MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

A Globally Convergent Algorithm for Range Image Registration Based on Consistency Evaluation of Rigid Transformation of Correspondences

Shingo YAMADA and Ikuko SHIMIZU
Dept. of Computer and Information Sciences,
Tokyo University of Agriculture and Technology, Japan.

Abstract

In this paper, a globally convergent algorithm for registration of two range images is proposed which finds the consistent combinations of the corresponding feature point pairs with largest similarity by formulating as graph-based optimization problem. Our algorithm evaluates the consistency of rigid transformations between triplets of feature point pairs. While the number of the all possible triplets of feature point pairs are very large, we reduce the candidates of the triplets based on the shape consistency, one-to-one correspondence assumption, and disntance consistency of all pairs of feature points. By introducing the graph kernel algorithm, the globally optimal combination of the triplets of feature point pairs is found by evaluating the similarity robustly using SIFT.

1 Introduction

3D shape modeling of the real world objects is important in various areas, such as architecture, manufacture, archaeology, and so on. To capture the 3D shape of the real world objects, range sensors are commonly used and images captured by range sensors are called range images. One range image contains only partial shape of the object because an range image is captured by observing the object from only one direction. To obtain the entire shape of the observed object, many range images are captured from different viewpoints. However, each range image is expressed in the coordinate system depending to the position and the orientation of the range sensor. Therefore, all range images need to be transformed to be represented in a common coordinate system. Estimating the rigid transformation between range images using the commonly observed region of the object in each range image is called range image registration.

In general, the ICP (Iterative Closest Point) algorithm by Besl and McKay [1] and Zhang [15], respectively, and its extensions [10][13] are very widely used for range image registration. However, they require sufficiently good initial values of the rigid transformation because they minimize the cost functions with many local minima by nonlinear optimization. Therefore, to obtain good initial values of the rigid transformation automatically, many coarse registration methods have been proposed in the literature.

The coarse registration methods usually based on the invariant features under rigid transformation, such as spin images [5], a point signature [2], a spherical attribute image [4], differential invariants [8], Color Cubic Higher-order Local Auto-Correlation (Color-CHLAC) [6] and so on. These methods are effective if the features can be sufficiently discriminated and their values can be accurately calculated. However, they matched features heuristically and could not guarantee global optimality.

On the other hand, Sakakibara et al. [11] proposed an algorithm by formulating range image registration problem using mixed integer linear programming. This method guarantee the global optimality and the combination of point pairs are consistent in the sense of rigid transformation. However, the computational time of this method was very large.

In addition, registration has been formalized as a discrete optimization task of finding the maximum strict sub-kernel in a graph [12][14]. The method proposed by Enqvist [3] also based on similar idea. This method finds the consistent pairs of feature points with the largest similarity. While the globally optimal solution can be obtained with this method, the solution depends on the quality of the features. However, invariant features such as curvatures are difficult to calculate stably because they are greatly affected by occlusion and discretization of the object surface. In addition, all point pairs in the solution is not necessary to consistent in the sense of rigid transformation because the rigid transformation cannot calculate using only one pair of points.

In this paper, we propose an algorithm for globally convergent registration of two range images by evaluating consistency of rigid transformation using both geometric and photometric features by using range images and the corresponding grayscale images, which are captured simultaneously with range images. Our algorithm guarantees the consistency of rigid transformations between triplets of feature point pairs. While the number of the all possible triplets of feature point pairs are very large, we reduce the candidates of the triplets based on the shape consistency, one-to-one correspondence assumption, and distance consistency of all pairs of feature points. By introducing the graph kernel algorithm, the globally optimal combination of the triplets of feature point pairs is found by evaluating the similarity. To evaluate the similarity robustly, the SIFT (Scale Invariant Feature Transform) [9], which is known as a robust photometric feature, is used.

2 Overview of the proposed algorithm

First, our algorithm selects feature points based on the SIFT [9] using each grayscale image. Next, the feature point pairs between range images are generated by evaluating the shape index [7] of each feature point, which are used to reduce the candidates of the triplets of feature point pairs. The shape index is roughly reflects the local shape around the point, such as convex,

Table 1. Three shape patterns classified by shape index s.

shape pattern	value of the shape index
Convex surface	$s < -\delta$
Planar surface	$-\delta \leq s \leq \delta$
Concave surface	$\delta < s$

concave, or flat. Therefore, our algorithm generate all possible feature point pairs between images and then check the shape indeces of pairs. If they are different, feature pairs are eliminated because their local shapes are not similar and they are not supposed to be the measurements of the same object point.

