

AUTOMATIC MULTIVIEW FINE REGISTRATION OF RANGE IMAGES ACQUIRED FROM UNKNOWN VIEW POINTS

Ajmal S. Mian, Mohammed Bennamoun and Robyn A. Owens

School of Computer Science and Software Engineering
The University of Western Australia
35 Stirling Highway, Crawley, WA 6009, Australia
{ajmal, bennamou, robyn}@csse.uwa.edu.au

ABSTRACT

In this paper, we present a fully automatic multiview fine registration algorithm capable of registering range images acquired from unknown view points. Our algorithm represents each range image with multiple tensors and indexes these tensors by a 4D hash table. Tensors of each view are simultaneously matched with the tensors of the remaining views by casting votes using the hash table and calculating the correlation coefficient between the pair of tensors with the highest votes. Matching pair of tensors are then used for the coarse registration of their corresponding range images. A graph building algorithm registers all the views in a common coordinate basis. Finally, a multiview fine registration algorithm is used in a hierarchical fashion to iteratively refine the initial coarse registration. Thus our algorithm avoids the use of a pairwise fine registration algorithm. Our results show that our algorithm is fully automatic and accurate.

1. INTRODUCTION

Many reverse engineering, medical diagnostic and robotics applications require geometrically accurate 3D models of free-form objects. Existing range imaging techniques can only acquire a partial surface of a free-form object from a single view point due to self occlusions. To construct a complete 3D model of the object, its multiple range images must be acquired from different view points to cover the entire surface of the object. These range images must then be registered in a common coordinate basis. According to the survey of Campbell and Flynn [5], registration is performed in two steps namely, coarse registration and fine registration. Automatic coarse registration of range images acquired from unknown view points is performed with a feature matching algorithm. This process is also known as automatic correspondence identification. Points on two range images which correspond to the same point on the object are known as corresponding points. These corresponding points are then used to derive a rigid transformation (rotation matrix and translation vector) that aligns the two range images by minimizing the distance error between them. Once the views are coarsely registered, the registration is refined using a pairwise registration refinement algorithm like the ICP [3] for example. In case of multiple range images, the registration errors may accumulate and result in large seams between some of the range images. To avoid this, a global registration algorithm is used after pairwise fine registration which distributes the registration errors evenly over the entire 3D model. Once the range images are registered they are integrated and reconstructed to form a single smooth surface.

We propose a two step algorithm which automatically performs the coarse registration followed by the fine registration of multiple range images of an object acquired from unknown view points. Our algorithm does not require the manual identification of corresponding points or the initial estimates of registration. Moreover, no information regarding which range image overlaps which other range image is required (i.e. the views are unordered). This is in contrast to existing multiview fine registration algorithms (see Section 2) which assume that the range images have already been pairwise registered. The novelty of our approach is that it does not use a pairwise fine registration algorithm at any stage. See Fig. 1 for a comparison between our approach and the traditional approach.

The rest of this paper is organized as follows. Section 2 gives a brief review of existing work in the area of pairwise and multiview registration. Section 3 gives an overview of our approach. Section 4 explains our tensor representation. Section 5 describes the hash table construction. Section 6 contains the details of our automatic multiview registration algorithm. Qualitative and quantitative results are given in Section 7. Finally, conclusions are given in Section 8.

2. EXISTING WORK

Existing work in the area of automatic pairwise coarse registration include the RANSAC-based DARCES (Data Aligned Rigidity Constrained Exhaustive Search) algorithm [6], the bitangent curve matching [25], the SAI (Spherical Attribute Image) matching [11], Roth's technique [20] and the spin image [14] matching algorithm. These algorithms are pairwise coarse registration algorithms and must perform an exhaustive search for correspondences between $N(N-1)/2$ (N is the total number of views) pairs of views in case they are unordered.

