Semantic Shape Context for the Registration of Multiple Partial 3D Views

Samir Khoualed[1]
samir.khoualed@lasmea.univ-bpclermont.fr
Umberto Castellani[2]
umberto.castellani@univr.it
Adrien Bartoli[1]
adrien.bartoli@gmail.com

[1] LASMEA
CNRS/Université Blaise Pascal
Clermont-Ferrand, France
[2] VIPS
University of Verona
Verona, Italy

Abstract

Point-to-point matching is a crucial stage of 3D shape analysis. It is usually solved by using descriptors that summarize the most characteristic and discriminative properties of each point. Combining local and global context information in the point descriptor is a promising approach.

We propose a new approach based on what we call *semantic shape context* to combine effectively *local* descriptors and *global* context information by exploiting the Bag of Words (BoW) paradigm for the representation of a single 3D point. Several local point descriptors are collected and quantized from the training set, by defining the visual vocabulary composed by a fixed number of visual words. Each point is then represented by a set of BoWs which encode the inter-relationship with all the other points of the object (i.e., the *context*).

Experiments were carried out on several 3D models. The proposed approach makes fully automatic 3D registration of partial views possible, and generally outperforms state-of-the-art methods in terms of robustness and accuracy.

1 Introduction

Estimating rigid transformations that align corresponding points of partial 3D views is a critical issue for various tasks in computer vision (e.g., object recognition, object tracking and 3D model reconstruction). The ICP algorithm [2] is the gold standard for pairwise view alignment, but it requires a sufficient overlap among the views and a coarse pre-registration to avoid getting stuck into a local minimum. In particular, according to the taxonomy proposed in [8], when an initial estimate is unknown and more than two views are involved, the problem is called *multiview surface matching*. Three main sub-problems need to be solved [8]: (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 address these issues. Focusing on (i) and (ii), two main rough categories of methods have been introduced: *local* methods [9, 13, 14], which are based on point-to-point correspondences, estimated based on a point signature describing local surface properties, and *global* methods [10, 16] which estimate directly the matching of the whole views by comparing

BMVC 2009 doi:10.5244/C.23.102

global surface characteristics. Combining local and global information is a promising approach, but only few methods use this strategy, to the best of our knowledge (see [12], for an extensive overview of 3D shape matching methods). An interesting and effective approach has been proposed in [7], which is an extension of the so called *shape context* [1] to the 3D domain. Shape context encodes the distribution over relative positions of a fixed point with all the other points of the shape. In this fashion, it summarizes the global shape in a rich local descriptor [1]. Besides, other approaches have been proposed [7, 9, 13, 14]. In the seminal work [9], the *Spin Images* were introduced as the 3D histogram of distances of neighbouring points from the normal vector and the tangent plane. In [13] an automatic method for correspondence establishment is proposed, by converting the views into multidimensional table representations, namely *tensors*. In [14] a geometric scale-space analysis of 3D models has been proposed from which a scale-dependent local shape descriptor is derived.

In this paper, we improve the basic idea of the shape context by combining local descriptors with the Bag-of Words (BoW) paradigm. We focus on the problem of multiview surface matching by addressing the early two above-described sub-problems. The proposed point description and matching approach is based on the following steps:

1. **Local points description**. Several local point descriptors are computed in order to capture the local shape variation in the point neighborhood [9, 15].

2. **Visual vocabulary construction**. The set of point descriptors collected from all the views of the same model are properly clustered in order to obtain a fixed number of 3D visual *words* (i.e., the set of cluster centroids) [9].

3. **Context definition**. Each local descriptor is assigned to a visual word, and a BoW representation is defined by counting the number of points assigned to each word. In particular, for a fixed point its context is defined as the set of BoWs computed on several regions which are defined by concentric shells centered on the fixed point itself.

4. **Point matching**. The matching between two points is computed by comparing their respective signatures and by taking into account the different kinds of descriptors. Both the *local* and *contextual* contributions are considered.

The main idea of this approach consists in the fact that the proposed context encodes not only the spatial relationship between points, but also their 'class' w.r.t. each local descriptor (i.e., points associated to the same cluster belong to the same class). We thus call this new representation *Semantic Shape Context* (SSC) (we use the term "semantic" to emphasize the fact that we learn the local shape of the point), where here the semantic is inferred by the point classification. It is worth noting that the choice of local point descriptors is not the focus of this work, since in principle any set of local descriptors can be used and cast in the proposed context. The effectiveness of the SSC is shown by proposing a multiview surface matching framework which implements a fully automatic model registration pipeline.

