14-26 MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

A Study on the Robustness of Shape Descriptors to common scanning artifacts

Ferran Roure[1], Yago Díez[1], Xavier Lladó[1], Josep Forest[1], Tomislav Pribanic[2] and Joaquim Salvi[1]

[1] ViCOROB Research Group - University of Girona {froure}{yago}{llado}{forest}{qsalvi}@eia.udg.edu
[2] Faculty of Electrical Engineering and Computing, University of Zagreb tomislav.pribanic@fer.hr

Abstract

Registration is a fundamental problem in a myriad of applications ranging from heritage reconstruction to industrial applications. Descriptors are an important part of the registration pipeline as well as a very active research field. However, the sets used to illustrate descriptor performance have often undergone several preprocessing steps such as noise filtering, hole filling or outlier removal. These steps simplify the problem but are not readily available in many applications. In this paper we compare the performances of 4 state of the art shape descriptors: SHOT [1], Spin Image [2], FPFH [3] and 3DSC [4]. Experiments were carried out with real as well as synthetic data paying special attention to issues commonly present in real data (noise, outliers and low overlap). The method obtaining a best result overall is SHOT, based mostly on the results with synthetic data. Experiments with real data showed how state of the art descriptors are not yet able to produce optimal results in the most challenging scenarios.

1 Introduction

3D registration is a very important problem in a variety of fields ranging from medical imaging to industrial applications. Aligning two or more views of the same object is a complicated task which often results in huge computational cost. This results from the size of the search space related to the problem: in order to register two point clouds, a minimum of three point correspondences are needed to determine a 3D rigid motion. Thus, the number of possible correspondences is in $O(n^6)$. Although current algorithms do manage to avoid exhaustive searches, the complexity of the problem as well as the sizes of the datasets (from tens of thousands of points to millions of points) demand algorithms that are as efficient as possible.

The current state of the art on registration can be organized in terms of the matching pipeline [5]. We divide the registration process in two main parts: Coarse and Fine Matching. The first provides an initial pose that is refined in the second. The Iterative Closest Point (ICP) algorithm [6] is a de-facto standard for many fine matching problems. Concerning coarse matching, although many strategies aiming at reducing computational costs [7] exist, the most active area has recently been Shape Descriptor functions. Although very different from each other, these functions share the common goal of providing a numerical representation of the shape of objects around points.

A key factor for shape descriptors is their resiliency to some of the most commonly appearing scanning artifacts. Although extensive works [8] studying the performance of shape descriptors exist, we feel that an in-depth study on the effects of these artifacts for state of the art descriptors for rigid registration has not yet been conducted. Consequently, in this paper we present a comparison of 4 state of the art shape descriptors. We use real and synthetic data in order to assess the effect of different levels of noise and different ratios of overlap in descriptor performance. We tested four different descriptors which are extensively used in the literature: SHOT[1], Spin Image[2], Fast Point Feature Histogram (FPFH) [3] and 3D Shape Context (3DSC) [4]. We use these approaches over five different models, both synthetic and real data. The synthetic data models are the Bunny, Buddha and Dragon models from the Stanford Repository [9], which allows us to gain further insight in the problem by controlling variables such as the amount of noise included, and the real data are Bust and Joints models, from our own scanning system [10], which makes it possible to show the behavior of algorithms in a realistic setting.

2 Methodology

The state of the art of 3D registration is mostly focused on detection/description. Local shape descriptors provide a numerical representations of the shape of the object around surface points. This already challenging situation is made even more difficult by the objects being represented as a discrete set of points. These points might contain noise and, thus, fail to precisely represent the object they belong to. Consequently, descriptors that are robust to small quantities of noise in points are sought. Additionally, any matching algorithm typically has to deal with the two objects being matched not being exactly identical but instead presenting a certain area of overlap. This is the case, for example, of the problem of reconstructing an object from several views. Descriptors that are able to focus on significant common areas are desirable. Finally, scanners sometimes produce spurious "outlier" points that do not really represent any part of the objects but nevertheless complicate registration. This is the case for example for metallic objects causing laser scanners to output incongruous points due to reflections.

2.1 Descriptors used

We selected four state of the art descriptor methods. For their implementation, we used the code provided by the Point Cloud Library (PCL) [11].