Pairwise fine registration algorithms include the classic ICP [3] algorithm and its many variants for example [21]. These algorithms iteratively establish correspondences between the nearest points of two views and minimize their mutual distance. Another example is the Chen and Medioni's algorithm [7] which minimizes the point to surface distances instead. Mutual information based approach [19] maximizes the mutual information between two views in order to register them. These fine registration algorithms however assume that the views have already been coarsely registered.

Multiview correspondence (coarse registration) is defined as a one-to-many matching approach in which one view is simultaneously matched with all the remaining views [1]. Existing work

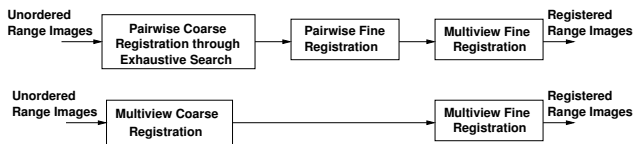


Figure 1: Block diagram comparing the traditional approach (top) with our approach (bottom).

in the area of multiview coarse registration includes [1]. Huber and Hebert [12][13] proposed a framework for automatic 3D modeling based on a claimed automatic multiview surface matching algorithm. Their algorithm however relies on an off-the-shelf automatic pairwise correspondence algorithm (i.e. the spin image matching [14]) to exhaustively search for correspondences between $N(N-1)/2$ pairs of views in order to initialize a graph of relative pose estimates. This process has a computational complexity of $O(N^2)$. The exhaustive search is followed by the claimed multiview surface matching algorithms which search for discrete edges (pairwise correspondences) in the graph and then verify them after global registration. This approach is not equivalent to multiview correspondence according to the definition in [1] since no simultaneous matching is performed. The computational complexity of this framework also adds an additional overhead to the initial exhaustive search.

A brief review of multiview fine registration algorithms includes Pulli's [15] algorithm, Williams and Bennamoun's algorithm [24], and Benjemma and Schmitt's algorithm [2]. As with the pairwise fine registration algorithms, these algorithms assume that the views have been coarsely registered or the correspondences between overlapping views have been identified through external means.

3. OVERVIEW OF OUR APPROACH

First, we use a variant of our multiview feature matching algorithm [1][17] for automatic coarse registration of the range images. The major differences are: (a) This variant does not use a pairwise fine registration algorithm at any stage (b) It does not try to make a minimum level spanning tree as opposed to [1] but tries to connect those views which have the best correspondence (see Section 6 for details). Once the views are coarsely registered, a multiview fine registration algorithm [15] is used to iteratively refine this coarse registration. A hierarchical approach is used during fine registration as follows. First, multiview fine registration is applied to the decimated meshes. Next, these multiview registration results are used to align the high resolution meshes and the multiview fine registration algorithm is repeated.

4. TENSOR REPRESENTATION

The range images (views) of an object in the form of point clouds are first converted into triangular meshes. Next, each mesh is decimated twice using Garland's algorithm [9] to form a hierarchy of three meshes. The lowest mesh in the hierarchy (the coarsest mesh) is used to select feature points in the next higher mesh which is used to compute tensors. The highest resolution mesh in the hierarchy (M_{hi} where h stands for high resolution and $i = 1 \dots N$) is used for multiview registration refinement and the final model construction. Each M_{hi} is decimated [9] once to 800 faces per

mesh (to get medium resolution meshes M_{mi}) and a second time to 400 faces per mesh (to get low resolution meshes M_{li}). Next, the closest vertices in M_{mi} to M_{li} are identified. These vertices of M_{mi} are paired together according to an angle constraint and a distance constraint. The angle constraint ensures that only those vertices whose normals make an angle (θ_d) greater than 5° are paired. The distance constraint ensures that only those vertices are paired which are within a certain distance (d_{min} and d_{max}) of each other. In our experiments we selected $d_{min} = 4b_s$ and $d_{max} = 6b_s$ (b_s is defined below).