Roadmap. This paper is organized as follows. Section 2 introduces the idea of SSC by describing the proposed approach to organize several descriptors into a compact point context. Section 3 gives details on the implemented multiview surface matching strategy. Section 4 shows exhaustive results and finally, conclusions are drawn in Section 5.

2 Point description and matching

The proposed approach is based on the four main components described in the introduction. Note that few and significant points are selected from each partial 3D view using the method proposed in [3], which extends the approach for salient point detection to the 3D domain. We thus process only a reduced number of *salient* points. Note that few and significant points are selected from each partial 3D view using the method proposed in [3], which exploits saliency measurements on meshes according to perceptual principles based on the extension of scale theory onto the 3D domain. In practice the salience measures how much a surface point is displaced after Gaussian filtering at different scales, and only points with local maximal saliency at different scales are detected. We thus process only a reduced number of salient points.

2.1 Local point description

Local descriptors aim at capturing local geometric properties in the neighborhood of a given point. In general, robustness against noise and invariance against rigid transformations are sought characteristics to robustify the matching. In this paper, starting from a set of oriented points (i.e., point with normal [15]), we focus on the following geometric measures to compute the descriptors for each feature point:

- **Shape Index** *si* [15]. The Shape Index is defined as:

$$si = -\frac{2}{\pi}\arctan\left(\frac{k_1+k_2}{k_1-k_2}\right) \quad k_1 > k_2$$

  where k_1, k_2 are the principal curvatures of a generic vertex. The Shape Index varies in $[-1,1]$ and provides a local categorization of the shape into primitive forms such as spherical cap and cup, rut, ridge, trough, or saddle [15].

- **Beta Value** *bv*. The Beta Value of vertex p is represented by the projection of the nearby vertex v to the normal $\vec{n}_p$ at p. In practice it is the distance between the surface point v and the plane identified by $\vec{n}_p$.

Finally, those two measurements are collected and accumulated separately onto concentric shells, centered on the feature point. This gives three-dimensional histograms $L_{si}(i,j)$ and $L_{bv}(i,j)$[1], where $i \in [1,..,N]$, $j \in [1,..,M]$ are the indices which identify the quantized geometric measures and the distance from the salient point, respectively.

2.2 Visual vocabulary construction

The proposed approach for learning point context is inspired from the BoW framework for textual document classification and retrieval. To this aim, a text is represented as an unordered collection of words, disregarding grammar and even word order. The extension of BoW to visual data requires one to build a *visual vocabulary*, i.e., a set of the visual analog of words. As for the 2D domain [4], the visual words are obtained by clustering local point descriptors (i.e., the visual words are the cluster centroids). In practice, the clustering defines a vector quantization of the whole point descriptor space, composed of all the feature

[1] In practice $L_{bv}(i,j)$ is the Spin Image [9].

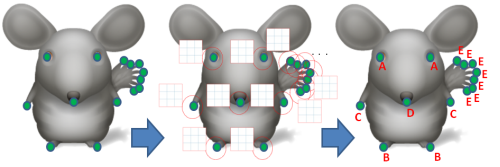


Figure 1: Selected points (left), the neighborhood of each point (center), and the point classification (right).

points extracted from all the views (i.e., the training set). In order to obtain the clustering, the K-means algorithm is employed [5]. The number of visual words is defined by fixing the parameter K. In this fashion, each point can be easily classified by assigning to it the visual word associated to the closest cluster centroid. Note that in our case, as in [4], the point classification is carried out by an unsupervised learning approach [5], but that more sophisticated classification techniques could be used. Figure 1 shows the selected points, the neighborhood of each point, and the point classification on a toy example. Similar points are assigned to the same visual word.

2.3 Context definition and global point description

Given one salient point, several sub-regions are defined by concentric shells, as for the Shape Context[2] [1]. Therefore, a BoW representation is defined for each sub-region by counting the number of points assigned to each word. Finally, the set of BoWs composes the SSC. Figure 2 shows an example of this phase.

