Fine Registration of 3D Point Clouds with Iterative Closest Point Using an RGB-D Camera

Jun Xie[1], Yu-Feng Hsu[2], Rogerio Schmidt Feris[3], and Ming-Ting Sun[1]

[1] Dept. of Electrical Engineering
University of Washington,
Seattle, WA, US

[2]Industrial Technology Research
Institute (ITRI), Taiwan

[3]IBM T. J. Watson Research Center
Hawthorne, NY, US

Abstract--We address the problem of accurate and efficient alignment of 3D point clouds captured by an RGB-D (Kinect-style) camera from different viewpoints. Our approach introduces a new cost function for the iterative closest point (ICP) algorithm that balances the significance of structural and photometric features with dynamically adjusted weights to improve the error minimization process. We also enhance the algorithm with a novel outlier rejection method, which relies on adaptive thresholding at each ICP iteration, using both the structural information of the object and the spatial distances of sparse SIFT feature pairs. The effectiveness of our proposed approach is demonstrated in challenging scenarios, involving objects lacking structural features, and significant camera view and lighting changes. We obtained superior registration accuracy than existing related methods while requiring low computational processing.

I. INTRODUCTION

3D object modeling is an active research topic, and has many practical applications such as animation, human computer interaction, virtual reality, and object manipulation by industrial robots [1, 2, 4]. With the birth of RGB-D cameras (such as Kinect), synchronized RGB and depth images can be captured at the same time, making 3D modeling of an object more robust and accessible.

In a typical 3D modeling process using an RGB-D camera, first, the 3D partial point clouds of the object from different views are pairwise registered through a coarse registration algorithm such as RANSAC (Random Sample Consensus) [5, 9], and then this initial registration is further refined by an iterative fine registration algorithm such as ICP (Iterative Closest Point) [6, 7]. After the fine registration, the 3D point cloud model can be transformed to other 3D representations for different applications.

The ICP convergence is sensitive to outliers. To improve the performance of ICP, many variants of ICP have been proposed [3, 11]. The variants cover the selection, matching, weighting, outlier rejection of the 3D points, and the minimization of the error metric. However, in some cases such as an object lacking salient structural features or under significant camera view and lighting changes, even if an almost perfect initial alignment is achieved, these ICP variants may actually converge to an incorrect alignment result since only the 3D structural information is used. Several color based ICP algorithms [12, 13] have been proposed to alleviate this issue, showing that adding the color information decreases the registration error when objects lack structural features. However, directly using color is not reliable since the color may not be the same for the same point in different views due to lighting, shadow, or reflection. In [14], SIFT descriptors have been incorporated into the ICP iteration process for improving registration. However, in this work the algorithm operates solely on sparse SIFT feature points, which is a very small subset of the point clouds in the 3D case. The performance of the algorithm is limited since the structural information from the 3D point clouds is not fully used. The algorithm could be extended to run on all the 3D points in the point clouds. However, it would require to compute the 128-dimensional SIFT descriptor for every point in the object in the search for the closest distance, which is very computationally inefficient. Also, it will have problems if the object lacks salient texture features. Furthermore, a fixed coefficient for weighting the closest distance and SIFT matching distance is utilized in this work, which may not provide the best performance.

In this paper, we propose a more robust and efficient point cloud registration approach by enhancing ICP with a new cost function that balances the significance of structural and photometric features with dynamically adjusted weights to improve the error minimization process. In addition, we introduce a novel outlier rejection method, which adaptively sets the outlier distance threshold at each ICP iteration by taking into account both 3D structure of the object and the spatial distances of sparse SIFT feature pairs. We show that our contributions enable ICP to achieve superior results than other related methods, in terms of both registration accuracy and efficiency. In particular, we demonstrate our approach in several challenging scenarios, involving symmetrical objects and alignment with large camera view and lighting changes.

The organization of the rest of the paper is as follows. In Section 2, we present our proposed approach. In Section 3, we discuss our setup to demonstrate the performance of our proposed algorithm. In Section 4, we perform simulations to show the effectiveness of the proposed techniques. In Section 5, we conclude the paper.

