Noname manuscript No.
(will be inserted by the editor)

Combination of Bag-of-Words Descriptors for Robust Partial Shape Retrieval

Guillaume Lavoué

Received: date / Accepted: date

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 very informative, 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, hence discarding any spatial information. A spatially-sensitive algorithm is also presented where the 3D shape is described by an histogram of pairs of visual words. We show that these two approaches are complementary and can be combined to improve the performance and the robustness of the retrieval. The performances have been compared against very recent state-of-the-art methods on several different datasets. For global shape retrieval our combined approach is comparable to these recent works, however it clearly outperforms them in the case of partial shape retrieval.

Keywords Shape Retrieval · 3D model · Bag of Words

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, local deformation or cropping.

In that context we introduce a new combined 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 four main reasons:

- The regular sampling of the feature points is mostly independent of the connectivity, geometry or topology of the model (on the contrary, with a protrusion detector or a segmentation algorithm some feature points or 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 discards most of the structural information of the feature points, hence it is intrinsically

Guillaume Lavoué
Université de Lyon, CNRS
INSA-Lyon, LIRIS UMR 5205, France
E-mail: glavoue@liris.cnrs.fr

invariant to isometric deformations or topological changes.
- Our approach combines a standard Bag-of-Words descriptor with a spatially-sensitive one, hence reinforcing the robustness.

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 [40,39,11,9, 16].
A preliminary version of this work, describing only the standard BoW descriptor has been presented at the 3DOR 2011 conference [19].

The paper is organized as follows, section 2 presents the related state of the art on 3D shape retrieval. Section 3 provides a recall on the Bag of Words principles and an overview of our approach. Section 4 describes our feature point detector and descriptor while section 5 presents our indexing/retrieval methods using BoW. Finally, section 6 presents some experiments, which evaluate the robustness of our method and provide a comparison with state of the art methods.

2 State of the art

In this section we review the existing works on 3D shape retrieval; starting with methods based on *global* descriptors we then present *local* frameworks usually based on salient feature point detection and description. Finally we detail the *Bag of Words* approaches, which constitute the most recent class of techniques for 3D shape retrieval.

2.1 Global methods

The earliest techniques introduced to tackle the problem of 3D shape retrieval were based on global descriptors; the first were only robust to rigid deformations [6, 13], while more recent ones are also invariant to non rigid deformations, like isometry or skeletal articulation. Except the work of Gal et al. [15] which relies on histograms of local shape diameter values, most of these recent invariant descriptors are based on some spectral embeddings: Reuter et al. [31] and Marini et al. [26] describe the shape by the eigenvalues of the Laplace-Beltrami operator while Rustamov [33] considers the eigenvectors of this operator, similarly, the approach from Jain and Zhang [17] relies on the eigenvectors of the affinity matrix; besides, the *conformal factor* descriptor from Ben-Chen et al. [2] and the diffusion distances introduced by Bronstein et al. [5] are also based on the Laplace-Beltrami operator (eigendecomposition for [5] and integration into a sparse linear system for [2]). 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.

2.2 Local detectors and descriptors

To face the hard robustness issues not handled by global methods, some researchers turned their attention to *local* descriptors associated with salient feature points (or keypoints), following successful approaches in 2D image recognition like SIFT [25]. In such keypoint-based 2D image recognition techniques [28], the object to recognize is represented by a set of salient local features (usually sparse) associated with local descriptors, then the recognition consists in 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 [14] 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. Li and Guskov introduced feature point detector and descriptor inspired by SIFT and applied them for rigid alignement of point sets. Their feature points, combined with RANSAC were recently applied for rigid partial matching of archaeological objects by Itskovich and Tal [16]. Sun et al. [37] 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. [5] as a global descriptor. Very recently Sun et al. [36] combined the HKS-based feature points with a matching framework based on fuzzy geodesics while Dey et al. [9] 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. [1] 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. [38] also create regions related to surface protrusions (detector from Tierny et al. [39]) and then describe them using a set of curves. Ruggeri et al. [32] 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. Lastly, Sipiran et al. [35] generalize the Harris point detector for 3D meshes.

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.