The SSC is computed for each kind of descriptor by defining $SSC_{si}(u,v)$ and $SSC_{bv}(u,v)$, where $u \in [1,..,R]$, $v \in [1,..,K]$ are the indices which identify the distance from the salient point and visual word, respectively. Note that in this case the region of influence is extended to the whole shape, by summarizing the global shape from the 'point of view' of the salient point.

Finally, the global point descriptor is defined by a vector $G_{pd} = [L_{si}, L_{bv}, SSC_{si}, SSC_{bv}]$ concatening the contribution of both, local descriptors and SSCs.

2.4 Point matching

Point matching between two points P_1 and P_2 is carried out by comparing their respective global point descriptors G^1_{pd} and G^2_{pd} using standard metric for histograms (i.e., the χ^2 distance). A composite error function is defined to take into account the multiple components of G_{pd}. In particular the error of the local description part is defined as:

$$El_h(L^1_h, L^2_h) = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(L^1_h(i,j), L^2_h(i,j))^2}{(L^1_h(i,j) + L^2_h(i,j))}. \quad (1)$$

where $h \in \{si, bv\}$. Similarly, the error for the SSC is computed as:

$$Essc_h(SSC^1_h, SSC^2_h) = \sum_{u=1}^{R} \sum_{v=1}^{K} \omega(u) \cdot \frac{(SSC^1_h(u,v), SSC^2_h(u,v))^2}{(SSC^1_h(u,v) + SSC^2_h(u,v))}. \quad (2)$$

[2]Here the space is not split in sectors as in [1, 2] due to the instability of defining a full local reference system.

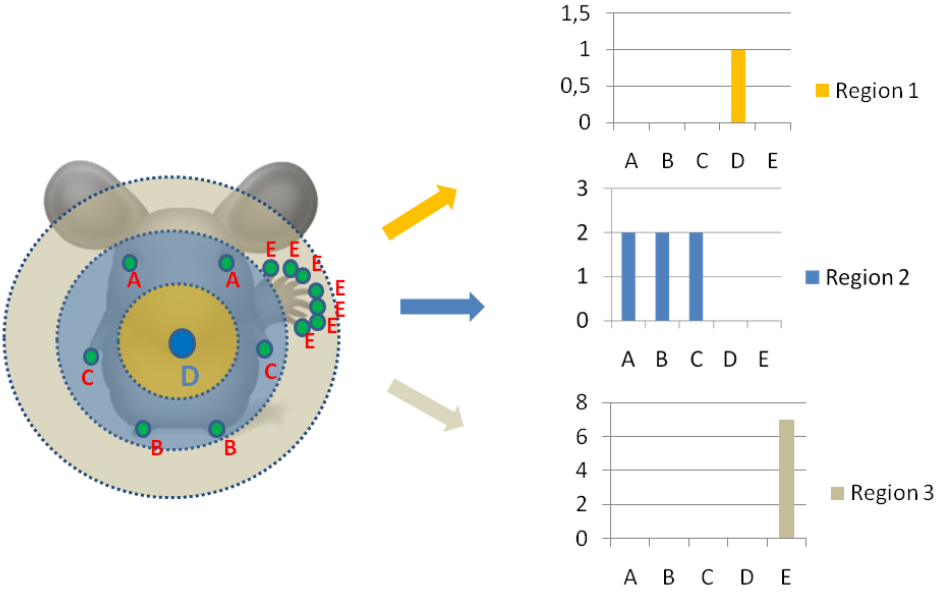


Figure 2: Context definition for a given point (colored in blue). Three sub-regions are identified by three shells. The BoW representation is computed for each shell. The set of three BoWs forms the SSC.

again with $h \in \{si, bv\}$. Note that here a new weight function $\omega(\cdot)$ has been introduced which is related to the sub-regions identified by u. The idea is to increase the influence of close regions and vice-versa. This approach is especially useful in the context of partial view matching since furthest points are likely to be occluded. More details on $\omega(\cdot)$ are given in Section 3.1.