Spin Image [2] is a widely used descriptor. The method considers a point a_i and its associated normal vector $\vec{n_i}$. The local shape around a_i is codified using two variables: distance α between each point and support line of the normal vector $\vec{n_i}$, and the distance β between each point and the tangent plane P.

Signature of Histograms of OrienTations (SHOT) [1] stresses the importance of choosing a good local reference frame (RF). For each point, its RF is chosen via eigendecomposition of the covariance matrix of neighboring points. The eigenvectors with higher eigenvalues are used as the axes of the RF. Then, the method makes a histogram of the angles between the normal vectors of the query point and the neighbors inside a supporting sphere.

Fast Point Feature Histogram (FPFH) [3] Given a query point p, all points inside a ball of radius r centered at this point are connected with the others via a fully interconnected (complete) mesh. For each query point with normal vector $\vec{n}$, the algorithm selects another neighboring point q holding that the angle between $\vec{n}$ and $\vec{pq}$ is minimum. Then angular information is calculated and stored (see the paper for details). A histogram divided into bins is computed according to the value of the angles between normal vectors of each pair of neighbors.

Finally, **3D Shape Context (3DSC)** [4] describes points in relation to all the other points in the object (not only in a neighborhood). Due to the size of the data sets, this algorithm only uses randomly sampled points instead of the full-sized data. Specifically, a coarse histogram is computed using a binned sphere-space division centered at p.

To sum up, we have considered a widely used descriptor as is Spin Image along with three more recent descriptors. SHOT, is an example of a descriptor favoring local reference frames that shows a remarkable resiliency to noise [1]. FPFH is representative of histogram-based descriptors that obtains good results with sets of overlap as low as 45% [3]. 3DSC is a descriptor with a totally different approach as not only points in local neighborhoods are used. We believe that this selection covers the different approaches that have proven successful in one way or another in the recent development of shape descriptors.

2.2 Data used

We use two types of data, *processed* and *real data*. Processed models have either been created synthetically or have had some kind of post-processing such as de-noising or smoothing. We consider real models to be those created with raw data acquired directly from scanners. The registration of these objects is challenging due to noise, occlusions and possible outlier points.

Processed Data

Three of the most well-known objects in the literature are the Bunny, Buddha and Dragon models the from Stanford Repository [9] (Figure 1). These datasets consist of several views that seem to have undergone noise and outlier filtering after acquisition with a a range scanning system. The Bunny model is the simplest model with ≈37000 points per view. All features are clearly defined, without noise or outliers. The degree of overlap between consecutive views is generally high (≈90%). The Buddha model is a little more challenging because of its larger size (≈75000 points per view). Additionally, it presents finer detail and some symmetries. Specifically, the base of the figure is a cylindrical (and thus symmetrical) pedestal which is a large source of local symmetries and hampers normal space analysis. Finally, the Dragon model presents ≈42000 points per view, with a considerably number of symmetries along its transversal axis.

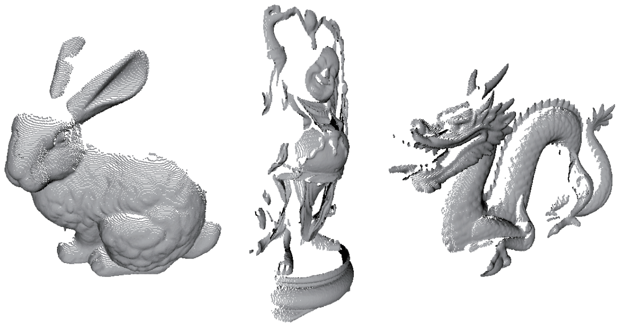

Figure 1. Views of Bunny, Buddha and Dragon models from the Stanford Repository.

Real Data

In these models no post-processing steps were used (Figure 2). The challenges here arise from the presence of noise, low overlap ratios and outliers. The Bust model was acquired using a structured light system [12]. The views of this model contain ≈450000 points. The overlap ratio is ≈50%, depending on the view. We also focused on the head section of the model in order to obtain a model with less points (≈100000 points), without losing point density. The most complex registration problem studied in this work is the Joints model (≈520000 points). This scenario is an instance of the "bin picking" problem, where a robot arm is expected to pick an industrial part from an unstructured heap of similar parts. This data was obtained using a range scan composed by a laser and a single camera and presents abundant noise and outliers. The model consists of a big heap of unsorted parts as well as a single joint model to be found within the heap.

