Eurographics Workshop on 3D Object Retrieval (2011)
H. Laga, T. Schreck, A. Ferreira, A. Godil, I. Pratikakis, and R. Veltkamp (Editors)

Bag of Words and Local Spectral Descriptor for 3D Partial Shape Retrieval

Guillaume Lavoué

Université de Lyon, CNRS
Insa-Lyon, LIRIS UMR 5205, France

Abstract

This paper presents a 3D shape retrieval algorithm based on the Bag of Words (BoW) paradigm. For a given 3D shape, the proposed approach considers a set of feature points uniformly sampled on the surface and associated with local Fourier descriptors; this descriptor is computed in the neighborhood of each feature point by projecting the geometry onto the eigenvectors of the Laplace-Beltrami operator, it is highly discriminative, robust to connectivity and geometry changes and also fast to compute. In a preliminary step, a visual dictionary is built by clustering a large set of feature descriptors, then each 3D shape is described by an histogram of occurrences of these visual words. The performances of our approach have been compared against very recent state-of-the-art methods on several different datasets. For global shape retrieval our approach is comparable to these recent works, however it clearly outperforms them in the case of partial shape retrieval.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval —

1. Introduction

After Image and Video in the 90s, three-dimensional data (mostly represented by polygonal meshes) constitute the emerging multimedia content; large collections of 3D models are now available and thus the need for efficient tools to filter, search, and retrieve this 3D content becomes more acute. Hence in recent years, the problem of content-based shape retrieval (CBIR) has attracted the interest of scientists. The objective of such system is to retrieve, from a given 3D query, the most similar 3D models from a given database; a similar issue consists in classifying a given shape into the correct category.

This problem is not easy since, to be really efficient, such retrieval/classification system has to be robust to common 3D shape variations like connectivity change, non-rigid deformation (isometry), local deformation or cropping.

A lot of methods have been introduced, based on global descriptors; some of them are only robust to rigid deformations [CTSM03, FMK*03], while more recent ones are also invariant to non rigid deformations, like isometry or skeletal articulation; except the work of Gal et al. [GSCO07] which relies on histograms of local shape diameter values, most of these recent invariant descriptors are based on some spectral embeddings: Reuter et al. [RWP06], and Marini et al. [MPSF10] describe the shape by the eigenvalues of the Laplace-Beltrami operator while Rustamov [Rus07] considers the eigenvectors of this operator, similarly, the approach from Jain and Zhang [JZ07] relies on the eigenvectors of the affinity matrix; besides, the *conformal factor* descriptor from Ben-Chen et al. [BCG08] and the diffusion distances introduced by Bronstein et al. [BBK*09] are also based on the Laplace-Beltrami operator (eigendecomposition for [BBK*09] and integration into a sparse linear system for [BCG08]). Even if these global descriptors provide a good invariance to non-rigid, quasi-isometric transforms, most of them are not adapted to deal with partial similarity and by extension local deformation or cropping. Moreover only few of them can deal with topological changes.

Hence, to face these hard robustness issues, some researchers turned their attention to *local* descriptors associated with salient feature points (or keypoints), following successful approaches in 2D image recognition like SIFT [Low04]. In such keypoint-based 2D image recognition techniques [MS05], the object to recognize is represented by a set of salient local features (usually sparse) associated with local descriptors, then the recognition consists in

DOI: 10.2312/3DOR/3DOR11/041-048

finding a correspondence between the sets of feature points from the model and the scene objects respectively, using techniques like RANSAC (rigid matching) or some graph-matching algorithms. For 3D recognition, Funkhouser and Shilane [FS06] introduced such a local approach, their descriptor is based on Spherical Harmonics, while the matching is derived from RANSAC; to select a minimal set of distinctive features a quite complex process computes their respective predicted retrieval performances using a training set of classified 3D models. Recently, Sun et al. [SOG09] introduced a multi-scale local descriptor, the Heat Kernel Signature (HKS), computed via an eigendecomposition of the Laplace Beltrami operator. Basically the HKS is defined for each vertex x as a function of t and intuitively relates to the amount of heat that remains at point x after time t; the HKS also allows to select salient feature points since its extrema correspond to protrusions on the surface. This signature is quite related to the diffusion distance also used by Bronstein et al. [BBK*09] as a global descriptor. Very recently Sun et al. [SCF10] combined the HKS-based feature points with a matching framework based on fuzzy geodesics while Dey et al. [DLL*10] filter them according to their *persistence* to obtain a more robust set of feature points which is then integrated into a region matching algorithm. Similarly Agathos et al. [APP*09] introduce a feature point detector also related to surface protrusions to create regions and then match them using a graph matching technique based on the Earth Mover's Distance. Tabia et al. [TCDV10] also create regions related to surface protrusions (detector from Tierny et al. [TVM09]) and then describe them using a set of curves. Ruggeri et al. [RPSS09] introduced another keypoint detector also based on an eigendecomposition of the Laplace-Beltrami operator, as well as an associated local descriptor consisting of the geodesic shape distribution around the point; these feature points are then used to smartly sample the object before applying a bipartite graph matching for recognition.