Finally, the final error of point-to-point matching is given by

$$E = \alpha_{l_{si}} El_{si} + \alpha_{l_{bv}} El_{bv} + \alpha_{ssc_{si}} Essc_{si} + \alpha_{ssc_{bv}} Essc_{bv}. \quad (3)$$

where the α are the contribution weights. We have introduced such weights because in principle we should need to tune them in order to balance properly the contribution of each term and possibly normalize the effect of different numerical ranges of the various error components. In practice we find good results with equal weights:

$$\alpha_{l_{si}} = \alpha_{l_{bv}} = \alpha_{ssc_{si}} = \alpha_{ssc_{bv}} = 0.25 \quad (4)$$

Note that, as we already mentioned before, both the local and contextual contributions are considered.

3 Multiview surface matching

In this section we test the effectiveness of the proposed point descriptor to address the problem of multiview surface matching. As customary [8], registration is carried out pairwise first, and then a simple strategy for treating multiple views is implemented.

3.1 Pairwise registration

We propose our point matching strategy to solve robustly the issue of pairwise view pre-alignment. Given a pair of 3D views, salient-point-correspondence estimation is carried out by comparing the descriptors of each salient point of the first view, with all the salient points of the second view. In this fashion a graph of point-to-point similarities is built and the correspondences are estimated by a bipartite graph matching algorithm [5]. Note that in this case, to compute the total error between two points (*i.e.*, equation. (3)), the weighting function $\omega(\cdot)$ of equation. (2) is defined as a Gaussian function. In order to estimate the scale, we choose the scale value σ_{best} which minimizes the pairwise registration error[3]. To this aim we implement a greedy approach evaluating a range of values $\sigma_s \in I$, which are fixed in order to allow the SSC to be influenced from the whole view to a small neighborhood.

In order to remove possible wrong correspondences, the standard RANSAC algorithm is implemented. It imposes the rigid constraint among the two views [6]. The output of RANSAC provides the final salient-point-correspondences which are used to estimate the pre-alignment.

3.2 Treating multiple views

The extension of the registration to multiple views requires the estimation of the pairwise overlap to determine the pairs of views that can be registered [8]. To this aim we build the *registration matrix* $M_{x,y}$ which stores the registration error between each pair of views V_x and V_y after the pre-alignment. Moreover, for each pairwise registration we store also the estimation of best scale $\sigma^{x,y}_{best}$. In fact, the underlying idea is that the magnitude of $\sigma^{x,y}_{best}$ can be used as an estimate of the overlap between V_x and V_y (i.e., the larger $\sigma^{x,y}_{best}$ the larger the overlap between V_x and V_y). Therefore, we propose a strategy for estimating the best pairs of views to be registered. We binarize the matrix $M_{x,y}$ by using a threshold ≈ 0.05 from the mean registration error. Then we estimate the best path which connects all the views and maximize the sum of values $\sigma^{x,y}_{best}$ (*i.e.*, similar to [14] by maximum spanning tree). Finally, the registrations of the selected pairs are refined by ICP [2] and all the views are put to the global reference system by simply concatenating subsequent pairwise rigid transformations.

4 Results

We performed our experiments on range data obtained from The Stuttgart Range Image Database[4] by using range images of different types. We used several models with various features, planar surfaces, symmetries and curvatures. In our experiments, the scale of our

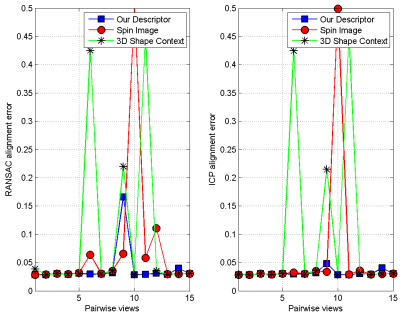
[3] The registration error is computed by the summing the residual errors of all corresponding points between the two views after the pre-alignment, where the correspondences are computed by closest point.