II. TECHNICAL APPROACH

The standard ICP algorithm aligns two point clouds by iteratively associating points through nearest-neighbor search and estimating the transformation using a mean square cost function. In our approach, to overcome the problem associated with the case of objects lacking structural features, we add a SIFT-based term into the cost function for error minimization. To utilize the structural information of the 3D points without intensive computation of the SIFT descriptor for every 3D point, we propose to add a constraint involving the spatial distances of the SIFT feature corresponding pairs which are readily available in the iterations. This added term effectively constrains the convergence to the correct direction which minimizes the spatial distances of points with structural features and texture features. In addition to this constraint, we use a new dynamic weight to properly balance the significance of structural and photometric terms. Moreover, since the initial alignment using SIFT matching and RANSAC is not

978-1-4673-5762-3/13/$31.00 ©2013 IEEE

perfect, some of the SIFT points may not be exactly paired. Thus, some of the correspondence pairs with large distances should be rejected. Therefore, we propose a new outlier rejection method which adaptively utilizes both the statistics of structural characteristics and the spatial distances of SIFT correspondence pairs. Our proposed approach is also effective to improve the performance of ICP under the situation of significant camera view changes. In this situation, the overlapping region is relatively small. If the threshold of the outlier rejection only depends on the statistics of the closest distances, the threshold will be relatively large due to the large number of outliers, meaning fewer outliers will be rejected. This may cause inaccurate registration results. Our outlier rejection method makes the threshold tighter under this situation which improves the performance of the registration.

The procedure of the proposed 3D registration approach is described below, followed by a detailed discussion and explanation of the equations in our formulation.

Initial Registration:

Given the RGB and depth images of two views from the RGB-D camera, we obtain two 3D point clouds $p = \{p_1 \dots p_N\}$ and $q = \{q_1 \dots q_M\}$, where N and M are the numbers of points in p and q. The SIFT feature points are extracted from the two RGB images. After the initial alignment with the RANSAC process, we find the set of SIFT feature correspondence 3D points as $cf = \{(pf_1, qf_1) \dots (pf_L, qf_L)\}$ where pf_i and qf_i are the corresponding SIFT feature 3D points in p and q, and L is the number of matched SIFT feature pairs. Define $T^{(k)}$ as the transform matrix after the kth iteration. $T^{(0)}$ is the transform after RANSAC and before the ICP iteration.

Fine Registration:

For the kth iteration (k=1, 2, …) in the process:

(i) For each point p_i in the point cloud p, find its corresponding point $q_i^{*(k)}$ in q with the closest distance:

$$q_i^{*(k)} = arg_{q_j}[min_{q_j \in \{q\}}(\|p_i \cdot T^{(k-1)} - q_j\|)], \quad (1)$$

The associated closest Euclidean distance with p_i is $cd_i^{(k)} = \|p_i \cdot T^{(k-1)} - q_i^{*(k)}\|$.

(ii) Compute the mean and standard deviation of all the closest distances. Define the statistic inlier $s^{(k)}$ as a subset of p satisfying:

$$s^{(k)} = \{p_i \mid cd_i^{(k)} < mean(cd^{(k)}) + 3 \cdot std(cd^{(k)})\}. \quad (2)$$

(iii) Calculate the adaptive threshold for outlier rejection:

$$t^{(k)} = c \cdot \sqrt{er^{(k)}} \cdot df^{(k)}, \quad (3)$$

where c is a constant (we set c = 30 in all simulations).
$df^{(k)}$ is the average spatial distance of m% SIFT feature correspondence pairs with shorter closest distances (we set m = 30 in all simulations).
$er^{(k)}$ is the root mean square of the closest distances for $s^{(k)}$ defined in (2):

$$er^{(k)} = \sqrt{mean_{p_i \in s^{(k)}}(\|p_i \cdot T^{(k-1)} - q_i^{*(k)}\|^2)},$$

