Automatic multi-view surface matching

Simone Fantoni[1], Umberto Castellani[1], Andrea Fusiello[1,2]

[1]University of Verona, Dipartimento di Informatica [2]University of Udine, DIEGM

Abstract

In this paper we tackle the problem of automatically aligning an unordered set of range views. We propose a full pipeline that goes from the scans to the complete 3D model. The emphasis is on the automation – no manual intervention is require – and on the fact that no knowledge on the acquisition sequence is assumed. The contribution is twofold: in the pre-alignment phase a voting scheme is proposed that discovers the overlapping relationship among views; in the final refinement step we extend the Levenberg Marquardt-ICP to work with multiple views, in order to solve for the absolute pose of all images simultaneously.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Curve, surface, solid, and object representations

1. Introduction

Three dimensional registration of range images acquired by a 3D scanner is still a critical issue for reliably obtaining a complete 3D model of objects or buildings [HH03, BM92, Pul99]. In particular, when several scans (i.e., more than two views) are involved and initial pose estimates are unknown, the problem is called *multiview surface matching*. Three main interrelated sub-problems need to be solved [HH03]: i) determining which views overlap, ii) determining the relative pose between each pair of overlapping views, and iii) determining the absolute pose of the views. Many works have been proposed to address these issues but few of them address all such three sub-problems at the same time. Sub-problem i) aims at improving the automation of the process. The output of this stage can be encoded in a *adjacency* matrix that contains overlap information. Sub-problem ii) is called in general *pairwise* registration. The ICP algorithm [BM92] represents the gold standard for this problem. As observed in [HH03], there is a mutual dependency between the overlap and relative poses. If the relative poses are known the overlap can be easily computed, and vice-versa. Therefore, in general these two phases are computed in a cooperative fashion. Finally, once the adjacency matrix is known and relative poses are available, Sub-problem iii) is addressed by multiple view registration approaches [Pul99], where the solution of the absolute pose estimations is computed simultaneously for all views. In [HH03] the authors focused mainly on the automation of the matching (i.e., Sub-problem 1) by proposing a graph-based optimization process to determine the best view ordering. Similar approaches exploit *local* point signatures [KCB09, BSL11] to obtain a robust pre-alignment. In [KMSRJ05] proposed to address the automatic alignment as a location-recognition-problem by integrating range data with 2D intensity images. In [AMCO08] a fast and robust technique is introduced for pairwise registration based on the alignment between coplanar and congruent 4-points sets randomly extracted from the two views. To address Sub-problem 3, in [Pul99] a method is proposed that uses the estimated pairwise transformations as constraints in a global multi-view step. More recently, [TBC10] proposed a new global registration framework based on the well-known Generalized Procrustes Analysis which is adapted to implement the ICP algorithm. In this paper we address all the three aforementioned sub-problems by proposing a multiview surface matching pipeline which deals with both automation of the process and accuracy of results. Our pipeline is composed by three stages: i) feature points detection and description, ii) overlap estimation by feature points matching, and iii) multiple view refinement.

2. Keypoint extraction and description

We adopt a *feature*-based approach which is composed of two main phases: i) keypoint extraction, and ii) keypoint description.

2.1. Keypoint extraction

Keypoint extraction aims at detecting few and significative feature points from the shape. To this aim we employ the method proposed in [CCFM08], that consists in three main steps: (i) multiscale representation, (ii) 3D saliency measure definition, and (iii) feature points detection. The multiscale representation is obtained by applying N Gaussian filters on the mesh M^d, obtaining N multidimensional filtering maps $\{F_i^d\}, i = 1, \ldots, N$. The *neighborhood region* of a vertex v, over which the filtering is applied, is built by expanding a n-rings search starting from v, and collecting all those vertices displaced within a distance equal to 2.5σ, where σ is the standard deviation of the Gaussian kernel. The Difference-of-Gaussians (DoG) operator is defined as:

$$F_i^d(v) = g(v, \sigma_i) - g(v, 2\sigma_i) \quad (1)$$

where σ_i is the value of the standard deviation associated to scale i. Six scales of filtering have been fixed, corresponding to standard deviation values $\sigma_i \in \{1\varepsilon, 2\varepsilon, 3\varepsilon, 4\varepsilon, 5\varepsilon, 6\varepsilon\}$, where ε amounts to 0.1% of the length of the main diagonal located in the bounding box of the model. It is worth noting that $F_i^d(v)$ is a 3D vector which denotes how much the vertex v has been moved from its original position after the filtering, and this can be taken as a saliency measure. In order to reduce the displacement vector $F_i^d(v)$ to scalar quantity it is projected to the normal $n(v)$ of the vertex v. In this fashion the *scale* map M_i^d is obtained as:

$$M_i^d(v) = ||n(v) \cdot (g(v, \sigma_i) - g(v, k\sigma_i))||. \quad (2)$$

Each map is then normalized by adopting the Itti's approach [CCFM08] which increases the evidence of the highest peaks. A saliency map is obtained by simply adding the contribution of each scale map. Finally, salient points are obtained as maxima of the saliency map: a point is detected if it is a local maximum and its value is higher than the 30% of the global maximum.