2.3 Bag of Words approaches

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 [8,10], 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. [29] and Lian et al. [22] 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. [23] and Li and Godil [21] introduce BoW algorithms based on Spin Image descriptors computed on a dense set of feature points (uniformly sampled on the surface). Bronstein et al. [4] 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. [37]. Differently, Toldo et al. [40] 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 [4,21]. 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. [40], or you consider a dense collection like [4]; 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 [40] 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 [4].

To resolve these issues, our algorithm relies on a connectivity-independent uniform sampling of the feature points based on Lloyd relaxations and on a very discriminative spectral descriptor.

3 Overview of our method

Figure 1 illustrates our approach. 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.

Note that in the following technical sections, scalars are represented by lower case letters, vectors are represented as bold lower case letters and matrices are represented as capital letters.

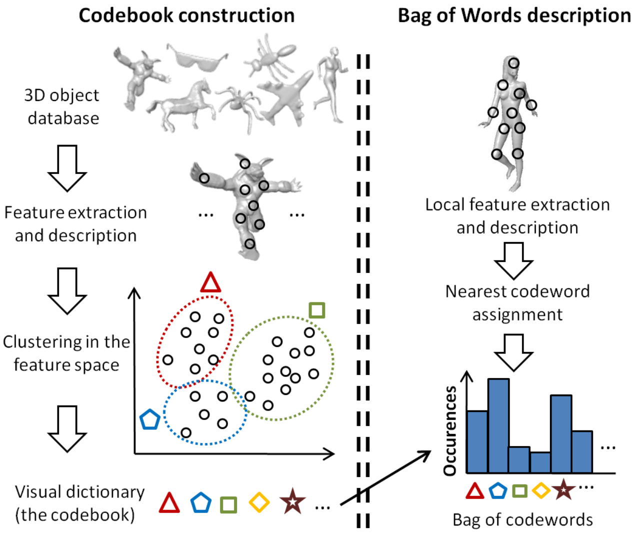
Codebook construction
Bag of Words description
3D object database
Feature extraction and description
...
...
Clustering in the feature space
Visual dictionary (the codebook)
Local feature extraction and description
Nearest codeword assignment
Occurences
...
Bag of codewords

Fig. 1 Flow chart of our Bag of Words approach.

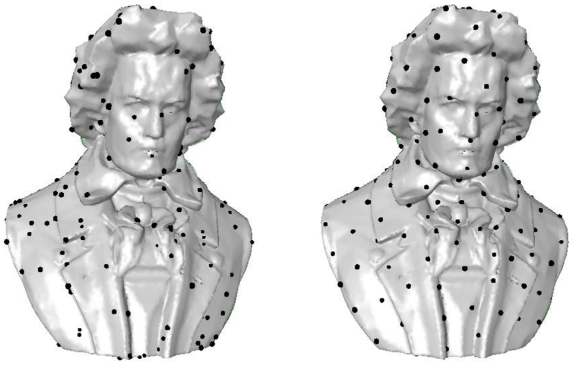

Fig. 2 Illustration of the Lloyd's relaxation algorithm. *Left:* 200 seeds randomly sampled. *Right:* result after 50 Lloyd's iterations

4 Feature point detection and description

4.1 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 [10] 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 [24] 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 [12]. 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.

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 [3].

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 overlapping 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 length of the bounding box of the object.

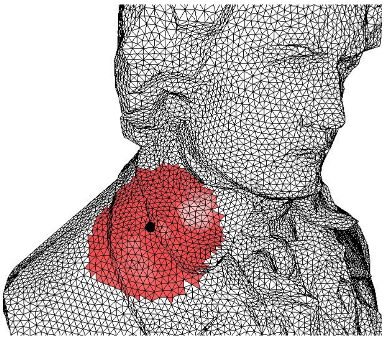

Fig. 3 A feature point (black sphere) and its associated patch (in red).