(iv) Define a new objective function $f(T)$ as

$$f(T) = \sum_{p_i \in p} \alpha_i \|p_i \cdot T - q_i^{*(k)}\|^2 + \sum_{(pf_i, qf_i) \in cf} \beta_i \cdot \|pf_i \cdot T - qf_i\|^2, \quad (4)$$

where we use α_i to reject the outliers:

$$\alpha_i = \begin{cases} \sqrt{\dfrac{0.01}{|\sigma(p_i) - \sigma(q_i^{*(k)})|}} & if\ cd_i^{(k)} < t^{(k)} \\ 0 & otherwise. \end{cases} \quad (5)$$

Here $\sigma(*)$ denotes the surface variation defined in [8]:

$$\sigma(p_i) = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2},$$

$\lambda_i (\lambda_0 \le \lambda_1 \le \lambda_2)$ are eigenvalues of covariance matrix:

$$C = \begin{bmatrix} p_{i1} - \overline{p_i} \\ \dots \\ p_{ir} - \overline{p_i} \end{bmatrix} \cdot \begin{bmatrix} p_{i1} - \overline{p_i} \\ \dots \\ p_{ir} - \overline{p_i} \end{bmatrix}^T,$$

$p_{i1} \dots p_{ir}$ are the closest r points around p_i and $\overline{p_i}$ is the centroid of these local neighbors.
The dynamically adjusted weight β_i is:

$$\beta_i = c' \frac{\sqrt{mean_{(pf_i, qf_i) \in cf}\left(\|pf_i \cdot T^{(k-1)} - qf_i\|^2\right)}}{matching\_dist(pf_i, qf_i) \cdot er^{(k)}} \quad (6)$$

where c' is a constant which we empirically set to 60, and matching_*dist(pfi,qfi)* is the 2D SIFT matching distance between *(pfi, qfi)* which is available from the SIFT feature matching in the RANSAC initial registration stage.

(v) Find the transformation $T^{(k)}$ by minimizing the objective function $f(T)$: $T^{(k)} = min_T(f(T))$. Also, we delete points p_i from p with $cd_i^{(k)} > 10 * t^{(k)}$ so that in next iteration, we just need to search for the closest distance for those remaining points. Thus, it can reduce much computation time.

(vi) The iteration terminates after the root mean square (RMS) of the closest distances of the inliers is smaller than a set threshold, or until a fixed number of iterations is reached (here we set the iteration number to 18 in our experiments).

In (3), we make the adaptive threshold $t^{(k)}$ for outlier rejection based on the average spatial distance of *30*% SIFT feature correspondence pairs with shorter spatial distances instead of all the feature pairs because even after the initial RANSAC based alignment, some of the feature points may not be exactly paired due to the resolution limitation or inaccurate matching. Using only a subset of feature correspondence pairs with shorter distances, we can ensure the accuracy of the chosen feature correspondences.

The first term in the objective function of (4) is the weighted mean square of the closest distances of the inliers. In (4) we show the Point-to-Point distance as the error metric. In the simulations, we also tried the Point-to-Plane distance [6] as the error metric, and the results are similar. The second regularization term is the weighted mean square of 3D spatial distances of the feature correspondence pairs. In addition to outlier rejection, we also use the local surface variation to control the weight α_i (5). As mentioned in [8], the surface variation is closely related to the curvature but needs much less computation than the curvature calculation. Here, we introduce the 3D local surface variation features

because the local features can have a better representation of the surface structure. If two points with the closest distance have similar surface variations, the weight is adjusted with more confidence and vice versa.