2.2. Keypoint description

Keypoint description aims at attaching a descriptor to each keypoint that must be: i) distinctive of the point, ii) invariant to rigid transformations, and iii) resilient to as much nuisances as possible (noise, clutter, partial views, sampling rate, and so on.). We use spin-images [JH99], a well-known surface representation that have been successfully employed in shape matching and object recognition. There are three parameters that control the generation of spin-images: bin size, support distance, and support angle. Following [JH99] the bin size have been set to 1.5 times the mesh resolution, and the support angle to 75deg. On the contrary, the support distance have been set to 1/10 of the size of the model, in order to force the spin-image to be a local descriptor.

3. View Matching

The objective is to find a set of matching keypoints in any pair of views. Keypoints have already been extracted (Section leading to our 2.1) and a descriptor has been attached (Section 2.2).

The first step is to identify in a computationally efficient way (linear in the number of views n) views that potentially share a good number of keypoints, instead of trying to match keypoints between every view pair, as they they are $O(n^2)$. We follow the approach of [BL03] for 2D image mosaicing. In this *broad phase* we consider only a constant number of descriptors in each view (we used 300, where a typical view contains thousands of keypoints). Then, each keypoint descriptor is matched to its ℓ nearest neighbors in feature space (we use $\ell = 6$). This can be done efficiently by using a k-d tree to find approximate nearest neighbors (we used the ANN library[†]). A 2D histogram is then built that registers in each bin the number of matches between the corresponding views. Then, in the *narrow phase* every view will be matched only to the m views (we use $m = 8$) that have the greatest values in the 2D histogram. Hence, the number of views to match is $O(n)$, being m constant. The output of this phase is a $n \times n$ *preliminary* adjacency matrix $\hat{A}$.

Precise view-to-view matching follows a nearest neighbor approach, with rejection of those keypoints for which the ratio of the nearest neighbor distance to the second nearest neighbor distance is greater than a threshold (set to 1.5 in our experiments). These matches are then use to compute the rigid transform that aligns the view pairs using MSAC [TZ00]. Some view matches can be rejected at this stage, if MSAC fails to compute a valid alignment or if the number n_i of remaining inlier matches between two views is less than a threshold:

$$n_i > 5.9 + 0.22 n_f \quad (3)$$

where n_f is the number of original matches. The derivation of the formula can be found in [BL03].

Finally, the rigid transform between the two views is refined with Iterative Closest Point (ICP) [BM92] on the whole set of points (whereas before we were considering keypoints only), and outlier points are singled out using a robust statistics called X84 [CFM02]. The output of this matching step is a $n \times n$ symmetric matrix A that contains in the entry (i, j) the weight of the matching between view i and view j, where 0 means no matching, and 1 represent a 100% overlap (possible only in case of identical views).

4. Global registration

There are two stages of global registration: first a global alignment is produced by combining the pairwise rigid transformations found in the previous section; then this alignment is refined with a multiview ICP that considers all the views simultaneously (resembling a "bundle adjustment").

[†] Ann library is available at http://www.cs.umd.edu/ mount/ANN

Graph-based alignment. A weighted graph is constructed, whose vertices are the views and edges links overlapping views with weight from the matrix A, which represents the adjacency matrix of the graph. Given a *reference* view chosen arbitrarily, which sets the global reference frame, for each view i, the transformation that aligns it with the reference view r is computed by chaining transformations along the shortest weighted path from i to r. This is equivalent to computing the (weighted) minimum spanning tree (MST) with the reference view as root. The idea (as in [MFM04]) is that this yield a global alignment of the views with the least accumulation error among the solutions based on chaining pairwise registrations.

Multiview LM-ICP. The graph-based alignment can be further improved by defining a global registration schema which estimates all the absolute poses at the same time. In particular, we extend the LM-ICP [Fit03] to deal with multiple views simultaneously.