4.2 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 Fourier spectra of the patch, computed by projecting the geometry onto the eigenvectors of the Laplace-Beltrami operator. 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. [31,33,17,26]), while we consider the spectral transform coefficients (after projection of the 3D signal onto the eigenvectors). Surprisingly this has never done before whereas these spectral coefficients are particularly discriminative and also robust to noise and connectivity changes like pointed in [42].
- 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 saves a lot of 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 eigen-problem can be discretized and simplified within the finite element modeling framework [20]:

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

$\mathbf{h}^k$ denotes the vector $[H_1^k, ...H_m^k]$ 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 defined as:

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

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

$\beta_{i,j}$ 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 [41], 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 are then calculated as the inner product between the geometry of the surface and the sorted eigenvectors. Let $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ be the m-dimensional vectors containing respectively the x, y and z values of the m vertices. For $\mathbf{x}$ (resp. $\mathbf{y}$,$\mathbf{z}$):

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

The k^{th} ($k = 1..m$) spectral coefficient amplitude is then defined as the norm of $[\tilde{x}_k, \tilde{y}_k, \tilde{z}_k]$:

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

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.

This descriptor owns some interesting theoretical robustness properties [42]: under a translation, only the first coefficient c_0 is modified, hence we do not consider c_0 in our descriptor and thus obtain translation robustness. Meanwhile, it can be easily proven that the manifold harmonics bases are kept unchanged under isometric transformations. Therefore, a rotation in the spatial domain x,y,z yields the same rotation in the spectral domain $\tilde{x}$,$\tilde{y}$,$\tilde{z}$, without any influence on the coefficient amplitudes c_k. It can also be demonstrated that under a uniform scaling with a factor s, all the spectral coefficients will be scaled by s^2. Hence this descriptor is not robust to scaling but that does not constitute a problem since the whole 3D object is normalized before processing.

We have also studied experimentally the discriminative power and the robustness of the descriptor: we have considered one arbitrary surface patch from the Stanford bunny and applied several strong distortions on it (noise addition and simplification). Figure 4 illustrates the first 30 spectral amplitudes c_k of the resulting patches. We can observe the very high stability of the descriptor regarding these distortions. On the contrary when considering other patches with different shapes (see figure 5) then the descriptors are very different, hence implying a very good discriminative power.

5 3D object representation and matching

5.1 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

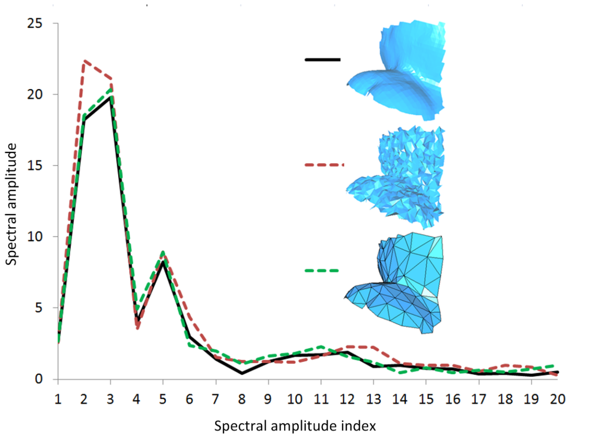

Fig. 4 Spectral amplitudes of a surface patch (in black) and distorted versions (dotted lines) under strong noise addition and simplification.

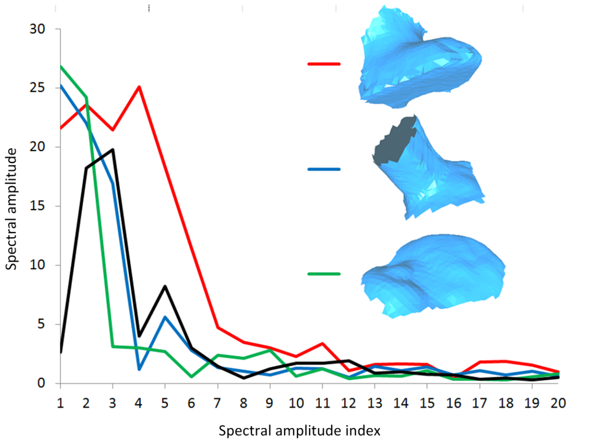

Fig. 5 Spectral amplitudes of several different surface patches, including the patch from figure 4 (in black).

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.

5.2 BoW representation and matching

5.2.1 Standard BoW

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:

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

$$b^i_j = 0 \ \ otherwise \tag{8}$$

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 6, 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.

The distance between two shapes is thus computed using a simple $L1$ distance between their bag of words.