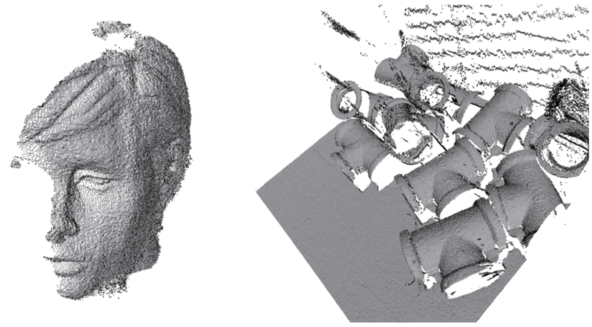

Figure 2. Real data models. Left: Head of the Bust model. Right: Joints model.

3 Experimental Evaluation

The descriptor-based coarse matching algorithm used in this evaluation is the following. Given two input point clouds $\mathcal{A}$ and $\mathcal{B}$, we select a *wide base* [13] $B_{\mathcal{A}}$ made up of 3 points from set $\mathcal{A}$. For each point in $B_{\mathcal{A}}$, we search for point correspondences in $\mathcal{B}$. Possible correspondences are sorted in terms of descriptor

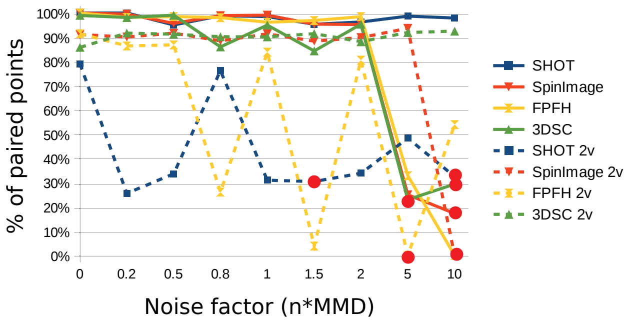

Figure 3. Results with Bunny model.

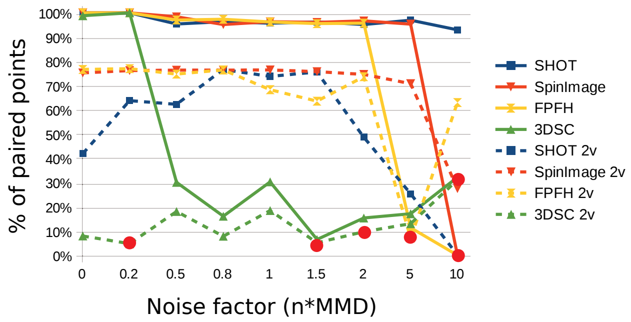

Figure 5. Results with Dragon model.

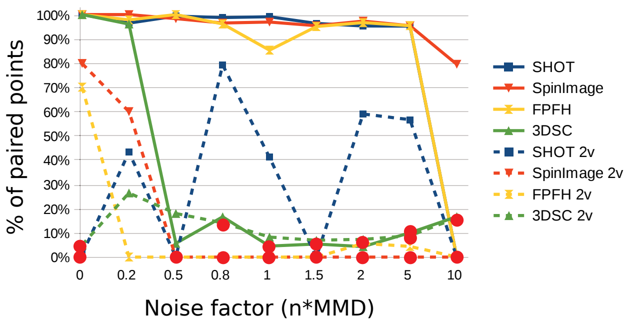

Figure 4. Results with Buddha model.

values. In order to discard obviously wrong bases from set $\mathcal{B}$, the distances between base points are checked. The bases that pass this test, are used to compute an alignment movement. Then, the percentage of paired points and the residue are computed to evaluate the quality of the coarse alignment. Finally, ICP is applied in order to refine the movement and complete the registration process.