[4] University of Stuttgart, http://range.informatik.uni-stuttgart.de/htdocs/html/

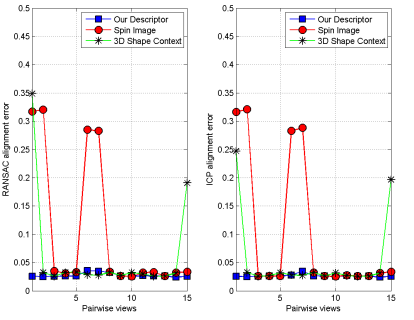
descriptor was automatically adjusted, $\sigma_s \in \{0.1,\ 0.6,\ 1.1,\ 1.6, \ldots, 4.6\}$ and the other parameters kept constant. The experiments have shown, in contrast to the *Spin Image* and 3*D Shape Context* descriptors, that the setting for all but the scale parameter does not affect the performance of our descriptor . Therefore, setting only one parameter is easily done in practice. The most important parameter, *the number of visual word* was set to $K = 50$, which is an estimate of the number of salient points per view. The proposed descriptor was tested with two geometric measures, (1) *shape index* and (2) *beta value*. The four components were weighted to contribute equally (*i.e.*, equation. (4)).

The descriptor was evaluated according to the *alignment error*. We define it as the residual error obtained after applying RANSAC-based and ICP-based alignment. Our test proceeds by computing the registration matrix $M_{x,y}$ that contains the alignment error for each pair of views among given sets of range images (in our case, we used 16 views of each model). We select the best σ_s denoted σ_{best} by inspecting the RANSAC alignment error. We record the corresponding RANSAC and ICP alignment errors as well as the associated transformation. We find the best path of a weighted graph (*i.e.*, the registration matrix $M_{x,y}$) by using the *maximum spanning tree* algorithm. Figure 3 illustrates the results of our descriptor with 16 views of 4 models (Bunny, Bull, Dino and Dragon). It is easy to notice the considerable difference in the approximate pairwise registration (*i.e.*, RANSAC-based) obtained with our descriptor compared to those obtained by *spin image* and 3*D shape context*. Moreover, Figure 3 shows that *spin image* and 3*D shape context* failed in many cases to give an accurate enough pre-alignment (*i.e.*, cases where the pre-alignment error is higher than ≈ 0.2). Our results are more accurate and they can be directly refined by any ICP-based registration algorithm. These results justify our claim that our descriptor contain rich discriminative information. That is the main reason for the lack of accuracy of the other descriptors.

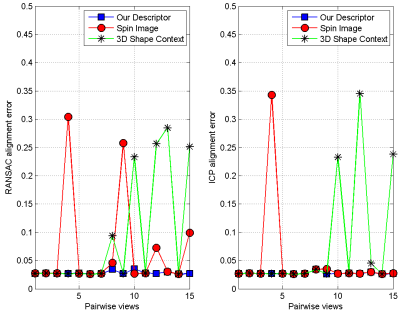
In order to demonstrate the robustness of our descriptor according to the overlap area, we computed the best scale σ_{best} for each pairwise alignment. The corresponding alignment error is reported. We use 'large','medium', or 'small' to define the amount of overlap which is proportionally estimated from σ_{best} (*i.e.*, the larger σ_{best} the larger the overlap). The computed pre-alignment error for each pairwise view is compared with the corresponding *spin image* and also with 3*D shape context* values. Table 1 summarizes the results. For each model we selected 4 different pairwise alignments according to the σ_{best} value (*i.e.*, large, medium, or small overlap). Note that the *spin image* and 3*D shape context* descriptors give more accurate alignments for large overlap. In contrast, for the small to medium overlaps, our descriptor seems to be significantly more accurate. The main reason of failure for the other approaches, is that they are based on a single source of information. Therefore, any small local surface deformation can produce a significant error on the resulting descriptor. This means that the error may come from the type of used information and may not be related to the descriptor representation (*i.e.*, histogram-like). In fact, there is no preferred information to build a descriptor that provides a high matching rate. Therefore, combining local descriptors with global context can resolve the ambiguities that may appear in each other (*e.g.*, locally where a view has similar surface elements). By using multiple sources of information (*e.g.*, here we used *shape index* and *beta value*), we can compensate for the apparent or intrinsic defect of one component (*e.g.*, *beta value*) by retrieving information from another: local descriptor or global context. Another strong point in our approach is related to the relaxed constraints on the invariance rotation for the histogram implementation. It means that we can use a simplified histogram-like implementation. Therefore, we avoid a high loss of information [□] (*i.e.*, going from a 3D to a 2D representation) and gain more in robustness. The mentioned arguments may explain the superior performance of our