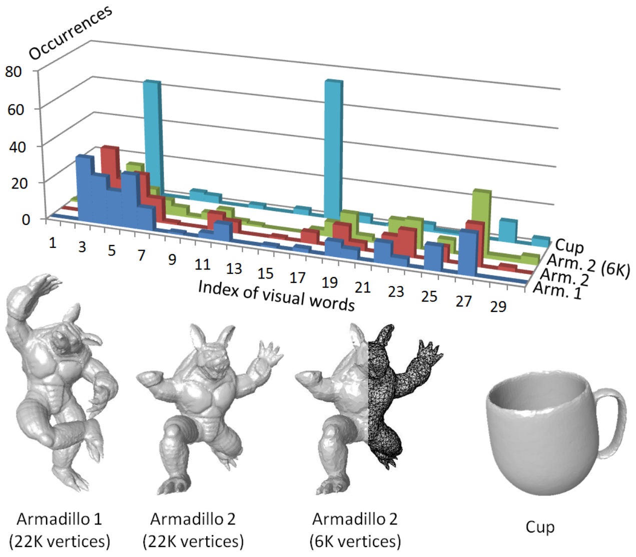

Fig. 6 Bag of Words examples.

5.2.2 Spatially-sensitive BoW

Like Bronstein et al. [4], we introduce here a spatially-sensitive version of our BoW descriptor. In this version, instead of a histogram of visual words, we construct a histogram of pairs of spatially-close visual words. The objective is to slightly take into account the spatial relations between the features. Our *spatial* bag of words B^M is a $(n_w \times n_w)$-dimensional matrix defined as follows:

$$B^M = \sum_{u=1}^{n_p}\sum_{v=1}^{n_p} \delta_{uv}\mathbf{b}^u(\mathbf{b}^v)^T \tag{10}$$

where δ_{uv} defines the following proximity rule:

$$\delta_{uv} = 1 \; if \; P_u \; and \; P_v \; are \; direct \; neighbors. \tag{11}$$

$$\delta_{uv} = 0 \;\; otherwise \tag{12}$$

The neighborhoods of the patches are extracted from the Voronoi regions associated to their seeds (see figure 7). Since the Voronoi diagram is centroidal, each patch owns 6 direct neighbors on average.

As for the 1-dimensional standard Bag of Words, the distance between two shapes represented by these 2-dimensional *spatial* BoWs is computed using a simple $L1$ distance between the matrices.

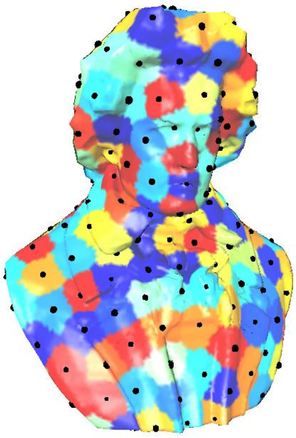

Fig. 7 Voronoi cells of the seeds, used to compute the neighborhood of the patches.

5.2.3 Hybrid BoW

We believe that the *standard* and *spatial* BoW descriptors presented in sections above could be complementary; indeed the standard BoW gives its best results with a quite high number of visual words n_w (200 as shown in the experimental section), while the spatial BoW provides optimal results for a much lower number of visual words (a typical value is 30); this low value appears logical, indeed for $n_w = 30$ then the real size of the vocabulary of pairs of words is $n_w^2 = 900$ which is already a high value. Hence the standard BoW really focuses on the discriminative power of the patches while the spatial one rather considers a coarser level of accuracy but associates it with a piece of spatial information.
Hence we have chosen to combine these *standard* and *spatial* BoWs into an *hybrid* one. This hybrid representation is simply the combination of both BoWs; then the distance between two models M_1 and M_2 using this hybrid representation is the combination of the $L1$ distances between both BoWs:

$$d(M_1, M_2) = 6 \times \left|\mathbf{b}^{M_1} - \mathbf{b}^{M_2}\right| + \left|B^{M_1} - B^{M_2}\right| \tag{13}$$

The factor *6* is introduced to balance the masses between both BoWs. Indeed since the typical neighborhood size of the patches is 6 then the mass of a spatial BoW B^M is $6 \times n_p$ while the mass of a standard BoW $\mathbf{b}^M$ is n_p.