Some details on the design decisions for this algorithm follow. First of all, using wide bases makes the search more robust than using randomized selection [13]. In order to increase the quality of the bases $B_{\mathcal{A}}$, we select their points from a reduced group of keypoints obtained using a Intrinsic Shape Signature (ISS) detector [14]. The number of descriptor-sorted correspondences for each point is set to be, $k = 50$ for processed data, and $k = 100$ for real models, after preliminary testing. Finally, and in order to circumvent having chosen bases falling in non-overlap areas, the whole search is repeated for $r = 10$ different bases and only the best result is outputted. The computational cost of the algorithm is thus, $O(r\,k^3\,m\,log\,n)$. With processed models the algorithm is able to find solutions after exploring few candidates (small k), therefore the computation time is rather small ($\bar{t} = 3min$). However, for more complicated objects, the algorithm tends to explore all possible candidates and, thus, the running time increases greatly ($timeout = 20hours$). For each model, we compute the Mean Minimum Distance (MMD) between its points in order to fairly apply random noise for the tests.

Processed Data

For this datasets we performed extensive tests using different levels of noise and overlap. For each model we considered two views. For each of them we produced 8 noisy copies. We added Gaussian noise with varying intensity. For each point, we added random noise up to $n * MMD$ with n taking eight values ($n \in \{0.2, 0.5, 0.9, 1, 1.5, 2, 5, 10\}$). We registered each model with noisy copies of itself and with noisy consecutive view. In the first case, only noise affects registration while in the second lower overlap is an added factor. Figures 3, 4 and 5 show the results separately for every model. Continuous lines correspond to registrations of the same view while dashed lines stand for registrations of consecutive views. X-axes contain the level of noise in each set while Y-axes contain the percentage of matched points of the first view over the second view, after coarse matching (and before ICP). Finally, red dots represent the cases where the approximation is not good enough for ICP to succeed. As the main aim of coarse matching algorithms is for this not to happen, we consider a lower "number of red dots" to be a measure of algorithm success.

We observed that the overlap between models plays a more crucial role than noise. All descriptors performed well even under high amounts of noise when objects were registered to (noisy) copies of themselves (see full lines in Figures 3, 4 and 5). Regarding whether or not ICP was able to converge to the optimal solution after the descriptor based algorithms, SHOT succeeded in all 27 registrations, FPFH failed in 1 of them, Spin Image in 2 and 3DSC in 5. However, when different views of the same object were considered the performance decreased. Lower overlap decreases the areas in the two objects that are actually useful for matching. Consequently, descriptors must do a better job at singling out areas that do actually correspond to matching parts of the two objects. Regardless of this, the random choice of the bases $B_{\mathcal{A}}$ plays a significant role. Whenever $B_{\mathcal{A}}$ is chosen to contain a point outside the overlap region the matching is doomed to fail. Consequently, the figures should be interpreted in terms of general tendencies and not particular performances. These factors produce the unstable results depicted in figures 3, 4 and 5 on the dashed lines. Specifically, SHOT failed to produce usable approximations for ICP in 7 of the 27 cases, 3DSC in 8 and FPFH and Spin Image in 9. The higher number of problems were detected in the Buddha model were 22 of the 36 registration run by all descriptors failed. To sum up, SHOT obtained the best results in both tests. Even with the problems detected for noise and low overlap most approximations (160/202) were enough for ICP refinement. Regarding data, complex objects with symmetries such as the Buddha proved to be the most challenging.

Real Data

In order to test the descriptor methods in realistic scenarios we performed a set of experiments where pairs of views of the Head and Bust models were registered. The Joints model case is substantially different because, instead of align two different views, we try to identify a certain shape in a unsorted heap. In these tests, noise is inherent to the scanning process. Table 1 shows shows the overlap between the views, the overlap after coarse registration and whether or not ICP succeeded. For the Joints model we provide both overlapping ratios to notice the size difference of this particular case.

The performance of descriptor methods with real data was much worse than with synthetic data. Positive results were not achieved for all models. Although we set timeout to be 20 hours and increased the number of candidate to $k = 100$, in some cases, descriptors are not able to find a good approximation or just they do not find any result. In any case, these times are too large to be considered usable in real applications. Concerning descriptor comparison, best results were obtained by 3DSC and SHOT. FPFH obtained lower pairing ratio after coarse matching but succeeded in their main goal of producing an alignment that ICP could use to produce the correct fine matching, but only in Head model, which is smaller than Bust.

Table 1. Real data results. Original overlap is given by two values: $\mathcal{A}$ respect to $\mathcal{B}$ and vice-versa. The paired points are only $\mathcal{A}$ over $\mathcal{B}$.