There exist two main problems with these approaches based on 3D feature points and direct matching: 1) the repeatability (i.e. the invariance in location) of salient feature points, regarding connectivity or topological changes is not so obvious and 2) the graph matching is often a quite complex process (inexact graph isomorphism is NP-complete) particularly when a high level of invariance is required (i.e. isometry, local deformation, cropping). Sub-graph isomorphism (i.e. for partial shape similarity) is even more difficult.

The fact is that the intra-class variation involved in 3D model recognition is much higher than for specific 2D object recognition in images; hence existing 2D recognition techniques are difficult to directly transpose into the 3D world. However another kind of techniques also based on feature points was introduced in computer vision, specifically designed for higher intra-class variations (used in the case of object-category recognition rather than specific-object recognition): the Bag of Words (BoW) framework. In this kind of approaches [FFP05], each feature point from a given image is associated to the nearest visual word in a given visual dictionary (we assume that the dictionary has been preliminary built using clustering techniques in the descriptor space); the image is then represented as an histogram (i.e. the *bag*) of occurrences of the visual words.

Few works based on Bag of Words (BoW) have been introduced for 3D object recognition. Ohbuchi et al. [OOF*08] and Lian et al. [LGS10] present similar approaches, the 3D model is represented by a set of 2D views which are indexed using bags of 2D SIFT features. Liu et al. [LZQ06] and Li and Godil [LG09] introduce BoW algorithms based on Spin Image descriptors computed on a dense set of feature points (uniformly sampled on the surface). Bronstein et al. [BBGO11] also consider a dense set of feature points (every vertex of the mesh) and describe them using the Heat Kernel Signature from Sun et al. [SOG09]. Differently, Toldo et al. [TCF09] do not sample feature points on the 3D model but segment it into regions; then each region is associated with several descriptors and thus several visual words. These existing 3D BoW methods provide quite good results however in our opinion they still suffer from some drawbacks: first, the descriptors which are used in these works are quite poor regarding their equivalent in computer vision; this makes necessary the addition of spatial information between feature points like in [BBGO11, LG09]. A second problem comes from the sampling of the feature points, two possibilities exist like for 2D images: either you select a sparse set of points (or regions) like Toldo et al. [TCF09], or you consider a dense collection like [BBGO11]; in case of a sparse set, keypoints have to be stable regarding connectivity or topological changes and that is a very difficult problem; for instance the segmentation used in [TCF09] seems quite dependent of the topology. In case of a dense set of keypoints, you have to insure that they are evenly distributed over the whole surface even in case of very irregular connectivity and that is not the case if you consider each vertex as a keypoint like in [BBGO11].

In that context we introduce a new Bag of Words approach for 3D shape recognition; our algorithm relies on a uniform sampling of the feature points based on Lloyd relaxations; each feature point is described using a rich spectral descriptor. Our approach is highly robust to connectivity change, non-rigid or local deformations and cropping, due to three main reasons:

- The regular sampling of the feature points is quite independent of the connectivity, geometry or topology of the model (on the contrary with a protrusion detector or a segmentation algorithm some feature points/regions may disappear after even a small topological change).
- The descriptor associated with each feature point is the local Fourier spectrum computed over a large neighborhood of the point (after projection of the geometry onto the spectral bases). This descriptor is very discriminative and moreover is quite robust to noise or connectivity change.
- Our approach completely discards the structural informa-

tion of the feature points, hence it is intrinsically invariant to isometric deformations or topological changes.

The proposed approach is particularly robust to cropping or local deformations and thus is particularly efficient for partial shape similarity which is a particularly difficult task, still tackled by few methods [TCF09, TVM09, DLL*10]. The paper is organized as follows, section 2 provides a recall on the Bag of Words principles and an overview of our approach, section 3 describes our feature point detector and descriptor while section 4 presents our indexing/retrieval methods using BoW. Finally, section 5 presents some experiments which evaluate the robustness of our algorithm and provide a comparison with state of the art methods.

2. Overview of our approach

Figure 1 illustrates the main steps of our algorithm. Basically we model a 3D object as a collection of local feature points; each point is associated with a local patch on which we compute a descriptor. According to its descriptor, each patch is then associated with the nearest visual word from a given visual dictionary (i.e. the codebook). Hence, the object is finally described by the corresponding distribution of codewords (an histogram of occurrences). The visual dictionary is preliminary built by clustering a huge set of feature point descriptors computed over a large collection of 3D models. The centroids of the clusters represent the codewords of the dictionary.