Then, consistent sets of feature point pairs are generated by constructing an unoriented graph. In this step, the feature point pairs are checked based on one-to-one correspondence assumption and distance consistency. By checking the consistency among feature point pairs, the number of triplets of pairs are much reduced in the next step.

Then, all triplets of pairs are generated by selecting three pairs from the same consistent set of feature point pairs. Note that the pairs which belong to the different consistent sets are not in the same triplet because they are assumed to be inconsistent.

Finally, our algorithm uniquely determine the largest consistent triplets of pairs by the graph-based optimization applying the strict sub-kernel(SSK) algorithm [12] based on the rigid transformation consistency. Our algorithm guarantees that the all corresponding point pairs are consistent in the sense of rigid transformation and the number of consistent pairs are the largest among all the possible combinations of point pairs.

3 Feature point extraction

First, feature points are extracted by a photometric feature, SIFT, using the grayscale image corresponding to each range image. The position of the a feature point can be located in the range image based on the grayscale image because a pixel in a range image and its corresponding grayscale image reflect the same position of the object. In addition to SIFT, shape index of the feature point, which is a geometric feature, is calculated using range image for checking the consistency of local shape roughly between feature point pairs.

Shape index s ($0 \leq s \leq 1$) is calculated from neighboring pixels of the feature point. It is calculated by principal curvatures k_1 and k_2 ($k_1 \geq k_2$) as $s = \arctan \frac{k_1+k_2}{k_1-k_2}$. Because it is difficult to calculate the differential features of the measured point in range image, the value of the shape index is not reliable in usual. Therefore, we do not use its value itself but use its sign which reflects local shape roughly. In concrete, we classify the local shape of each feature point into three patterns: convex, planar, and concave using its shape index. The shape index of almost planar surface is around 0, while those of the convex or concave surface are positive or negative, respectively, as shown in Tab.1.

4 Extraction of the sets of consistent feature point pairs

We extract sets of consistent feature point pairs using an unoriented graph $\mathcal{D}_1$. This graph $\mathcal{D}_1$ is constructed in the same way as the unoriented graph in [12]. Note that $\mathcal{D}_1$ is constructed only for extracting sets of the consistent point pairs, and our algorithm does not optimize this graph $\mathcal{D}_1$ but optimize the graph $\mathcal{G}$ which is constructed from the triplets of feature points in Sec.5.

Suppose that feature points $\boldsymbol{x}_i (i = 1, \dots N_1)$ and $\boldsymbol{y}_j (j = 1, \dots N_2)$ are extracted from two range images, respectively. The number of all possible combinations of feature points pairs is $N_1 \times N_2$. We remove the apparently wrong pairs using the shape patterns based on the shape index as shown in Tab.1 because the feature points on the concave surface could not correspond to the feature points on the convex surface. We generate unoriented graph $\mathcal{D}_1 = (P_1, E_1)$, whose vertex set P_1 represents putative correspondences $p = (\boldsymbol{x}_i, \boldsymbol{y}_j)$. Note that $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ should have the same shape pattern.

There are two kinds of edges in the set E_1: one-to-one correspondence assumption and distance constraint.

Edges based on the one-to-one correspondence assumption joins two vertices are generated as follows: If point $\boldsymbol{x}_1$ is matched to $\boldsymbol{y}_1$ then $\boldsymbol{x}_1$ cannot be matched to any other point in $\boldsymbol{y}_j (j = 2, \dots, N_2)$. Hence, vertices $p = (\boldsymbol{x}_1, \boldsymbol{y}_1)$ and $q_i = (\boldsymbol{x}_i, \boldsymbol{y}_1)$, $i \neq 1$ are connected by edges and $p = (\boldsymbol{x}_1, \boldsymbol{y}_1)$ and $s_j = (\boldsymbol{x}_1, \boldsymbol{y}_j)$, $j \neq 1$ are also connected by edges.

Edges based on the distance consistency joins two vertices are generated as follows: If point $\boldsymbol{x}_i$ is matched to $\boldsymbol{y}_j$ and $\boldsymbol{x}_k$ is matched to $\boldsymbol{y}_l$, the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_k$, $\|\boldsymbol{x}_i - \boldsymbol{x}_k\|$, must be almost same as the distance between $\boldsymbol{y}_j$ and $\boldsymbol{y}_l$, $\|\boldsymbol{y}_j - \boldsymbol{y}_l\|$. If the distances are differ, the two vertices are connected.

Then, we find the consistency sets of pairs as the vertices in P_1 which are not connected each other. In the next step, we select the triplets of pairs which belongs to the same consistency set.

5 Correspondences of the triplets of feature point pairs

Our method generate triplets of feature point pairs because the rigid transformation can be calculated and evaluated using triplets of pairs. In other word, we generate the triangle-to-triangle correspondences based on the point-to-point correspondences.