Let be $V^1, \ldots, V^n$ the set of acquired views. Let be $\mathbf{a}_1, \ldots, \mathbf{a}_n$ the set of parameter vectors which encode the absolute poses. Give the binarized adjacency $\tilde{A}$ whose entries $\tilde{A}(h,k) = 1$ if view V^h can be registered to view V^k, and $\tilde{A}(h,k) = 0$ vice-versa. Therefore, every pair of views (h,k) leads to an aligned error between view V^h and V^h into the global coordinate system:

$$E(\mathbf{a}_h, \mathbf{a}_k) = \sum_{i=1}^{N_h} \tilde{A}(h,k) D^k_\varepsilon(T(\mathbf{a}_h, \mathbf{a}_k, \mathbf{d}^h_i)), \qquad (4)$$

where D^k_ε is the distance transform of V^k, $\mathbf{d}^h_i$ is a point of view V^h, and $T(\mathbf{a}_h, \mathbf{a}_k, \mathbf{d}^h_i)$ is the transform function which maps $\mathbf{d}^h_i$ to D^k_ε by the absolute poses $\mathbf{a}_h$ and $\mathbf{a}_k$. Indeed, the total error from all the overlapping views is defined as:

$$E(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \sum_{(h,k)} \sum_{i=1}^{N_h} \tilde{A}(h,k) D^k_\varepsilon(T(\mathbf{a}_h, \mathbf{a}_k, \mathbf{d}^h_i)), \qquad (5)$$

The error is minimized with Levemberg-Marquardt, with analytic Jacobian. Rotations are represented as quaternions whose norm is set to unity at each iteration.

Note that the distance transform requires the use of a large amount of memory, thereby imposing some limitations on the number of views. On the other hand, distance transform can be avoided by computing finite differences at the cost of a reduced speed of the minimization procedure.

5. Results

In this section we report results for both synthetic and real views acquired by a 3D scanner. The `Bunny` 3D model – available from the Stanford 3D scanning repository – have been used to create 24 synthetic partial views. `Gargoyle` and `Madonna` are sets range images (27 and 196 views respectively) courtesy of CNR-ISTI. All the parameters have been kept fixed in all the experiments.

Table 2 reports rotation and translation errors for `Bunny` after graph-based alignment and our multiview LM-ICP. The table shows that the latter clearly improves on graph-based alignment. Furthermore, in order to appreciate the advantage of multiview LM-ICP a detail of `Bunny` is shown in Figure 1. A clear misalignment can be observed for the graph-based alignment method, whereas the same detail is correctly aligned by multiview LM-ICP.

Table 1: *Average closest points distance (in mm) before and after LM-ICP.*

Dataset	Before LM-ICP	After LM-ICP
Madonna	3.734	3.709
Gargoyle	0.460	0.459
Bunny	0.095	0.093

Table 2: *Rotation and translation errors for both Graph-Based alignment and Multiview LM-ICP.*

Method	Rot. err. (deg)	Tran. err.
Graph-Based alignment	0.3899	0.0727
Multiview LM-ICP	0.0336	0.0054

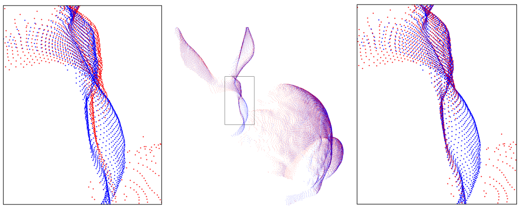

Figure 1: *Visual accuracy evaluation. Two views of* `Bunny` *onto the global coordinates system. Graph-Based alignment approach (top) and Multiview LM-ICP (bottom). We zoomed to a detail of bunny face for which the advantage of Multiview LM-ICP is clearly shown.*

The results on real data are shown in Fig.2. In this case the evaluation is only visual, as the ground truth is not available. The *binarized* adjacency matrices are shown in Fig. 3.

6. Conclusions

In this paper we propose a fully automatic method for 3D registration of multiple views. We have shown the effectiveness of *feature*-based approach to improve the estimation of views overlap, combined with a *global* view matching strategy. In order to improve the final accuracy, a multiple view registration is eventually carried out. Future work will address the application of the proposed framework on large scale scenarios.

Acknowledgments

This paper was supported by the “LIRE” Joint-Project.

