differences especially in 100-200, 200-300 and 1400-1500 sections. Next, Fig. 4 shows the ATE results from the point-to-plane ICP methods and it also indicates the proposed point-to-plane method mostly decreases the ATE and definitely cuts down in 0-100, 200-300, 1700-1800 and 1800-1900 sections.

Table 2 shows the weight and GMT values related to the minimum ATE for each section. As we can find in Table 2, the proposed method has the high weights in the section representing the definite differences and the low weights in the section indicating the minor differences. This suggests that the weight should be appropriately selected depending on the property of the images in the section.

Based on this analysis, we will study on selecting the adaptive weights according to the property of the color and depth images in the future.

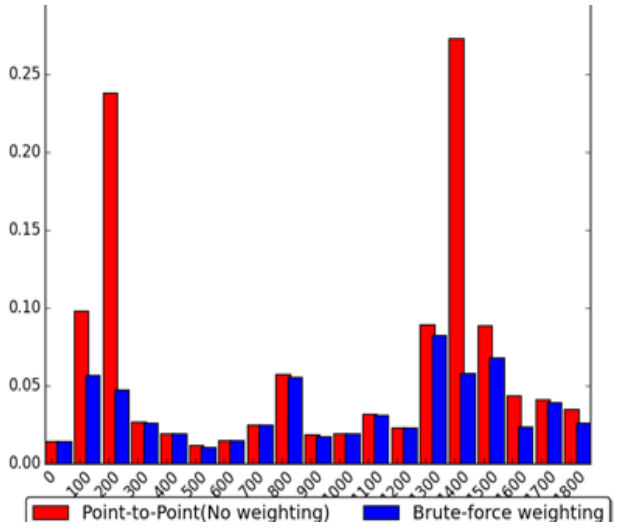

Fig. 3 ATE graph from the point-to-point ICP methods

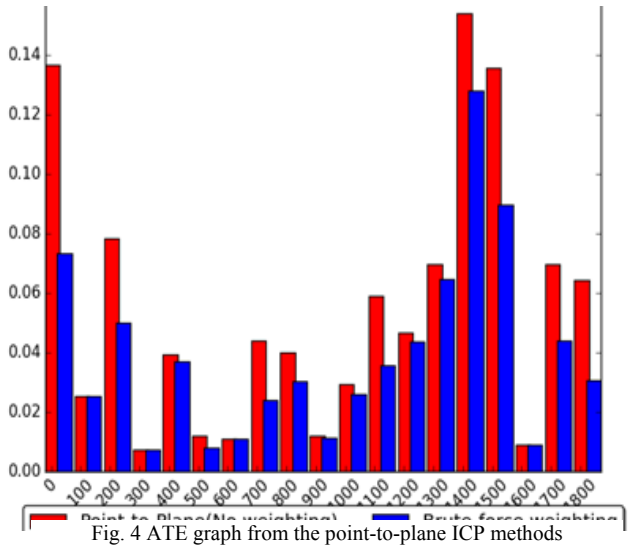

Fig. 4 ATE graph from the point-to-plane ICP methods

Table 2. Weight and GMT values related to the minimum ATE for each section

	Section	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800	800-900	900-1000	1000-1100	1100-1200	1200-1300	1300-1400	1400-1500	1500-1600	1600-1700	1700-1800	1800-1900
Proposed Point-to-Point	Weight	0.1	100	100	10	0.1	0.1	0.1	0.01	0.01	10	0.01	0.01	0.1	100	100	100	100	100	10
	GMT	2	10	5	2	2	5	2	5	10	5	2	2	2	10	2	5	5	10	10
Proposed Point-to-Plane	Weight	10	0.01	100	0.01	100	10	0.01	100	100	10	10	10	100	100	100	10	0.01	100	100
	GMT	10	5	2	2	5	5	2	5	2	2	5	5	2	5	5	2	2	5	5