We generate the second unoriented graph $\mathcal{D}_2 = (P_2, E_2)$, whose vertex set P_2 represents all combinations of triplets of feature point pairs $r_i = (p_i(1) = (\boldsymbol{x}_{j1}, \boldsymbol{y}_{k1}), p_i(2) = (\boldsymbol{x}_{j2}, \boldsymbol{y}_{k2}), p_i(3) = (\boldsymbol{x}_{j3}, \boldsymbol{y}_{k3}))$ where $p_i(j)$ belongs to the same consistent set of pairs in P_1. In this step, the congruence of the triangles $(\boldsymbol{x}_{j1}, \boldsymbol{x}_{j2}, \boldsymbol{x}_{j3})$ and $(\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}, \boldsymbol{y}_{k3})$ is checked. It they are not congruent, the vertex is eliminated. We also eliminate the vertex whose rotation angle is too big because the measured regions from two view points should be overlapped each other.

There are also two kinds of edges in the set E_2: one-to-one correspondence assumption and geometric consistency. The one-to-one correspondence assumption

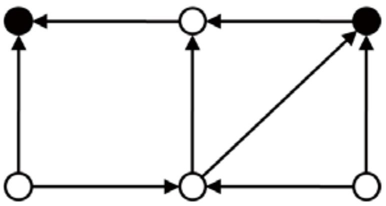

Figure 1. Example of SSK. The black vertices are the SSK for this graph.

Figure 2. A bear model and a chameleon models.

Figure 3. SIFT feature points of the bear model.

Figure 4. SIFT feature points of the chameleon model.

means the same as the first graph. The geometric consistency means whether they are consistent with the rigid transformation. The rigid transformation for the vertex r and that for s are different, they are inconsistent in the sense of the rigid transformation.

Then, we orient the edges of the graph $\mathcal{D}_2$ using the SIFT [9] by evaluating the similarity and obtain the oriented graph $\mathcal{G} = (P, A \cup A^*)$. A^* is the set of irreversible edges, A is the set of reversible edges, and $A \cap A^* = \emptyset$.

To compare the similarity of two triangle-to-triangle correspondences, we employ SIFT. The dimension of SIFT feature vector is 128 for each point. Because a vertex is a triangle-to-triangle correspondence, we connect three feature vectors simply, and dimension of feature vector become 384. The similarity of a vertex is a norm of this vector whose dimension is 384. If the vertex r and s are connected by edge, the similarity of the SIFT of each vertex's triplets of feature point pairs are calculated. If the similarity of s is much better than that of r, $(r, s) \in A^*$ and the whose direction is r to s. If the similarity of s is as same as that of r, the arc (r, s) is reversible and $(r, s) \in A$.

Finally, the maximum strict sub-kernel in a graph $\mathcal{G}$ is obtained by the SSK algorithm [12]. In Fig.1, SSK of the graph as marked by the black vertices.

6 Experimental Results

To test effectiveness of the proposed method, experimental results using real range images are shown in this section. Range images used in the experiments were captured by VIVID 910 [16]. The CPU of the PC used for this experiments were Intel XeonE5620 2.40GHz with 8GB memory.

Two models shown in Fig.2 are used. The bear model has less photometric feature than the chameleon model. The chameleon model has more complicated shape than the bear model. For each model, two range images are captured from two different viewpoints by rotating models using a turn table. The rotation angle between two viewpoints is 20 degree for both models.

In Fig.3 and 4, extracted SIFT feature points of models are shown. Result of this method using the bear model is shown in Fig.5 and that of the chameleon model is shown in Fig.6. Furthermore, result of ICP method implemented by [17] using results of this method as initial value is shown in Fig.7. In the experiments, $\delta = 0.1$.

For the bear model, 81 and 76 feature points were selected in the respective range images. The number of triplets was only 15 because the inconsistent triplets are reduced in Sec.4. For the chameleon model, 248 and 238 feature points were selected in the respective range images. The number of triplets was only 3166.

The estimated rotation angle of the bear model is 19.52 degree, and that of the chameleon model is 19.53 degree. For both models, the error of the rotation angle by our method is less than 0.5 degree.

For the bear model, computational times of feature point extraction for two range images are 0.59 second and 0.55 second, and that of registration is 0.5 sec. For the chameleon model, computational times of feature point extraction for two range images are 0.59 second and 0.60 second, and that of registration is 63.25 sec. Because the number of feature points of the chameleon model was larger than that of the bear model, the computational time was larger. However, total time was about one minute even for the chameleon model which had a large number of feature points.