Despite the above constraints the number of vertex pairs is generally large therefore 400 pairs are randomly selected. Each of these vertex pairs is then used to define a 3D coordinate basis as follows. The midpoint of the two vertices defines the origin and the average of their normals defines the z -axis. x -axis is defined by the cross product of the normals of the vertices. Finally, the y -axis is defined by the cross product of the z -axis and the x -axis. Each of these 3D coordinate bases is used to define a $15 \times 15 \times 15$ grid centered at its origin. The size of the bins b_s is selected such that the grid completely encloses the mesh. This is in contrast to our previous variants [1][18] in which each 3D grid enclosed only a local surface patch of a mesh as they further rely on ICP algorithm [3] for registration refinement. In our experiments we selected $b_s = D/15$ where D is the average dimension of the meshes.

Once the 3D grid is defined, the area of intersection of the mesh with its bins is recorded in a third order tensor. The value of each element of the tensor is equal to the surface area intersecting its corresponding bin in the 3D grid. Since most bins of the grid are likely to be empty, the resultant tensor has many zero elements. Therefore, in order to save memory, each tensor is compressed by eliminating all zero elements and retaining the non-zero elements and their index positions.

5. HASH TABLE CONSTRUCTION

The tensors of the meshes M_{mi} are indexed by a 4D hash table. Three dimensions of the hash table correspond to the i, j, k indices of the tensors whereas the fourth dimension corresponds to the θ_d of the tensor. θ_d is quantized at intervals of 5° . The hash table is filled up as follows. For each tensor of a mesh, the entry (mesh number, tensor number) is made in all the bins of the hash table corresponding to the i, j, k indices of the nonzero elements of the tensor and the θ_d of the tensor.

6. AUTOMATIC MULTIVIEW REGISTRATION

During multiview coarse registration, the tensors (T_r reference tensor) of each mesh (M_{mr} reference mesh) are used to cast votes to the tensors of the remaining meshes as follows. The i, j, k indices of the nonzero elements of the tensor and its θ_d are used to cast votes to all the tuples (mesh number, tensor number) present at i, j, k, θ_d index position in the hash table. To avoid redundant voting, a forward voting approach is used in which the tensors of a mesh M_{mr} can cast votes to the tensors of a mesh M_{mj} such that $r < j$ ($r = 1 \dots N-1$ and $j = 2 \dots N$). Tuples that receive less votes than half the number of nonzero elements of T_r are discarded.

Next, a linear correlation coefficient is calculated between T_r and each of its remaining corresponding tensors T_c . To cater for occlusions, the linear correlation coefficient is calculated in only

the overlapping region of the two tensors. For each calculation, a vector \mathbf{v} is extracted from \mathbf{T}_r which comprises only of those nonzero tensor elements which have a corresponding nonzero element in \mathbf{T}_c . A vector \mathbf{u} is extracted from \mathbf{T}_c in a similar way. Note that both \mathbf{u} and \mathbf{v} will have length equal to the number of intersecting elements n_v between the two tensors. The linear correlation coefficient C_c is then calculated between the two vectors using Eqn. 1.

$$C_c = \frac{n_v \sum v_i u_i - \sum v_i \sum u_i}{\sqrt{n_v \sum v_i^2 - (\sum v_i)^2} \sqrt{n_v \sum u_i^2 - (\sum u_i)^2}} \quad (1)$$

In Eqn. 1 all summations are from $i = 1 \dots n_v$. C_c is then multiplied by n_v to find the similarity S between the two tensors ($S = n_v C_c$). This information (\mathbf{M}_{mr} number, \mathbf{T}_r number, \mathbf{M}_{mc} number, \mathbf{T}_c number and S) is then recorded in a rank table.