	Method	**Orig. Ovlp**	**Pair. Pnts**	**ICP**
Head	SHOT	84% / 54%	33.18%	✓
	SP		2.13%	✗
	FPFH		5.75%	✓
	3DSC		75.15%	✓
Bust	SHOT	86% / 53%	23.12%	✓
	SP		3.29%	✗
	FPFH		-	✗
	3DSC		50.47%	✓
Joints	SHOT	4% / 73%	31.24%	✓
	SP		-	✗
	FPFH		-	✗
	3DSC		-	✗

4 Conclusions

In this paper we analyzed the performance of 4 state of the art descriptors when dealing with common scanning artifacts. In general terms, SHOT proved to be the most stable for synthetic data (failing only in 7 out of 54 registrations). In these cases, overlapping between views becomes more important than noise, because with single-view tests, almost all methods succeed. Additionally, objects presenting symmetries such as the Buddha proved to be more challenging to register as seen by the failure rate of 22 of its 36 2-view synthetic registrations. Concerning real data, 3DSC and SHOT obtained the best initial approximation. Here, the intrinsic noise from the scanning system becomes more problematic. Descriptors need bigger searching radii, and more candidates are needed to achieve a good alignment results. The need of good methods able to deal with raw data in a reasonable amount of time has been shown to be an open problem for the community as only two descriptors were able to output results usable for ICP within the 20 hours of time limit set.

References

[1] F. Tombari, S. Salti, and L. Di Stefano, “Unique signatures of histograms for local surface description,” *European Conf. on Computer Vision*, pp. 356–369, 2010.

[2] A. E. Johnson, *Spin-images: A representation for 3-D surface matching*. PhD thesis, Citeseer, 1997.

[3] R. B. Rusu, N. Blodow, and M. Beetz, “Fast point feature histograms (fpfh) for 3d registration,” in *Robotics and Automation, 2009. ICRA’09. IEEE International Conference on*, pp. 3212–3217, 2009.

[4] M. Körtgen, G.-J. Park, M. Novotni, and R. Klein, “3d shape matching with 3d shape contexts,” in *Central European Seminar on Comp. Graph.*, vol. 3, p. 5, 2003.

[5] Y. Díez, F. Roure, X. Lladó, and J. Salvi, “A qualitative review on 3d coarse registration methods,” *ACM Computing Surveys (CSUR)*, vol. 47, no. 3, p. 45, 2015.

[6] P. J. Besl and N. D. McKay, “A method for registration of 3-d shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[7] Y. Díez, J. Martí, and J. Salvi, “Hierarchical normal space sampling to speed up point cloud coarse matching,” *Pattern Recognition Letters*, vol. 33, pp. 2127 – 2133, 2012.

[8] A. e. Bronstein, “Shrec 2010: robust feature detection and description benchmark,” *Eurographics Workshop on 3D Object Retrieval*, vol. 2, no. 5, p. 6, 2010.

[9] “http://graphics.stanford.edu/data/3dscanrep.”

[10] F. Roure, Y. Díez, X. Lladó, J. Forest, T. Pribanic, and J. Salvi, “An experimental benchmark for point set coarse registration,” in *Int. Conf. on Computer Vision Theory and Applications*, 2015.

[11] “http://www.pointclouds.org.”

[12] T. Pribanić, S. Mrvoš, and J. Salvi, “Efficient multiple phase shift patterns for dense 3d acquisition in structured light scanning,” *Image and Vision Computing*, vol. 28, no. 8, pp. 1255–1266, 2010.

[13] D. Aiger, N. J. Mitra, and D. Cohen-Or, “4-points congruent sets for robust pairwise surface registration,” in *ACM Transactions on Graphics*, vol. 27, p. 85, 2008.

[14] Y. Zhong, “Intrinsic shape signatures: A shape descriptor for 3d object recognition,” in *IEEE Int. Conf. on Computer Vision Workshops*, pp. 689–696, 2009.

ACKNOWLEDGEMENTS

This work has been supported by the FP7-ICT-2011-7 project PANDORA-Persistent Autonomy through Learning, Adaptation, Observation and Replanning (Ref 288273) funded by the European Commission and the project RAIMON-Autonomous Underwater Robot for Marine Fish Farms Inspection and Monitoring (Ref CTM2011-29691-C02-02) funded by the Ministry of Economy and Competitiveness of the Spanish Government. Ferran Roure is supported by a FPI scholarship associated to the RAIMON project.