Also in (4), β_i is a dynamic weight for the feature correspondence pair (pf_i, qf_i). If the weight is set too large, the transformation mainly depends on the relatively small number of feature correspondences. This could cause problems when the feature correspondences are not completely reliable. On the other hand, if the weight is set too small, since the number of feature point pairs is much smaller compared to the number of the 3D point inliers in the first part of (4), the feature based regularization term becomes less significant. To resolve the above problems, we adaptively set β_i according to the SIFT matching distance as well as the ratio between the spatial distances of SIFT feature pairs and the closest distances of $s^{(k)}$. In (6), if the 2D SIFT feature matching distance is large, meaning the matching is less accurate, we assign less weights to the feature based regularization term. Also, we calculate the ratio between the RMS of the spatial distances of the SIFT feature pairs and the closest distances for $s^{(k)}$ in each iteration. If the RMS distance of the SIFT feature correspondences is relatively large compared to the RMS distance of $s^{(k)}$, we assign more weights to the feature based regularization term to have a more balanced minimization.

In the algorithm described above, the adaptive threshold t utilizes both the structural information and the SIFT features of the point clouds. Moreover, we introduce the SIFT feature matching constraint into the objective function with a dynamic weighting scheme. As a result, ICP will converge properly. Unlike the outlier rejection method described in [11], our proposed algorithm utilizes the texture feature information in the outlier rejections. Moreover, unlike the color based ICP algorithm in [12, 13], our method is more robust to lighting changes.

III. Experimental Setup

In this section, we show the setup we used to demonstrate the performance and illustrate the challenging scenarios of symmetrical objects where ICP produces inaccurate results.

The RGB-D images we used are from the RGB-D Object Dataset [10] from the Robotics and State Estimation Lab of University of Washington with an object placed on a turntable. We use only the SIFT features extracted from the object in the RGB images to perform RANSAC for the initial alignment. After the initial registration, we perform ICP on the food-can for the fine alignment to obtain the final transformation. Fig. 1 shows the partial 3D point clouds of a food-can registered from the two views after the initial RANSAC alignment and after ICP. It should be noted that most of the black area around the turntable belongs to the background. Since the food-can and the turntable are rigid and are fixed together, ideally, the transform should also be the same for the turntable. We use the red rectangle markers on the turntable as shown in the figure to demonstrate the problem. Since the markers are sharply defined and have a very distinct color from the turntable, it is easy to precisely extract the corners of the markers. These corner points serve as the ground truth points in our simulations for comparison. With the coordinates of the six ground truth points (corners), we can compute the distance of the ground truth points from two views after the fine registration. Since the markers are at a distance from the food-can, and ICP is performed only based on the 3D points of the food-can, the markers also serve the purpose of making the errors more visible.

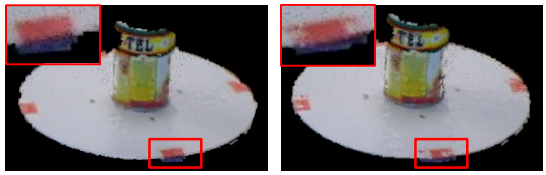


Fig. 1. 3D point clouds registered from two views. (left) Initial registration result. (right) Fine registration result after ICP.

From Fig. 1, we can see that after the coarse initial alignment, the red markers are well aligned (left). However, after ICP is performed for the fine registration (right), the markers are no longer aligned (mixed with red and white colors). This is because the food-can lacks salient structural features and ICP in this case will be converged to a wrong direction even with a good initial alignment. In Fig. 2, we plot the RMS of the closest distances of the 3D points and the RMS error of the distances of the ground truth points in each iteration. From the figure, we can see that although the RMS value of the closest distances of the 3D points is decreasing, the RMS error of the ground truth points is increasing, indicating that ICP is actually converging to a wrong position due to the lack of structural features. Moreover, in the scenario where the overlap region from two views is relatively small, we also find that ICP will encounter similar problems.

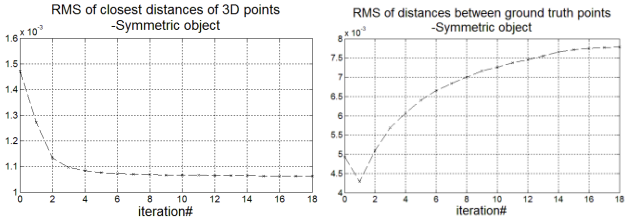
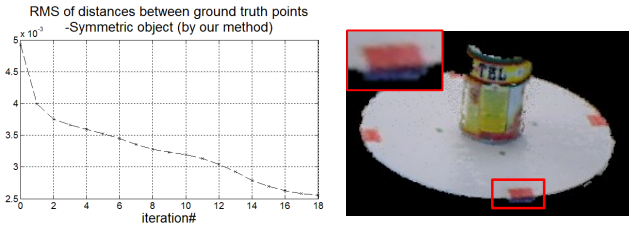


Fig. 2. Fine-registration error curve for a symmetrical object with ICP.