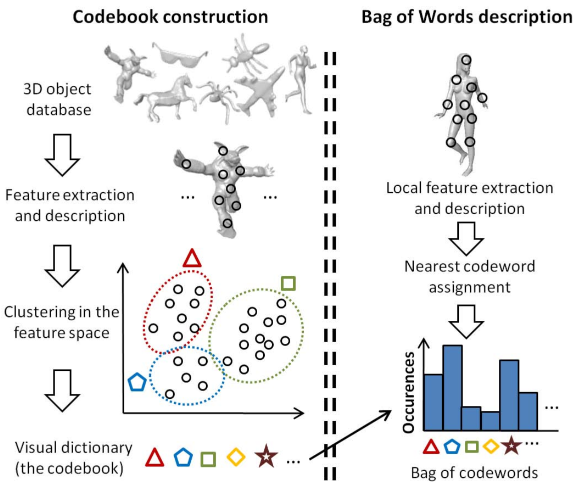

Figure 1: *Flow chart of our Bag of Words approach.*

3. Feature point detection and description

3.1. Detection

We consider a uniform sampling of the feature points on the mesh surface; the reason is that such uniform sampling gave very good results in the field of 2D image recognition (see results from [FFP05] for instance), moreover most of existing 3D salient point detectors provide collections of points which are either too sparse (i.e. protrusion detectors) or not so stable under complex geometrical or topological changes. To create the uniform sampling, we consider a random set of n_p vertices on the mesh as an initial set of *seeds* (see figure 2 on the left) and then we apply Lloyd relaxation iterations. Lloyd's algorithm [Llo82] is a fixed-point iteration that simply consists of iteratively moving the seeds to the centroids of their Voronoi cells; this algorithm was used by many authors to construct random and uniform sampling (i.e. blue noise sampling) on surfaces. Our algorithm is as follows:

1. Each vertex of the mesh is associated to the nearest seed; this creates a partitioning of the model into n_p regions, basically corresponding to the Voronoi regions associated to the n_p seeds.
2. The centroids of the n_p Voronoi regions are computed and become the new *seeds*.
3. Steps (1) and (2) are repeated until convergence.

The metric used is simply the 3D Euclidian distance. This simple algorithm converges quickly and provides a uniform sampling of the seeds (i.e. the feature points) over the surface. Figure 2 illustrates the set of feature points before and after the Lloyd's relaxation.

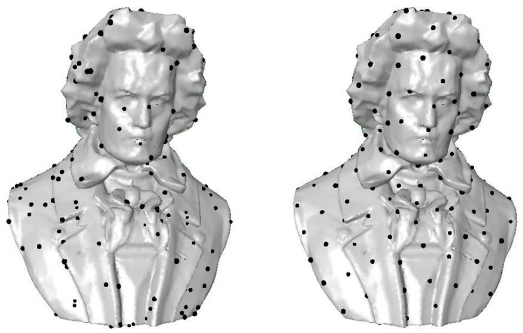

Figure 2: *Illustration of the Lloyd's relaxation algorithm. Left: 200 seeds randomly sampled. Right: result after 50 Lloyd's iterations*

This distribution has the benefit to cover uniformly the whole surface of the object, even in the case of irregular connectivity; of course this uniformity is limited by the fact that our feature points are necessarily on existing vertex positions. This constraint was not a problem in our experiment, moreover it can be easily avoided by using some very recent blue noise sampling methods like [BWWM10].
Each feature point p_i is then associated with a local patch P_i on which we will compute a descriptor; we could have taken P_i as the Voronoi region associated to each point however we prefer to extract larger overlapped regions. Hence, for each feature point we extract this local patch P_i by considering the connected set of facets belonging to a given sphere of center p_i and of a given radius r. We construct it by a region growing approach. Figure 3 illustrates a local patch for $r = 10\%$ of the bounding box length.

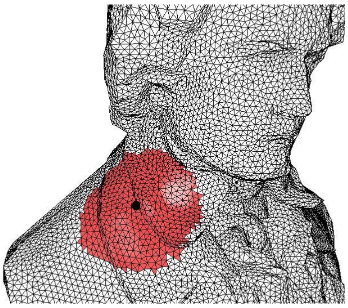

Figure 3: *A feature point (black sphere) and its associated patch (in red).*

3.2. Description

Each feature point is associated to a descriptor computed on its patch. Our objective is to propose a rich (i.e. informative) descriptor which is also fast to compute. Our idea is to use the local Fourier spectra of the patch, computed by projecting the geometry of the patch onto the eigenvectors of the Laplace-Beltrami operator (solved locally on the patch). The use of spectral tools for 3D shape retrieval has proven its efficiency (see section 1), however the proposed descriptor owns some original properties regarding the state-of-the-art:

- A lot of methods consider directly eigenvalues or eigenvectors of the Laplace operator for retrieval (e.g. [RWP06, Rus07, JZ07, MPSF10]), while we consider the spectral transform coefficients (after projection of the 3D signal onto the eigenvectors). Surprisingly this has never been done before whereas these spectral coefficients are particularly discriminative and also robust to noise and connectivity changes like pointed in [WLBD09].
- While all other methods uses spectral descriptors computed over the whole mesh, we compute our spectral transform locally, i.e. patch-by-patch. Since the eigendecomposition is a very costly process, this may save some computation time.