6 Experiments

We have conducted a set of experiments to evaluate the complementarity of our methods and their performances regarding the state of the art. The first experiment studies the influence of the parameters; a second experiment evaluate the performance of our methods in term of global shape retrieval while the last one considers a partial shape retrieval scenario. In all these experiments we compare our standard, spatial and hybrid BoW algorithms as well as several recent state-of-the-art methods. The next section describes the databases which were used in these experiments.

6.1 Databases and measures

To test our algorithms we have considered three existing databases:

- The McGill Database [1]. 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 [2]. 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 [3]. 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 assess the efficiency of the methods we use the following measures, using the tools from [34]:

- Nearest Neighbor (NN): The percentage of queries for which the closest match belongs to the query's class.

[1] http://www.cim.mcgill.ca/s̃hape/benchMark/
[2] http://watertight.ge.imati.cnr.it/
[3] http://partial.ge.imati.cnr.it/

- 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.

6.2 Influence of the parameters

Our algorithms are 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. Table 1 presents the corresponding performances in term of the First Tier measure, for our standard algorithm. 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 to : $n_c = 40$, $n_w = 200$ and $n_p = 200$. For the spatial version, we use the same values, except for the size of the dictionary which has to be smaller; indeed in that case we build histograms of word pairs hence the practical size of the vocabulary is n_w^2. We have conducted experiments similar to the McGill tests above and found that a size $n_w = 30$ provides the best results (actually results were very similar for a range $n_w \in [20, 40]$).

6.3 Global Shape Retrieval

Firstly we have compared the performance of our own methods on the McGill and Shrec 2007 Databases. Tables 2 and 3 show the comparisons between the standard, spatial and hybrid Bag of Words algorithms. Two main conclusions can be drawn from these two tables: (1) the standard and spatial approaches provide similar results on both databases, with a little advantage for the standard one and (2) these two approaches seem complementary. Indeed the hybrid algorithm always provides the best results (in bold in the tables). Actually the spatial version is more efficient when the topology of the 3D models is more discriminative than their local geometry patterns, hence the complementarity between them. For instance the spatial version produces better results than the standard one for the *crabs*, *octopuses*, *spiders* and *snakes* classes of the McGill database and the *chairs*, *octopuses*, *springs* and *tables* classes of the Shrec 2007 database; all these classes own a rather poor local geometric information but can be discriminated using higher level topological notions such as the fact that a tubular surface is linked to a planar surface in the case of the *tables* class for instance.

Table 1 First Tier measure of our standard method for different parameter settings.

n_c	30	40	50
$n_p = 200 \Vert n_w = 200$	0.621	0.629	0.624

n_w	100	200	300
$n_p = 200 \Vert n_c = 40$	0.619	0.629	0.630

n_p	100	200	300
$n_w = 200 \Vert n_c = 40$	0.611	0.629	0.634

Table 2 Average retrieval statistics of our BoW algorithms for the McGill database.

Method	NN	FT	ST	DGC
Standard BoW	**95.7**	62.9	77.5	87.9
Spatial BoW	93.3	62.5	78.3	87.9
Hybrid BoW	**95.7**	**63.5**	**79.0**	**88.6**

Table 3 Average retrieval statistics of our BoW algorithms for the Shrec 2007 database.

Method	NN	FT	ST	DGC
Standard BoW	90.2	59.0	73.4	84.1
Spatial BoW	89.7	56.7	71.5	83.3
Hybrid BoW	**91.8**	**60.0**	**74.0**	**84.7**

We have then compared the performance of our hybrid BoW method with two recent algorithms on the McGill Database: the graph-based approach from Agathos et al. [1] and the hybrid 2D/3D approach from Papadakis et al. [30]. Table 4 presents the results; for each row, the algorithms are positioned according to their respective performances. The first remark is that the graph-based algorithm [1] provides the best results, that is logical since the database considers only skeletal articulation deformations without topology changes hence it is particularly suited for graph-based representation. However we can notice that our method, whereas considering almost no structural information, provides quite good results, almost always better than [30].

Table 4 Retrieval statistics for the McGill database.