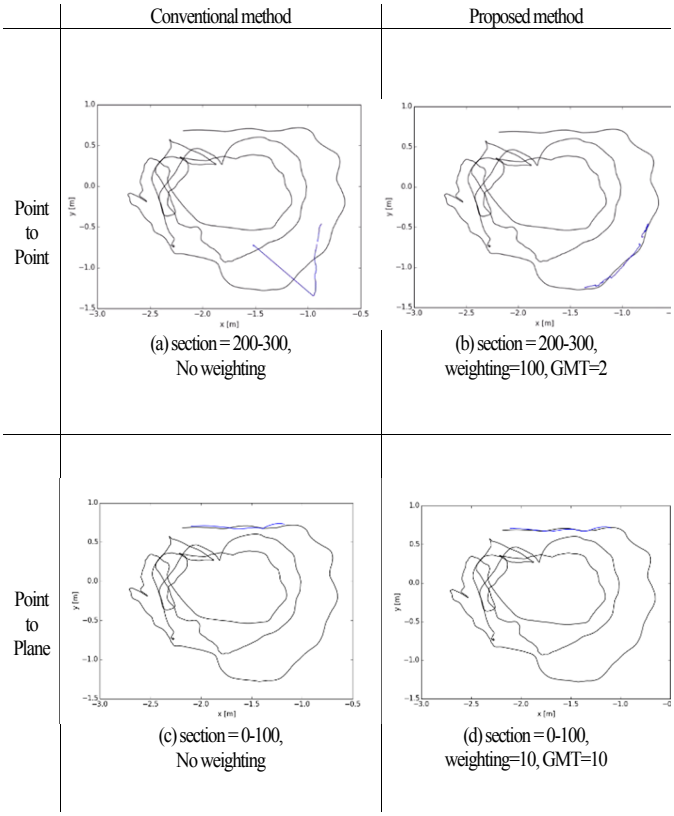

Fig. 5 Estimated camera trajectories

Next, the estimated camera trajectories using ICP methods are shown in Fig. 5. The black curve represents the ground truth camera trajectory and the blue curve means the estimated camera trajectory from the ICP methods. Among all sections, we exhibited specific results having the definite differences between the convention and proposed methods.

We showed the results from the 200-300 section by the point-to-point ICP methods in Fig. 5(a) and Fig. 5(b). While we can find the incompatible results with the ground truth data in Fig. 5(a), the proposed point-to-point ICP method shows a more suitable result in Fig. 5(b). The results from the proposed point-to-plane ICP method in Fig. 5(d) also shows the more accurate result than the conventional point-to-plane ICP method from Fig. 5(c).

For the more visual comparison, we showed the reconstructed 3D model from each ICP methods. In Fig. 6(a) and Fig. 6(b), we represented the result from the point-to-point ICP methods for the 200-300 section. We can notice that the accumulated camera trajectory errors deteriorate the consistency of the reconstructed 3D model. On the other hand, Fig. 6(b) shows the relatively consistent 3D model since the camera trajectory errors are reduced.

Similarly, we represented the result from the point-to-plane ICP methods for the 0-100 section in Fig. 6(c) and Fig 6(d). We can also find the object drift problem in the result from the conventional point-to-plane ICP method while the result from the proposed point-to-plane ICP method shows the accurate 3D model.

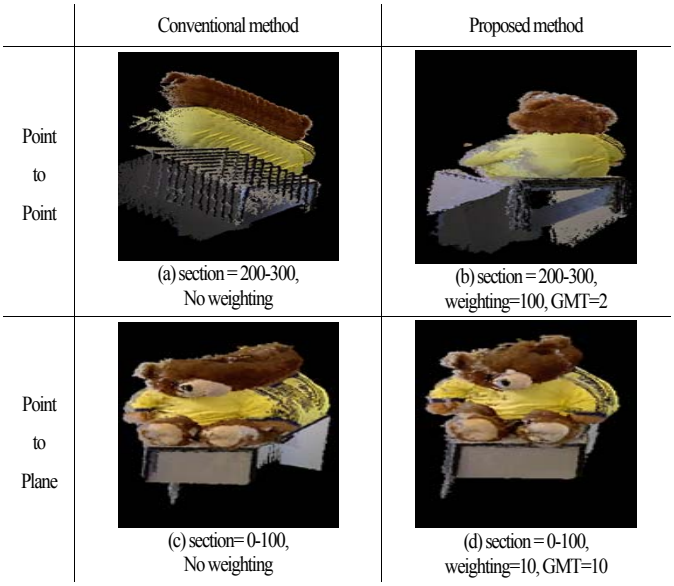

Fig. 6 Reconstructed 3D models

IV. Conclusions & Future Works

In this paper we proposed the new iterative closest points methods using the photometric information from SIFT features for 3D object reconstruction. The proposed method solved the object drift problem by reducing the accumulated camera trajectory error and estimated the trajectory close to the ground truth data. Though the minimum ATE from the results using all feasible weight combinations by the proposed method was compared with the ATE from conventional method, we will study the adaptive weight method depending on the property of the section for the future works.

Acknowledgement

This work was supported by the 'Civil-Military Technology Cooperation Program' grant funded by the Korea government.

References

[1] D. W. Shin and Y. S. Ho, "Implementation of 3D object reconstruction using a pair of Kinect cameras," APSIPA, pp. 1-4, 2014.

[2] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In Proceedings of the 24th annual ACM symposium on User interface software and technology, pp. 559–568, 2011.

[3] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets.," PAMI, vol. 9, no. 5, pp. 698-700, 1987.

[4] K. L. Low, "Linear least-squares optimization for point-to-plane icp surface registration," Chapel Hill, 2004.

[5] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints.," IJCV, vol. 60, no. 2, pp. 91-110, 2004.

[6] http://vision.in.tum.de/data/datasets/rgbd-dataset/download