The Laplace-Beltrami operator Δ is the counterpart of the Laplace operator in Euclidian space. It is defined as the divergence of the gradient for functions defined over manifolds. The eigenfunction and eigenvalue pairs (H^k, λ_k) of this operator satisfy the following relationships:

$$-\Delta H^k = \lambda_k H^k \tag{1}$$

In the case of a 2-manifold triangular mesh the above eigenproblem can be discretized and simplified within the Finite Element Modeling framework (Neumann boundary conditions) [LZ10]:

$$-Q\mathbf{h}^k = \lambda_k D\mathbf{h}^k \tag{2}$$

$\mathbf{h}^k$ denotes the vector $[H_1^k, ...H_m^k]^T$ where m is the number of vertices of the patch. D is the Lumped Mass matrix, it is a $m \times m$ diagonal matrix defined by $D = diag(\sum_{t \in \aleph(v_i)} |t|)$ with $\aleph(v_i)$ the set of neighboring triangles from vertex v_i. Q is the Stiffness matrix and is defined as:

$$Q_{i,j} = (cotan(\beta_{i,j}) + cotan(\beta'_{i,j}))/2 \tag{3}$$

$$Q_{i,i} = -\sum_j Q_{i,j} \tag{4}$$

$\beta_{i,j}$ and $\beta'_{i,j}$ are the two angles opposite to the edge between vertices v_i and v_j.
To resolve this discrete eigenproblem we use the fast algorithm from Vallet and Lévy [VL08], based on a band-by-band approach and an efficient eigen-solver; hence we obtain the eigenvectors (i.e. the manifold harmonic bases) and the associated eigenvalues. The spectral coefficients $\tilde{x}_k$ (resp. $\tilde{y}_k$, $\tilde{z}_k$) are then calculated as the inner product between the patch geometry $\mathbf{x}$ (resp. $\mathbf{y}$,$\mathbf{z}$) and the sorted eigenvectors $\mathbf{h}^k$:

$$\tilde{x}_k = <\mathbf{x}, \mathbf{h}^k> = \sum_{i=1}^{m} x_i D_{i,i} H_i^k \tag{5}$$

The k^{th} $(k = 1..m)$ spectral coefficient amplitude is then defined as:

$$c_k = \sqrt{(\tilde{x}_k)^2 + (\tilde{y}_k)^2 + (\tilde{z}_k)^2} \tag{6}$$

Hence, for a given patch P_i around a feature point p_i, our descriptor is the spectral amplitude vector $\mathbf{c}^i = [c_1^i, ...c_{n_c}^i]$, with c_k^i, the k^{th} spectral coefficient amplitude of the patch P_i. We consider only the n_c first spectral coefficients to limit the descriptor to low/medium frequencies hence bringing more robustness.

4. 3D object representation and matching

4.1. Codebook construction

Given a 3D object containing a set of patches P_i associated with descriptors $\mathbf{c}^i$, the next step is to represent it as a distribution of visual words from a given dictionary. To create the visual dictionary $\Gamma = (\bar{\mathbf{c}}^1, ...\bar{\mathbf{c}}^{n_w})$, we apply a simple *k-means* clustering (n_w clusters) on a huge dataset of descriptors and keep the n_w centroids $\bar{\mathbf{c}}^k$ of the clusters as visual words. Each visual word $\bar{\mathbf{c}}^k$ is a n_c-dimensional vector.

4.2. BoW representation and matching

For a given model M, each patch P_i is associated with its closest visual word; practically we associate each patch P_i with a vector $\mathbf{b}^i$, of size n_w, such as:

$$b_j^i = 1 \quad if\ j = argmin_{k \in [1..n_w]} ||\mathbf{c}^i - \bar{\mathbf{c}}^k|| \tag{7}$$

$$b_j^i = 0 \quad else \tag{8}$$

Then the bag of words $\mathbf{b}^M$ of the whole model M is a n_w-dimensional vector containing the distribution of the visual words over all its patches:

$$\mathbf{b}^M = \sum_{i=1}^{n_p} \mathbf{b}^i \tag{9}$$

Some examples of bag of words are presented in figure 4, for $n_p = 200$ patches, $n_w = 30$ clusters and $n_c = 40$ spectral coefficients. We can observe that whereas the two armadillo models have strong differences of pose their BoWs are very similar; even a strong simplification (from 22K vertices to 6K vertices) does not significantly change the BoW. On the contrary the BoW of the cup model is significantly different.

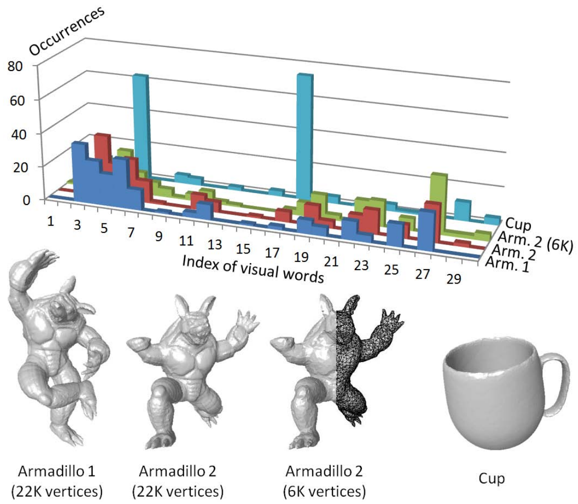

Figure 4: *Bag of Words examples.*

5. Experiments

We have conducted a set of experiments to evaluate our method. The first experiment studies the influence of the parameters; a second experiment evaluate the performance of our method in term of global shape retrieval while the last one considers a partial shape retrieval scenario. The next section describes the databases which were used in these experiments.