Once the voting process is over, the rank table contains a complete list of matching pairs of tuples along with the similarity between them. The rank table is sorted according to the decreasing value of S and is then fed to a graph building algorithm which makes a spanning tree graph from it whose nodes represent the meshes (views) and arcs represent the rigid transformation between its end nodes. The graph building algorithm picks pairs of corresponding tensors starting from the one with the maximum value of S , and uses them to calculate the rigid transformation between their corresponding meshes using Eqn. 2 and Eqn. 3.

$$\mathbf{R} = \mathbf{B}_c^T \mathbf{B}_r \quad (2)$$

$$\mathbf{t} = \mathbf{O}_r - \mathbf{O}_c \mathbf{R} \quad (3)$$

In Eqn. 2, \mathbf{B}_r and \mathbf{B}_c are the 3×3 matrix of x, y, z coordinate vectors of \mathbf{T}_r and \mathbf{T}_c respectively. In Eqn. 3, \mathbf{O}_r and \mathbf{O}_c are the 1×3 vectors of coordinates of the origins of \mathbf{T}_r and \mathbf{T}_c respectively. \mathbf{R} and \mathbf{t} are the 3×3 rotation matrix and 1×3 translation vector that aligns the two coordinate bases and hence the corresponding mesh with the reference mesh. The transformation is used to initialize the spanning tree with two nodes and an arc. The graph building algorithm then searches for another pair of corresponding tensors (with maximum value of S) which connect another node (mesh) to the spanning tree. A rigid transformation is calculated from these tensors (Eqn. 2 and Eqn. 3) and used to connect another node to the spanning tree. This process continues until all the meshes have been added to the spanning tree. Each time a new mesh is added to the spanning tree, the transformation is verified by checking if vertices of the corresponding mesh are transformed into the active space of the sensor (free space of the sensor or the space between the reference mesh and the sensor) or not. The new mesh is added to the spanning tree only if the ratio of such vertices is negligible. The spanning tree is then used for the coarse registration of all the meshes by concatenating transformations.

Once the views are coarsely registered, Pulli's multiview fine registration algorithm [15] is used to iteratively refine the coarse registration. A hierarchical approach is used during multiview fine registration. First, multiview fine registration [15] is applied to the decimated meshes \mathbf{M}_{mi} . These multiview registration results are then used to align the high resolution meshes \mathbf{M}_{hi} and the multiview registration algorithm is repeated. The hierarchical approach has two advantages: (a) It ensures quick convergence of the multiview registration algorithm. (b) It minimizes the chances that the

algorithm will get stuck in a local minimum. It may be noted that in our automatic multiview registration algorithm described above, no pairwise registration refinement has been used. The multiview coarse alignments were used as constraints in Pulli's algorithm [15] as opposed to pairwise fine alignments¹.

The multiview fine registration is followed by integration and reconstruction to complete the 3D model. We used VripPack [22] for this purpose which uses the volumetric integration algorithm by Curless and Levoy [8] for integration and the marching cubes algorithm [16] for reconstruction.

7. RESULTS

We performed two types of experiments. The first experiment was performed on real data i.e. range images of free-form objects namely the angel and the bird [4] acquired with a Minolta 3D scanner. 19 views of the angel and 18 views of the bird were separately fed to our algorithm without any order for multiview registration. Fig. 2 and Fig. 3 show the qualitative results of our algorithm. For quantitative comparison, the same views were also registered using the traditional approach (top block diagram of Fig. 1). The coarse registration was performed manually in this case. A variant of the ICP algorithm [21] was used for pairwise fine registration and Pulli's algorithm [15] was used for multiview fine registration. Since the ground truth was unknown in this case, we compared the rotational misalignment between the views when registered using our approach and when registered using the traditional approach (see Eqn. 4).

$$\Delta\theta_i = \cos^{-1} \left(\frac{\text{trace}(\mathbf{R}_{i1} \mathbf{R}_{i2}^{-1}) - 1}{2} \right) \times \frac{180}{\pi} \quad (4)$$