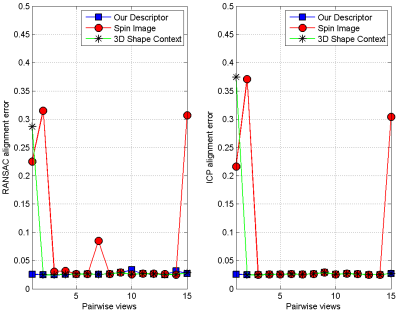
(a) Bunny



(b) Bull



(c) Dino



(d) Dragon

Figure 3: Comparison of pairwise alignment errors obtained by our descriptor with those obtained with *spin image* and 3*D shape context*. Note: the number of bins of *spin image* is 10×10 whereas the number of bins of 3*D shape context* is $10 \times 10 \times 10$. For each model: Left: RANSAC alignment error (*i.e.*, pre-alignment error), Right: ICP alignment error. The x axis shows the pairwise alignment with a reference view ordered from 1 to 15 (*i.e.*, we used 16 views of each model).

approach in the case of small overlap illustrated in Table 1.

Finally, we conducted an automatic multiple views registration. We tested a collection of 10 models, consisting of 16 views of each model. Note that those views are chosen in such a way that ICP would diverge by requiring a robust pre-aligning strategy. Figure 4 shows the results of 4 models. Note that the registered models (second and third columns of each row) were further refined with an ICP-based algorithm. The registered models in the second and third columns of each row do not have noticeable defects. This means that the approximate alignment (*i.e.*, with the RANSAC-based algorithm) obtained is very accurate and can be directly and successfully refined with an ICP-based algorithm[5].

[5] An ICP-based algorithm requires a sufficient overlap among the views and a coarse pre-registration to avoid getting stuck into a local minimum.

(a) Bunny

			RANSAC alignment error		
Pairwise alignment	σ_{best}	overlap	**Our Descriptor**	**Spin Image**	**3D Shape Context**
08 ← 07	3.1	large	0.0310	0.0283	0.0388
14 ← 12	1.6	medium	0.0292	0.0583	0.4479
03 ← 02	0.6	small	0.0301	0.0637	0.4248
09 ← 10	4.6	large	0.0284	0.0306	0.0304

(b) Bull

			RANSAC alignment error		
Pairwise alignment	σ_{best}	overlap	**Our Descriptor**	**Spin Image**	**3D Shape Context**
07 ← 08	0.1	small	0.0256	0.3168	0.3489
11 ← 12	4.1	large	0.0260	0.0329	0.0333
16 ← 02	1.6	medium	0.0267	0.0323	0.0299
09 ← 10	3.6	large	0.0256	0.0348	0.0256

(c) Dino

			RANSAC alignment error		
Pairwise alignment	σ_{best}	overlap	**Our Descriptor**	**Spin Image**	**3D Shape Context**
16 ← 02	0.1	small	0.0269	0.0727	0.2568
03 ← 04	3.6	large	0.0271	0.0260	0.0260
02 ← 03	1.6	medium	0.0275	0.0301	0.2847
07 ← 08	4.6	large	0.0261	0.0276	0.0277

(d) Dragon

			RANSAC alignment error		
Pairwise alignment	σ_{best}	overlap	**Our Descriptor**	**Spin Image**	**3D Shape Context**
04 ← 05	4.1	large	0.0263	0.0263	0.0263
13 ← 14	2.1	medium	0.0258	0.2252	0.2869
14 ← 15	1.1	small	0.0246	0.3149	0.0247
03 ← 04	2.6	medium	0.0258	0.0846	0.0258

Table 1: The approximate pairwise alignment.

5 Conclusion

In this paper, we introduced a new approach to computing 3D point descriptors. This is based on combining local information with the corresponding shape context information by exploiting the Bag of Words (BoW) paradigm. We derived a novel descriptor robust to the overlap and extremely discriminative. Our *Semantic Shape Context* descriptor is effective yet simple to implement. It resolves the problem of pre-alignemnt (*i.e.*, approximate alignment by postapplication of a RANSAC-based method) which becomes accurate enough to be directly refined by an ICP-based registration method. We also introduced an approach to automatically estimate the overlap area. We used the scale for determining which views overlap. Based on the best scale, we find the best pairs of views to be registered to each other for fully automatic registration. Our experiments demonstrated the efficiency of our descriptor for the case of a small overlap. In the future, we will investigate our approach in the 2D domain.