5.1. Databases and measures

To test our method we have considered three existing databases:

- The McGill Database †. It contains 255 objects divided into ten classes (Ant, Crabs, Hands, Humans, Octopuses, Pliers, Snakes, spectacles, Spiders and Teddy); the intra-class variations consist in non-rigid transforms applied to the models.
- The SHREC 2007 Watertight dataset ‡. It contains 20 categories each composed of 20 meshes; the intra-class variations are higher than for the McGill corpus, for instance the *FourLeg* category contains different animals (horse, dog, cow, ...).
- The SHREC 2007 Partial retrieval dataset §. It is composed of the SHREC 2007 Watertight dataset and a query set of 30 models, each one obtained by merging or removing several subparts of models belonging to the Watertight dataset.

To asset the efficiency of the methods we use the following measures, using the tools from [SMKF04]:

- Nearest Neighbor (NN): The percentage of queries for which the closest match belongs to the query's class.
- First Tier (FT): The recall for the $(C-1)$ closest matches, where C is the cardinality of the query's class.
- The Second Tier (ST): The recall for the $2(C-1)$ closest matches. It is similar to the *Bulls Eye* Score (recall for the $2C$ closest matches).
- The Discounted Cumulative Gain (DGC): This statistic gives more importance to correct detections near the front of the list; the objective is to reflect how well the overall retrieval would be viewed by a human.

5.2. Influence of the parameters

Our method is based on three parameters: the number of patches n_p, the number of coefficients n_c of the spectral descriptor and the number of codewords n_w. We have studied the influence of these parameters on the results by carrying out several retrieval tests on the McGill Database each time varying one of the parameters. The visual dictionary was built by clustering all the feature point descriptors of every 3D model of the dataset. Table 1 presents the corresponding performances in term of the First Tier measure. Several points are interesting to raise:

- When the number of spectral coefficients n_c increases from 30 to 40, the discriminative power of the Fourier descriptor increases hence the performances are better; however for $n_c = 50$ the FT measure is lower, the reason is that adding too high frequencies to the spectral descriptor removes a part of its robustness.
- Increasing the size n_w of the dictionary leads to an improvement of the performances; however a saturation effect appears, indeed the FT difference between $n_w = 200$ and $n_w = 300$ is very small.
- Increasing the number of patches n_p also leads to an improvement of the results however once again we can observe a saturation effect.

According to these observations and regarding the fact that higher are the values and higher are the indexing/retrieval times, we fix these parameters for all experiments to : $n_c = 40$, $n_w = 200$ and $n_p = 200$. It is interesting to notice that similar tests on different databases yield similar optimal values.

5.3. Global Shape Retrieval

We have compared the performance of our BoW method with two recent algorithms on the McGill Database: the graph-based approach from Agathos et al. [APP*09] and the

† http://www.cim.mcgill.ca/∼shape/benchMark/
‡ http://watertight.ge.imati.cnr.it/
§ http://partial.ge.imati.cnr.it/

Table 1: *First Tier measure for different parameter settings.*

n_c	30	40	50
$n_p = 200 \| n_w = 200$	0.621	0.629	0.624
n_w	100	200	300
$n_p = 200 \| n_c = 40$	0.619	0.629	0.630
n_p	100	200	300
$n_w = 200 \| n_c = 40$	0.611	0.629	0.634

hybrid 2D/3D approach from Papadakis et al. [PPTG08]. Table 2 presents the results; for each row, the algorithms are positioned according to their respective performances. The first remark is that the graph-based algorithm [APP*09] provides the best results, that is logical since the database considers only skeletal articulation deformations without topological changes hence it is particularly suited for graph-based representation. However we can notice that our method, whereas considering no structural information at all, provides quite good results, almost always better than [PPTG08]. In particular our method is particularly suited for nearest neighbor classification (its *NN* value is 100% for 7 categories).

Table 2: *Retrieval statistics for the McGill database.*

Class	Method	NN	FT	ST	DGC
Whole	[APP*09]	97.6	74.1	91.1	93.3
	BoW	94.5	62.9	77.6	88.1
	[PPTG08]	92.5	55.7	69.8	85.0
Ants	[PPTG08]	100	73.6	89.2	94.8
	BoW	96.7	58.3	86.7	89.1
	[APP*09]	96.7	54.9	79.7	88.4
Crabs	[APP*09]	100	98.2	99.8	99.9
	BoW	100	61.1	73.9	91.4
	[PPTG08]	100	55.2	71.8	88.7
hands	[APP*09]	95.0	83.9	88.9	95.2
	BoW	100	54.5	69.7	87.8
	[PPTG08]	90.0	43.4	57.6	77.8
humans	[APP*09]	96.6	93.5	96.4	98.1
	BoW	100	68.2	88.8	93.4
	[PPTG08]	100	47.0	63.8	83.1
Octopuses	[APP*09]	88.0	58.8	81.8	88.1
	[PPTG08]	56.0	29.5	45.0	68.9
	BoW	60.0	24.7	38.2	64.2
Pliers	[APP*09]	100	100	100	100
	BoW	100	92.1	98.4	99.2
	[PPTG08]	100	71.6	87.9	94.6
Snakes	[APP*09]	100	43.2	95.2	84.7
	[PPTG08]	80.0	23.7	28.7	62.4
	BoW	88.0	20.8	25.0	64.3
Spectacles	BoW	100	90.2	98.7	99.0
	[APP*09]	100	70.3	99.8	94.0
	[PPTG08]	96.0	53.5	63.3	85.9
Spiders	[APP*09]	100	87.2	100	98.4
	BoW	100	71.6	96.0	93.6
	[PPTG08]	100	71.5	91.0	93.7
Teddy	BoW	100	96.6	100	99.9
	[PPTG08]	100	90.3	98.4	99.1
	[APP*09]	100	45.3	63.2	83.9