IV. Experimental Results of Proposed Algorithm

In this section, we demonstrate the improved accuracy and efficiency for the challenging scenarios described in the previous section with objects lacking salient structural features.



(a) (b)

Fig. 3. Fine-registration errors for the symmetric object by our proposed method (a) error curves, (b) the associated visual result of registering the two point clouds from Fig. 1 (compare to Fig. 2(b)).

In Fig. 3, we show the alignment result for the case of an object with a symmetrical structure (the food-can case in Fig. 1) using our proposed algorithm. In this case, the convergence towards the ground truth points does not have any problem and the error continues decreasing to a much smaller value. Also compared to the original ICP visual result in Fig. 1(b), with our proposed approach in Fig. 3(b), the markers are aligned very well, which shows the effectiveness of our approach.

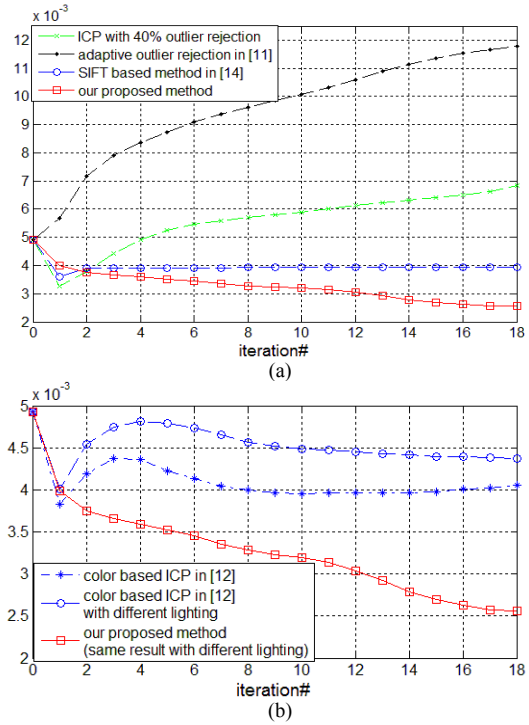


Fig. 4. Comparison of the RMS of distances between the ground truth points for the symmetric food-can case with (a) different methods. (b) with different lighting.

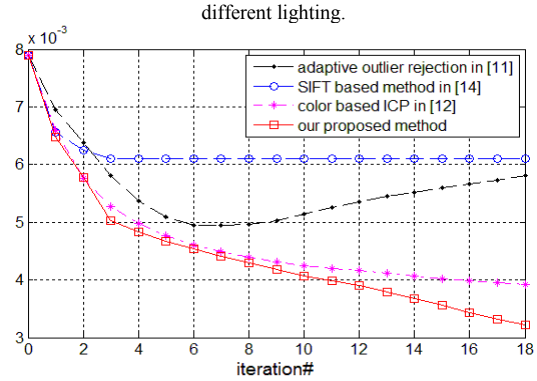


Fig. 5. Comparison of the RMS of distances between the ground truth points for the case of two views with small overlaps.

We compared our method with other previously mentioned methods in the above figures. In the food-can case, the approximate percentage of overlaps is about 60%, so we also draw the curves with a fixed 40% outlier rejection method (which gives better performance compared to other fixed percentages), the outlier rejection method in [11] and the SIFT based registration approach in [14] for comparison in Fig. 4(a). As can be seen from the figure, for all the previously mentioned approaches, the ICP registration results have larger mean square errors than ours. We also draw the curves of results from the color based ICP approach in [12] with different shading condition in Fig. 4(b). In the color based ICP results, the errors are much larger compared to the result of our approach and its accuracy varies significantly in different shading conditions. In our approach, since the SIFT descriptor is more robust to illumination changes, the inlier SIFT features do not change in this case so the varied shading does not affect the fine registration result. We also performed simulations for other cases such as when two views have small overlaps due to abrupt camera view changes as shown in Fig. 5. From it, we can see that our approach also performs better than other algorithms.