In Eqn. 4, \mathbf{R}_{i1} is the rotation matrix (that registers view i) calculated using the traditional approach and \mathbf{R}_{i2} is a similar matrix but calculated using our approach. $\Delta\theta_i$ is the rotational error between \mathbf{R}_{i1} and \mathbf{R}_{i2} . Fig. 4(a) shows the $\Delta\theta_i$ for each of the 19 views of the angel whereas Fig. 4(b) shows $\Delta\theta_i$ for each of the 18 views of the bird. Note that the rotational error (difference) between the two approaches is less than 1° in each case. The translation error between the two approaches was calculated using Eqn. 5.

$$\Delta t_i = \frac{\|\mathbf{t}_{i1} - \mathbf{t}_{i2}\|}{d_r} \quad (5)$$

In Eqn. 5, \mathbf{t}_{i1} and \mathbf{t}_{i2} are the translation vectors calculated using the two approaches of Fig. 1. The translation error is normalized with respect to resolution d_r of the mesh to make it scale independent. Fig. 4(c) and Fig. 4(d) show the translation errors (difference) in the views of the angel and the bird respectively. These results indicate that the errors (difference) in translation of the views are smaller than the resolution of the meshes.

The above experiments clearly indicate that there is little difference when the range images are registered using our approach and when using the traditional approach. However, our approach has two advantages. First, it can register unordered range images using multiview surface matching and without going into exhaustive. This makes our approach more efficient. Second, our approach does not require pairwise registration refinement which

¹The actual Pulli's algorithm [15] uses pairwise fine alignments as constraints.

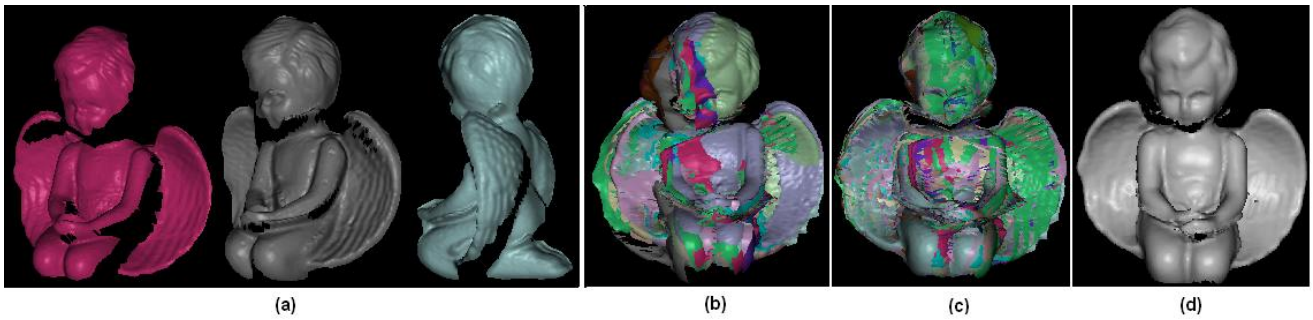


Figure 2: (a) Three example views of the angel. (b) Multiview coarse registration results of our algorithm for 19 views of the angel. (c) After multiview fine registration using Pulli's algorithm [15]. (d) The complete 3D model after integration and reconstruction using VripPack [22]. All models are rendered in Scanalyze [23]. This figure is best viewed in colour.