We have also tested our method on the SHREC 2007 Watertight dataset; figure 5 presents the Precision vs Recall plots of our method and the recent method from Toldo et al. [TCF09] which is also based on Bag of Words. The two algorithms present quite comparable performances; the method from Toldo et al. [TCF09] owns a better precision for low recall values while our algorithm is better for high recall values.

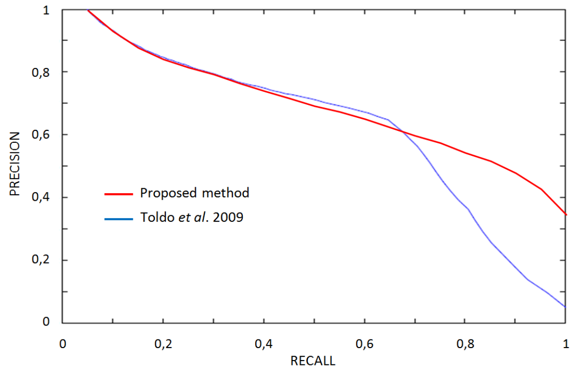

Figure 5: *Precision vs Recall curves, for the Shrec 2007 database, of the proposed approach and the BoW algorithm from Toldo et al. [TCF09]*

5.4. Partial shape similarity

Partial shape similarity is a quite complex problem tackled by few methods. We have tested and compared the performance of our algorithm on the SHREC 2007 Partial retrieval dataset. This dataset contains a *query set* of 30 shapes which are compared against a *testing set* of 400 models (the SHREC 2007 Watertight dataset). Each of the query models is composed of sub-parts from two or three models from the *testing set*; for each query object, a ground-truth classification of each model of the *testing set* is provided (highly relevant, marginally relevant or non relevant). The visual dictionary was built by clustering the descriptors of the 400 models of the *testing set*.
Figure 6 illustrates some query models and the top-8 results returned by our algorithm; we can observe that despite the difficulty of the task, all the retrieved objects are relevant except the Horse returned at 4^{th} position in the bottom row. In particular, in the bottom row, despite an important cropping and a large scale difference, the armadillo model is well recognized by our system.

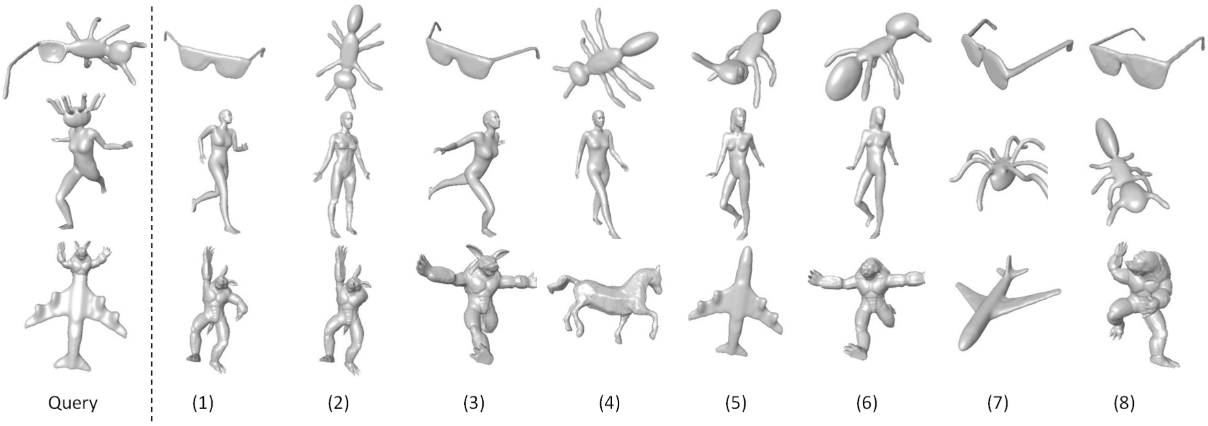

Figure 6: *Some examples of query objects from the SHREC 2007 Partial retrieval dataset and the top-8 retrieved models.*