Besides, we also conduct experiments on the computation comparison. We assume that the initial alignment (including SIFT matching, RANSAC etc.) is performed before ICP in all cases so we can just compare the computation in the fine registration stage. From the experimental results, the time to calculate the surface variation and the outlier rejection threshold in our method is not significant and the speed of our algorithm does not differ much from the original ICP. In the small overlap cases, the speed is even faster than any of the above methods, which demonstrates the efficiency of our algorithm.

V. Conclusion

In this paper, to improve the accuracy and robustness of the ICP algorithm, we introduced a regularization term incorporating the spatial distances of the SIFT feature pairs with dynamically adjusted weights to balance the errors in the error minimization process. We also proposed a new outlier rejection method which is based on adaptive thresholding and leverages the structure and sparse feature pairs from the texture of the RGB images as a constraint to keep the ICP iterations in a right convergent track. Simulation results demonstrate the effectiveness of the proposed approach compared to previous methods.

References

[1] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel and S. Thrun, “Performance Capture from Sparse Multi-view Video,” ACM Transactions on Graphics, vol. 27, no. 3, pp. 98:1-10, 2008.

[2] G. Klein and D. Murray, “Parallel Tracking and Mapping for Small AR Workspaces,” International Symposium on Mixed and Augmented Reality, pp. 225 – 234, 2007.

[3] S. Rusinkiewicz and M. Levoy, “Efficient Variants of the ICP Algorithm,” *3DIM*, pp. 145-152, May 2001.

[4] A. Kashani, W. S. Owen, N. Himmelman, P. D. Lawrence, and R. A. Hall, “Laser Scanner-based End-effector Tracking and Joint Variable Extraction for Heavy Machinery,” The International Journal of Robotics Research, vol. 29, no. 10, pp. 1338-1352, 2010.

[5] M. Fischler and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Image Understanding Workshop, pp. 71-88, April 1980.

[6] Y. Chen and G. Medioni, “Object Modeling by Registration of Multiple Range Images,” *ICRA*, vol. 3, pp. 2724-2729, April 1991.

[7] P. Besl and N. McKay, “A method for registration of 3-d shapes,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, pp. 239–256, February 1992.

[8] M. Pauly, M. Gross and L. Kobbelt, “Efficient Simplification of Point-Sampled Surface,” IEEE Visualization, pp. 163-170, November 2002.

[9] R. Inomata, K. Terabayashi, K. Umeda and G. Godin, “Registration of 3D Geometric Model and Color Images Using SIFT and Range Intensity Images”, International Symposium on Advances in Visual Computing, vol. 6938, pp. 325-336, 2011.

[10] K. Lai, L. Bo, X. Ren, and D. Fox, “A Large-Scale Hierarchical Multi-View RGB-D Object Dataset,” *ICRA*, pp. 1817-1824, 2011.

[11] Z. Zhang, “Iterative point matching for registration of freeform curves and surfaces,” International Journal of Computer Vision, vol. 13, no. 2, pp. 119–152, 1994.

[12] A.E. Johnson and S.B. Kang, "Registration and Integration of Textured 3-D Data," *3DIM*, pp. 234-241, May 1997.

[13] S. Druon, M. Aldon, and A. Crosnier, “Color constrained ICP for registration of large unstructured 3d color data sets,” International Conference on Information Acquisition, pp. 249-255, August 2006.

[14] R. Lemuz-López and M. Arias-Estrada, “Iterative Closest SIFT Formulation for Robust Feature Matching,” International Symposium on Advances in Visual Computing, vol. 4292, pp. 502-513, 2006.