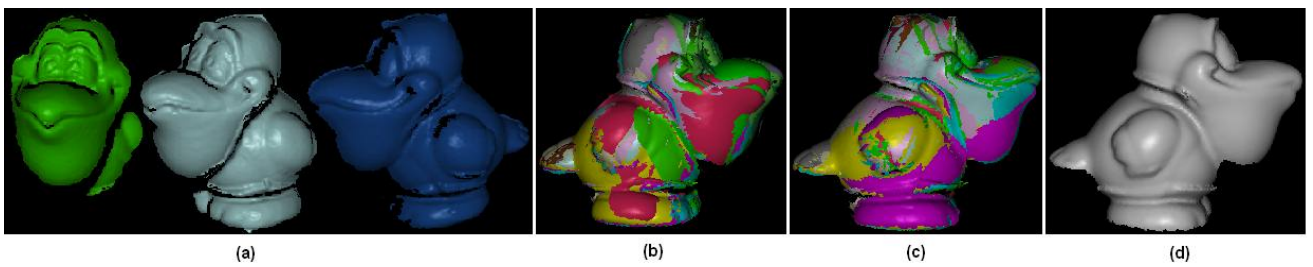


Figure 3: (a) Three example views of the bird. (b) Multiview coarse registration results of our algorithm for 18 views of the bird. (c) After multiview fine registration using Pulli's algorithm [15]. (d) The complete 3D model after integration and reconstruction using VripPack [22]. All models are rendered in Scanalyze [23]. This figure is best viewed in colour.

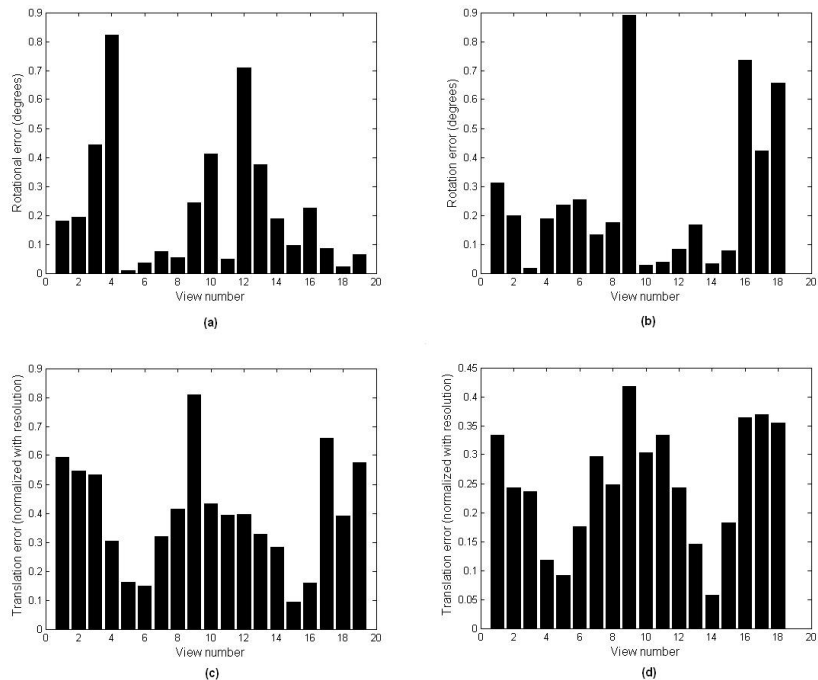


Figure 4: Rotation and translation errors (differences) when the real range images are registered using our approach and when using the traditional approach as shown in Fig. 1. (a) Rotation error of angel. (b) Rotation error of bird. (c) Translation error of angel. (d) Translation error of bird.

makes it simpler and further enhances its efficiency. Due to the unavailability of the ground truth, it cannot be determined as to which approach gives more accurate results. In our next experiment, we generated synthetic range data by rendering a 3D model of a dog (data courtesy of the University of Stuttgart [10]) from different views and reconstructing its visible surface using VripPack [22]. With this approach, we generated 30 synthetic range images of the dog (see Fig. 5) and recorded the ground truth rotations and translations that would register each view to a known reference. Next, these 30 views were registered once using our approach and a second time using the traditional approach. The rotation matrices and translation vectors calculated with the two approaches were then compared to the ground truth using Eqn. 4 and Eqn. 5. Fig. 6 and Fig. 7 shows the rotational and translation error respectively for each of the 30 views of the dog. These results show that both approaches have rotation errors of less than 1° and the translation errors of less than half the resolution of the meshes. The mean rotation error was 0.26° for our approach and 0.28° for the traditional approach. The mean translation error was 0.14 in both cases. Therefore, we can conclude that both approaches provide approximately the same accuracy.