We conducted a quantitative performance evaluation using the Normalized Discounted Cumulated Gain vector (NDCG) as proposed in the SHREC 2007 Partial retrieval contest; for a given query, the value NDCG[i] represents basically the relevance of the top-*i* results, it is recursively defined as:

$$DCG[i] = \quad G[i] \quad if\ i = 1 \qquad (10)$$

$$DCG[i] = \quad DCG[i-1] + G[i]log_2(i) \quad otherwise \qquad (11)$$

where $G[i]$ is a gain value depending on the relevance of the i^{th} retrieved model (2 for highly relevant, 1 for marginally relevant and 0 otherwise). The Normalized Discounted Cumulated Gain vector (NDCG) is then obtained by dividing the *DCG* by the ideal cumulated gain vector. Figure 7 illustrates the *NDCG* plots for our method and several methods from the state of the art:

- The BoW method from Toldo et al. [TCF09]
- The graph-based technique from Tierny et al. [TVM09]
- The best runs of the two methods from the SHREC 2007 Partial retrieval contest [VF07]: the extended reeb graphs (ERG) and the curve-skeleton based many-to-many matching (CORNEA).

Our method clearly outperforms the other algorithms, even most recent ones [TCF09, TVM09]. This is probably due to the fact that our method completely discards the structural information, hence the topological changes due to the subpart merging do not affect very much the bags of words. Moreover the descriptive power of our spectral descriptor efficiently discriminates the relevant regions of each model.

Our approach is also computationally efficient; for instance, for the Partial retrieval dataset where the average model size is 18K vertices, the whole indexing of a model (uniform point sampling, local spectral descriptor calculation, Bag of Word construction) takes an average of 25 seconds per model. Our implementation is based on the CGAL library (C++) and runs on a 2GHz laptop.

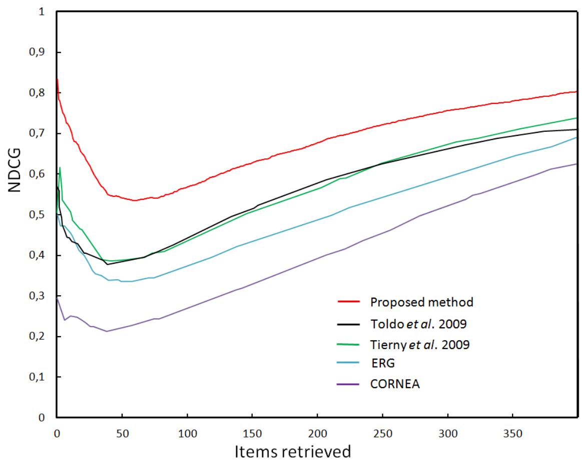

Figure 7: *NDCG curves of several methods, for the SHREC 2007 Partial retrieval dataset.*

6. Conclusion

We have presented a new robust 3D shape retrieval method based on the Bag of Words paradigm; the proposed approach relies on a uniform sampling of feature points associated with a new local Fourier descriptor both fast to compute and discriminative. Our algorithm is particularly suited for partial similarity scenarios where it clearly outperforms the state-of-the-art.

A weakness of our method is that, whereas it correctly retrieves a model from a partial query, it does not perfom the precise matching between the corresponding sub-parts. A solution to perform this matching would be to construct a graphical structure over the set of feature points and to apply some kind of fast approximate sub-graph isomorphism, robust to non rigid deformations.

Acknowledgements

We gratefully acknowledge Roberto Toldo for sending us his results. This work has been partly supported by the Cluster 2 of the Rhône-Alpes region within the LIMA Project.

References

[APP*09] AGATHOS A., PRATIKAKIS I., PAPADAKIS P., PERANTONIS S., AZARIADIS P., SAPIDIS N.: Retrieval of 3D Articulated Objects using a graph-based representation. In *Eurographics Workshop on Shape Retrieval* (2009), pp. 1–8. 2, 5, 6

[BBGO11] BRONSTEIN A. M., BRONSTEIN M. M., GUIBAS L. J., OVSJANIKOV M.: Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics 30*, 1 (2011), 1–20. 2

[BBK*09] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R., MAHMOUDI M., SAPIRO G.: A Gromov-Hausdorff Framework with Diffusion Geometry for Topologically-Robust Non-rigid Shape Matching. *International Journal of Computer Vision 89*, 2-3 (2009), 266–286. 1, 2

[BCG08] BEN-CHEN M., GOTSMAN C.: Characterizing shape using conformal factors. In *Eurographics Workshop on Shape Retrieval* (2008), Citeseer, pp. 1–8. 1

[BWWM10] BOWERS J., WANG R., WEI L.-Y., MALETZ D.: Parallel Poisson Disk Sampling with Spectrum Analysis on Surfaces. In *SIGGRAPH Asia* (2010). 3

[CTSM03] CHEN D., TIAN X., SHEN Y.-T., MING OUHYOUNG: On visual similarity based 3D model retrieval. *Computer Graphics Forum 22*, 3 (2003), 223–232. 1

[DLL*10] DEY T., LI K., LUO C., RANJAN P., SAFA I., WANG Y.: Persistent Heat Signature for Pose-oblivious Matching of Incomplete Models. *Computer Graphics Forum 29*, 5 (2010), 1545–1554. 2, 3