Acknowledgements

This paper was supported by the PRIN 2006 project 3-SHIRT and the Joint-Project INVIA.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (4):509–522, 2002.

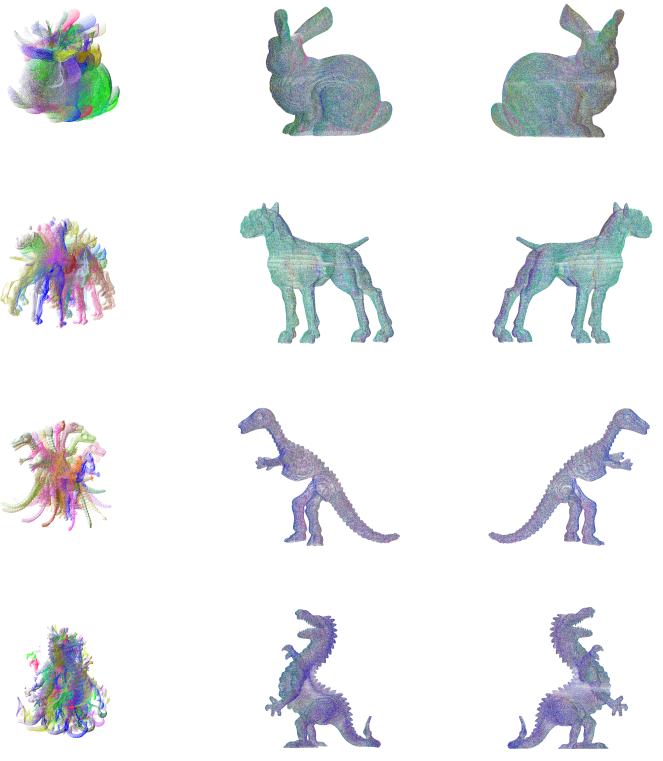


Figure 4: Automatic registration of a set of range images from the Stuttgart Range Image Database. First column shows the set of 16 (plotted in the same reference system) views used to construct each model. Each 3D model is visualized from two opposite angles (second and third columns). The registered models are very accurate.

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

[3] U. Castellani, M. Cristani, S. Fantoni, and V. Murino. Sparse points matching by combining 3D mesh saliency with statistical descriptors. In *Computer Graphics Forum*, volume 27, pages 643–652. Blackwell Publishing, 2008.

[4] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, volume 2004, 2004.

[5] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification*. Wiley New York, 2001.

[6] M. Fischler. Random Sample Consensus: A Paradigm for Model Fitting With Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[7] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik. Recognizing objects in range

data using regional point descriptors. *Lecture Notes in Computer Science*, pages 224–237, 2004.

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

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

[10] A. Makadia, A.I.V. Patterson, and K. Daniilidis. Fully Automatic Registration of 3D Point Clouds. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Volume 1*, pages 1297–1304. IEEE Computer Society Washington, DC, USA, 2006.

[11] S. Malassiotis and M.G. Strintzis. Snapshots: A Novel Local Surface Descriptor and Matching Algorithm for Robust 3D Surface Alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1285, 2007.

[12] A.S. Mian, M. Bennamoun, and R.A. Owens. Automatic correspondence for 3d modeling: An extensive review. *International Journal of Shape Modeling*, 11(2):253, 2005.

[13] A.S. Mian, M. Bennamoun, and R.A. Owens. A novel representation and feature matching algorithm for automatic pairwise registration of range images. *International Journal of Computer Vision*, 66(1):19–40, 2006.

[14] J. Novatnack and K. Nishino. Scale-Dependent/Invariant Local 3D Shape Descriptors for Fully Automatic Registration of Multiple Sets of Range Images. In *Proceedings of the 10th European Conference on Computer Vision: Part III*, pages 440–453. Springer-Verlag Berlin, Heidelberg, 2008.

[15] S. Petitjean. A survey of methods for recovering quadrics in triangle meshes. *ACM Computing Surveys*, 34(2), 2002.

[16] D.V. Vranic, D. Saupe, and J. Richter. Tools for 3D-object retrieval: Karhunen-Loeve transform andspherical harmonics. In *2001 IEEE Fourth Workshop on Multimedia Signal Processing*, pages 293–298, 2001.