From Fig.7, our algorithm provided sufficient good

Figure 5. Result by our method of the bear model.

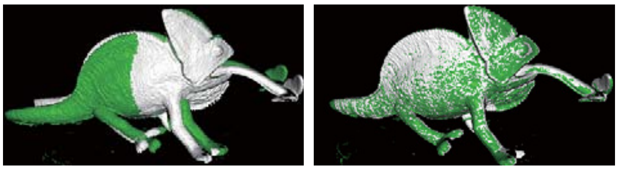

Figure 6. Result by our method of the chameleon model.

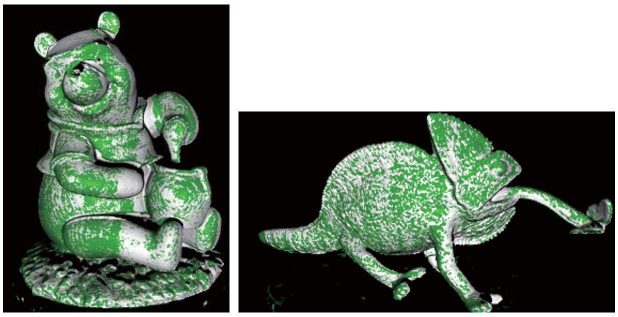

Figure 7. Results by ICP method with intial values by our method.

initial values for ICP algorithm. The rotated angles by ICP for the bear model and the chameleon model are 0.29 degree and 0.30 degree, respectively.

7 Conclusions

In this paper, we proposed a range image registration algorithm which guarantees the globally convergence based on the evaluation of the consistency of rigid transformation by combining shape and image features.

Feature points are selected based on the SIFT which is known as robust under the viewpoint change. Furthermore, to evaluate consistency of rigid transformation, all combinations of triplets of feature point pairs are generated. Though the number of all possible combinations of triplets are very large, we reduce the candidates using the one-to-one correspondences by roughly evaluating local shape patterns based on the shape index. Our algorithm finds the consistent triplets of pairs in the sense of the rigid transformation with the largest similarity by the SSK algorithm.

We applied our methods to real range images, and the effectiveness was shown. However, in the case that many feature points are extracted, the total of nodes increases and the computational time will increase. Our feature works include that to reduce spatial and temporal computational cost by reducing the candidates of feature point pairs using other features.

References

[1] P. J. Besl and N. D. McKay. A Method for registration 3-D shapes. *IEEE Trans. on PAMI*, 14(2):239–56, 1992.

[2] C. S. Chua and R. Jarvis. 3D free-form surface registration and object recognition. *IJCV*, 17(1):77–99, 1996.

[3] O.Enqvist, K. Josephson and F. Kahl. Optimal correspondences from pairwise constraints In *Proc. of ICCV*, 2009.

[4] K. Higuchi, M. Hebert, and K. Ikeuchi. Building 3-D models from unregistered range images. *GMIP*, 57(4):315–33, 1995.

[5] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. on PAMI*, 21(5):433–49, 1999.

[6] A. Kanezaki, T.Harada, and Y. Kuniyoshi. Partial matching of real textured 3D objects using color cubic higher-order local auto-correlation features. *Visual Computing*, 26:1269–1281, 2010.

[7] J. J. Koenderink, *Solid Shape*, MIT Press, 1990.

[8] P. Krsek, T. Pajdla, and V. Hlavac. Differential invariants as the base of triangulated surface registration. *CVIU*, 87:27–38, 2002.

[9] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[10] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. of 3DIM*, pages 145–52, 2001.

[11] S. Sakakubara, Y. Kounoike, Y. Shinano and I. Shimizu. Automatic Range Image Registration Using Mixed Integer Linear Programming. In *Proc. ACCV2007, Lecture Note in Computer Science*, pages 424–434, 2007.

[12] R. Sara, I. Shimizu Okatani and A. Sugimoto. Globally Convergent Range Image Registration by Graph Kernel Algorithm. In *Proc. of 3DIM*, pages 377–384, 2005.

[13] J. Salvi, C. Matabosch, D. Fofi and J. Forest. A review of recent range image registration methods with accuracy evaluation. IVC, 25(5):578–596, 2007.

[14] I. Shimizu, A. Sugimoto and R. Sara. Graph-Based Range Image Registration Combining Geometric and Photometric Features. In *Proc. of SCIA*, pages 542–552, 2007.

[15] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *IJCV*, 13(2):119–52, 1994.

[16] VIVID910 spec: `http://www.konicaminolta.jp/instruments/products/3d/vivid910/spec.html`

[17] Scanalyze: `http://www.graphics.stanford.edu/software/scanalyze`