8. CONCLUSION

We presented a fully automatic multiview registration algorithm that performs fine registration of range images acquired from unknown view points. We did not use a pairwise fine registration algorithm in our approach and presented quantitative results which clearly indicate that the accuracy of our algorithm is equivalent to the traditional approach which uses a pairwise fine registration algorithm. Therefore, our approach makes the 3D modeling process simpler without any compromise on accuracy.

9. ACKNOWLEDGMENTS

The authors would like to thank Stanford University for providing Scanalyze and VripPack code and Carnegie Mellon University for providing the mesh simplification software. The authors would also like to thank The Ohio State University and The University of Stuttgart for providing range data. This research is sponsored by ARC grant number DP0344338.

10. REFERENCES

- [1] A. S. Mian and M. Bennamoun and R. A. Owens. Automatic Multiview Coarse Registration for 3D Modeling. In *IEEE Conference on Cybernetics and Intelligent Systems*, December 2004.
- [2] R. Benjemma and F. Schmitt. Fast Global Registration of 3D Sampled Surfaces Using a Multi-Z-Buffer Technique. In *International Conference on Recent Advances in 3D Digital Imaging*, pages 113–120, 1997.
- [3] P. J. Besl and N. D. McKay. Reconstruction of Real-world Objects via Simultaneous Registration and Robust Combination of Multiple Range Images. *IEEE TPAMI*, 14(2):239–256, February 1992.
- [4] R. Campbell and P. Flynn. A WWW-Accessible 3D Image and Model Database for Computer Vision Research. pages 148–154. K. W. Bowyer and P. J. Phillips (eds.), IEEE Computer Society Press, 1998.

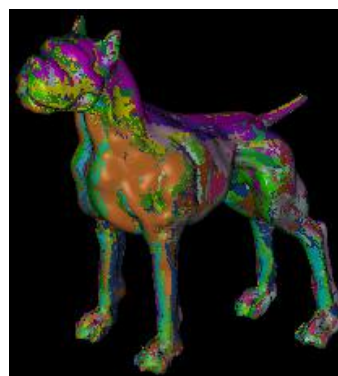


Figure 5: 30 synthetically generated views of the dog have been registered with our algorithm.

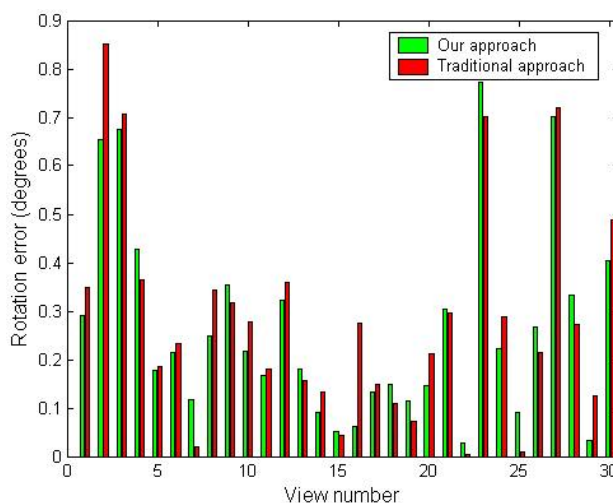


Figure 6: Ground truth rotation errors of the two registration approaches for the 30 views of the dog.

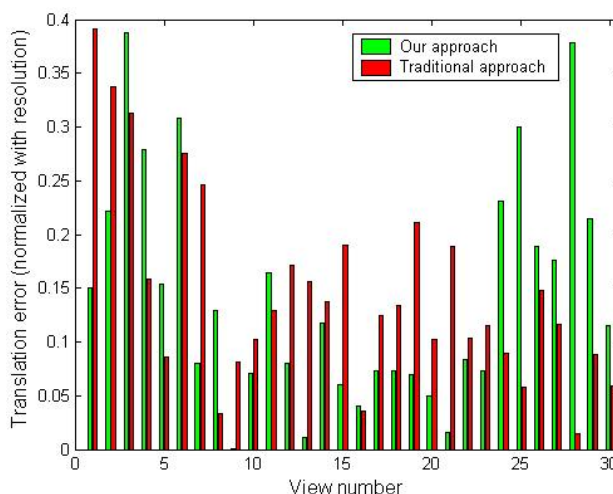


Figure 7: Ground truth translation errors of the two registration approaches for the 30 views of the dog.