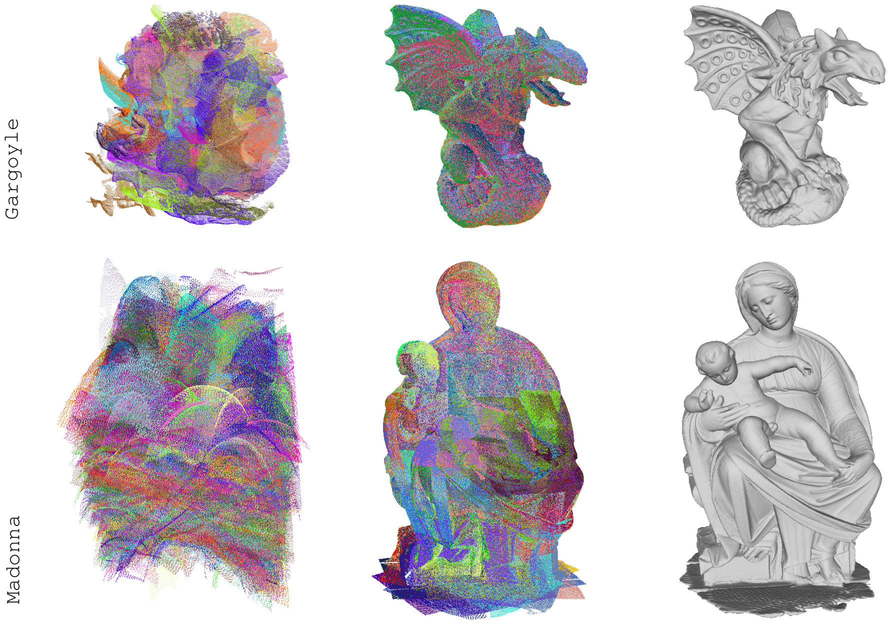

Figure 2: *Reconstructed models. Starting pose (left), aligned views (center), and reconstructed models with Poisson (right).*

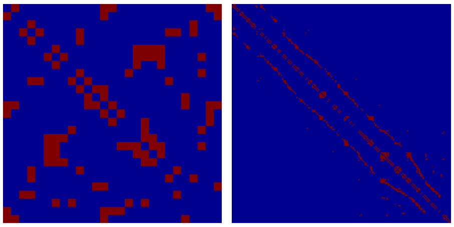

Figure 3: *The binarized adjacency matrices $\tilde{A}$. Gargoyle (left) and Madonna (right).*

References

[AMCO08] AIGER D., MITRA N. J., COHEN-OR D.: 4-Points Congruent Sets for Robust Pairwise Surface Registration. In *SIGGRAPH* (2008).

[BL03] BROWN M., LOWE D.: Recognising panoramas. In *International Conference on Computer Vision* (2003).

[BM92] BESL P., MCKAY H.: A method for registration of 3-D shapes. *IEEE Transactions on pattern analysis and machine intelligence 14*, 2 (1992), 239–256.

[BSL11] BONARRIGO F., SIGNORONI A., LEONARDI R.: A robust pipeline for rapid feature-based pre-alignment of dense range scans. In *International Conference on Computer Vision* (2011).

[CCFM08] CASTELLANI U., CRISTANI M., FANTONI S., MURINO V.: Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Computer Graphics Forum 27* (2008), 643–652.

[CFM02] CASTELLANI U., FUSIELLO A., MURINO V.: Registration of multiple acoustic range views for underwater scene reconstruction. *Computer Vision and Image Understanding 87*, 3 (2002), 78–89.

[Fit03] FITZGIBBON A.: Robust registration of 2D and 3D point sets. *Image and Vision Computing 21*, 13-14 (2003), 1145 – 1153.

[HH03] HUBER D., HEBERT M.: Fully automatic registration of multiple 3D data sets. *Image and Vision Computing 21*, 7 (2003), 637–650.

[JH99] JOHNSON A., HEBERT M.: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 5 (1999), 433–449.

[KCB09] KHOUALED S., CASTELLANI U., BARTOLI A.: Semantic shape context for the registration of multiple partial 3--D views. In *British Machine Vision Conference* (2009).

[KMSRJ05] KING B. J., MALIZIEWICZ T., STEWARD C., R. J R.: Registration of multiple range scans as a location recognition problem: Hypotesis generation, refinement and verification. In *3-D Digital Imaging and Modeling (3DIM)* (2005).

[MFM04] MARZOTTO R., FUSIELLO A., MURINO V.: High resolution video mosaicing with global alignment. In *International Conference on Computer Vision* (2004).

[Pul99] PULLI K.: Multiview registration for large data sets. In *3DIM '99: Proceedings of the Fifth International Conference on 3-D Digital Imaging and Modeling* (1999), pp. 160–168.

[TBC10] TOLDO R., BEINAT A., CROSILLA F.: Global registration of multiple point clouds embedding the generalized procrustes analysis into an icp framework. In *Symposium on 3D Data Processing, Visualization, and Transmission* (2010).

[TZ00] TORR P. H. S., ZISSERMAN A.: MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding 78*, 1 (2000), 138–156.