[FFP05] FEI-FEI L., PERONA P.: A bayesian hierarchical model for learning natural scene categories. *Computer Vision and Pattern Recognition* (2005), 524–531. 2, 3

[FMK*03] FUNKHOUSER T., MIN P., KAZHDAN M., CHEN J., HALDERMAN A., DOBKIN D., JACOBS D.: A search engine for 3D models. *ACM Transactions on Graphics (TOG) 22*, 1 (2003), 83. 1

[FS06] FUNKHOUSER T., SHILANE P.: Partial matching of 3D shapes with priority-driven search. In *Eurographics symposium on Geometry processing* (2006), Eurographics Association, p. 142. 2

[GSCO07] GAL R., SHAMIR A., COHEN-OR D.: Pose-oblivious shape signature. *IEEE transactions on visualization and computer graphics 13*, 2 (2007), 261–71. 1

[JZ07] JAIN V., ZHANG H.: A spectral approach to shape-based retrieval of articulated 3D models. *Computer-Aided Design 39*, 5 (2007). 1, 4

[LG09] LI X., GODIL A.: Exploring the Bag-of-Words method for 3D shape retrieval. *IEEE International Conference on Image Processing* (2009), 437–440. 2

[LGS10] LIAN Z., GODIL A., SUN X.: Visual Similarity based 3D Shape Retrieval Using Bag-of-Features. *Shape Modeling International* (2010). 2

[Llo82] LLOYD S.: Least squares quantization in PCM. *IEEE Transactions on Information Theory 28*, 2 (1982), 129–137. 3

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision 60*, 2 (2004), 91–110. 1

[LZ10] LÉVY B., ZHANG H.: Spectral mesh processing. *Siggraph 2010 Course* (2010). 4

[LZQ06] LIU Y., ZHA H., QIN H.: Shape Topics: A Compact Representation and New Algorithms for 3D Partial Shape Retrieval. *Computer Vision and Pattern Recognition* (2006), 2025–2032. 2

[MPSF10] MARINI S., PATANÉ G., SPAGNUOLO M., FALCIDIENO B.: Feature Selection for Enhanced Spectral Shape Comparison. *Eurographics Workshop on 3D Object Retrieval* (2010). 1, 4

[MS05] MIKOLAJCZYK K., SCHMID C.: Performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence 27*, 10 (2005), 1615–30. 1

[OOF*08] OHBUCHI R., OSADA K., FURUYA T., BANNO T., OTHERS: Salient local visual features for shape-based 3D model retrieval. *Shape Modeling International* (2008). 2

[PPTG08] PAPADAKIS P., PRATIKAKIS I., THEOHARIS T., G: 3D object retrieval using an efficient and compact hybrid shape descriptor. *Eurographics Workshop on 3D Object Retrieval* (2008). 5, 6

[RPSS09] RUGGERI M. R., PATANÈ G., SPAGNUOLO M., SAUPE D.: Spectral-Driven Isometry-Invariant Matching of 3D Shapes. *International Journal of Computer Vision 89*, 2-3 (2009), 248–265. 2

[Rus07] RUSTAMOV R.: Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In *Eurographics symposium on Geometry processing* (2007), pp. 225 – 233. 1, 4

[RWP06] REUTER M., WOLTER F., PEINECKE N.: Laplace-Beltrami spectra as Shape-DNA of surfaces and solids. *Computer-Aided Design 38*, 4 (2006), 342–366. 1, 4

[SCF10] SUN J., CHEN X., FUNKHOUSER T.: Fuzzy Geodesics and Consistent Sparse Correspondences For Deformable Shapes. *Computer Graphics Forum 29*, 5 (2010), 1535–1544. 2

[SMKF04] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The princeton shape benchmark. *Shape Modeling International* (2004), 167–388. 5

[SOG09] SUN J., OVSJANIKOV M., GUIBAS L.: A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. *Computer Graphics Forum 28*, 5 (2009), 1383–1392. 2

[TCDV10] TABIA H., COLOT O., DAOUDI M., VANDEBORRE J.: 3D-Shape Retrieval using Curves and HMM. *IEEE International Conference on Pattern Recognition* (2010). 2

[TCF09] TOLDO R., CASTELLANI U., FUSIELLO A.: Visual vocabulary signature for 3D object retrieval and partial matching. *Eurographics Workshop on 3D Object Retrieval* (2009). 2, 3, 6, 7

[TVM09] TIERNY J., VANDEBORRE J., M: Partial 3D shape retrieval by Reeb pattern unfolding. *Computer Graphics Forum 28*, 1 (2009), 41–55. 2, 3, 7

[VF07] VELTKAMP R. C., FRANK B. TER HAAR: SHREC2007: 3d shape retrieval contest, Technical Report UUCS- 2007-015, 2007. 7

[VL08] VALLET B., LÉVY B.: Spectral geometry processing with manifold harmonics. *Computer Graphics Forum 27*, 2 (2008), 251–260. 4

[WLBD09] WANG K., LUO M., BORS A., DENIS F.: Blind and robust mesh watermarking using manifold harmonics. In *IEEE International Conference on Image Processing* (2009), no. 3, pp. 211–226. 4