- [5] R. J. Campbell and P. J. Flynn. A Survey of Free-Form Object Representation and Recognition Techniques. *Computer Vision and Image Understanding*, 81(2):166–210, 2001.
- [6] C. Chen, Y. Hung, and J. Cheng. RANSAC-Based DARCES: A New Approach to Fast Automatic Registration of Partially Overlapping Range Images. *IEEE TPAMI*, 21(11):1229–1234, 1991.
- [7] Y. Chen and G. Medioni. Object Modeling by Registration of Multiple Range Images. In *IEEE ICRA*, pages 2724–2729, April 1991.
- [8] B. Curless and M. Levoy. A Volumetric Method for Building Complex Models from Range Images. In *SIGGRAPH*, 1996.
- [9] M. Garland and P. S. Heckbert. Surface Simplification using Quadric Error Metrics. In *SIGGRAPH*, pages 209–216, 1997.
- [10] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3D Object Recognition from Range Images using Local Feature Histograms. In *IEEE CVPR*, volume 2, pages 394–399, 2001.
- [11] K. Higuchi, M. Hebert, and K. Ikeuchi. Building 3-D Models from Unregistered Range Images. In *IEEE ICRA*, volume 3, pages 2248–2253, May 1994.
- [12] D. Huber and M. Hebert. 3D Modeling Using a Statistical Sensor Model and Stochastic Search. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 858–865, 2003.
- [13] D. F. Huber. *Automatic Three-dimensional Modeling from Reality*. PhD thesis, The Robotics Institute, Carnegie Mellon University, USA, July 2002.
- [14] A. E. Johnson and M. Hebert. Surface Registration by Matching Oriented Points. In *Int. Conf. on Recent Advances in 3-D Imaging and Modelling*, pages 121–128, 1997.
- [15] K. Pulli. Multiview Registration for Large Datasets. In *3D Digital Imaging and Modeling*, 1999.
- [16] W. E. Lorensen and H. E. Cline. A High Resolution 3D Surface Construction Algorithm. In *ACM SIGGRAPH*, pages 163–169, 1987.
- [17] A. S. Mian, M. Bennamoun, and R. A. Owens. A Novel Algorithm for Automatic 3D Model-based Free-form Object Recognition. In *IEEE SMC*, October 2004.
- [18] A. S. Mian, M. Bennamoun, and R. A. Owens. Matching Tensors for Automatic Correspondence and Registration. In *ECCV*, volume 2, pages 495–505, 2004.
- [19] A. Rangarajan, H. Chui, and J.S. Duncan. Rigid point feature registration using mutual information. *Medical Image Analysis*, 3(4):425–440, 1999.
- [20] G. Roth. Registering Two Overlapping Range Images. In *3DIM*, pages 191–200, 1999.
- [21] S. Rusinkiewicz and M. Levoy. Efficient Variants of the ICP Algorithm. In *3DIM*, pages 145–152, 2001.
- [22] Stanford Computer Graphics Laboratory. A Volumetric Range Image Processing Package. <http://graphics.stanford.edu/software/vrip/>, Aug 2001.
- [23] Stanford Computer Graphics Laboratory. Scanalyze. <http://graphics.stanford.edu/software/scanalyze/>, September 2004.
- [24] J. Williams and M. Bennamoun. Simultaneous Registration of Multiple Corresponding Point Sets. *CVIU*, 81(1):117–142, 2001.
- [25] J. V. Wyngaerd, L. V. Gool, R. Koth, and M. Proesmans. Invariant-based Registration of Surface Patches. In *IEEE ICCV*, volume 1, pages 301–306, 1999.