Class	Method	NN	FT	ST	DGC
Whole	[1]	97.6	74.1	91.1	93.3
	Hybrid BoW	95.7	63.5	79.0	88.6
	[30]	92.5	55.7	69.8	85.0
Ants	[30]	100	73.6	89.2	94.8
	Hybrid BoW	96.7	57.7	86.2	88.7
	[1]	96.7	54.9	79.7	88.4
Crabs	[1]	100	98.2	99.8	99.9
	Hybrid BoW	100	60.7	81.0	92.3
	[30]	100	55.2	71.8	88.7
hands	[1]	95.0	83.9	88.9	95.2
	Hybrid BoW	100	51.1	68.9	85.8
	[30]	90.0	43.4	57.6	77.8
humans	[1]	96.6	93.5	96.4	98.1
	Hybrid BoW	100	70.8	90.6	93.8
	[30]	100	47.0	63.8	83.1
Octopuses	[1]	88.0	58.8	81.8	88.1
	[30]	56.0	29.5	45.0	68.9
	Hybrid BoW	68.0	26.8	39.3	67.5
Pliers	[1]	100	100	100	100
	Hybrid BoW	100	88.7	97.6	98.6
	[30]	100	71.6	87.9	94.6
Snakes	[1]	100	43.2	95.2	84.7
	Hybrid BoW	92.0	24.7	27.0	66.2
	[30]	80.0	23.7	28.7	62.4
Spectacles	Hybrid BoW	100	86.7	97.5	98.4
	[1]	100	70.3	99.8	94.0
	[30]	96.0	53.5	63.3	85.9
Spiders	[1]	100	87.2	100	98.4
	Hybrid BoW	100	77.7	98.7	95.2
	[30]	100	71.5	91.0	93.7
Teddy	Hybrid BoW	100	96.1	100	99.8
	[30]	100	90.3	98.4	99.1
	[1]	100	45.3	63.2	83.9

We have also compared our hybrid method on the Shrec 2007 Watertight dataset. Figure 8 presents the Precision vs Recall plots of our method and the recent method from Toldo et al. [40] which is also based on Bag of Words. The two algorithms present quite comparable performances, however our algorithm is slightly better, indeed its precision is higher for low and high recall values.

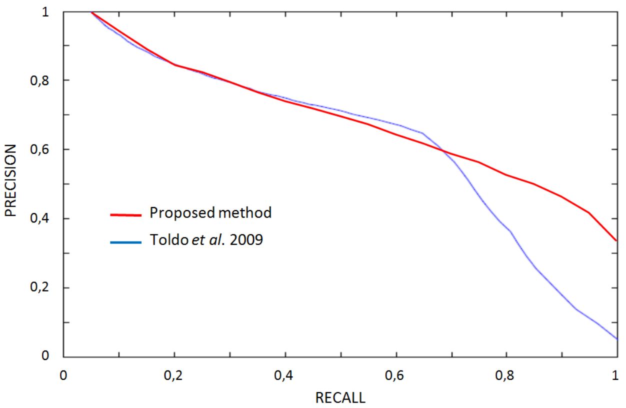

Fig. 8 Precision vs Recall curves, for the Shrec 2007 database, of the proposed hybrid approach and the BoW algorithm from Toldo et al. [40]

6.4 Partial shape similarity

Partial shape similarity is a quite complex problem tackled by few methods. We have tested and compared the performance of our algorithms 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 subparts 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).

Figure 9 illustrates some query models and the top-8 results returned by our hybrid algorithm; we can observe that despite the difficulty of the task, almost all the retrieved objects are relevant. In particular, in the bottom row, despite an important cropping, the giraffe model is well recognized by our system, as well as the plane.

We conducted a quantitative performance evaluation using the Normalized Discounted Cumulated Gain vector (NDCG) [18]. For a given query, the value NDCG[i] represents basically the relevance of the top-i results, it is recursively defined as:

$$DCG[i] = \qquad G[i] \qquad if\ i = 1 \tag{14}$$

$$DCG[i] = DCG[i-1] + G[i]log_2(i)\ otherwise \tag{15}$$

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 10 illustrates the respective performances of our algorithms. It is interesting to notice that the standard

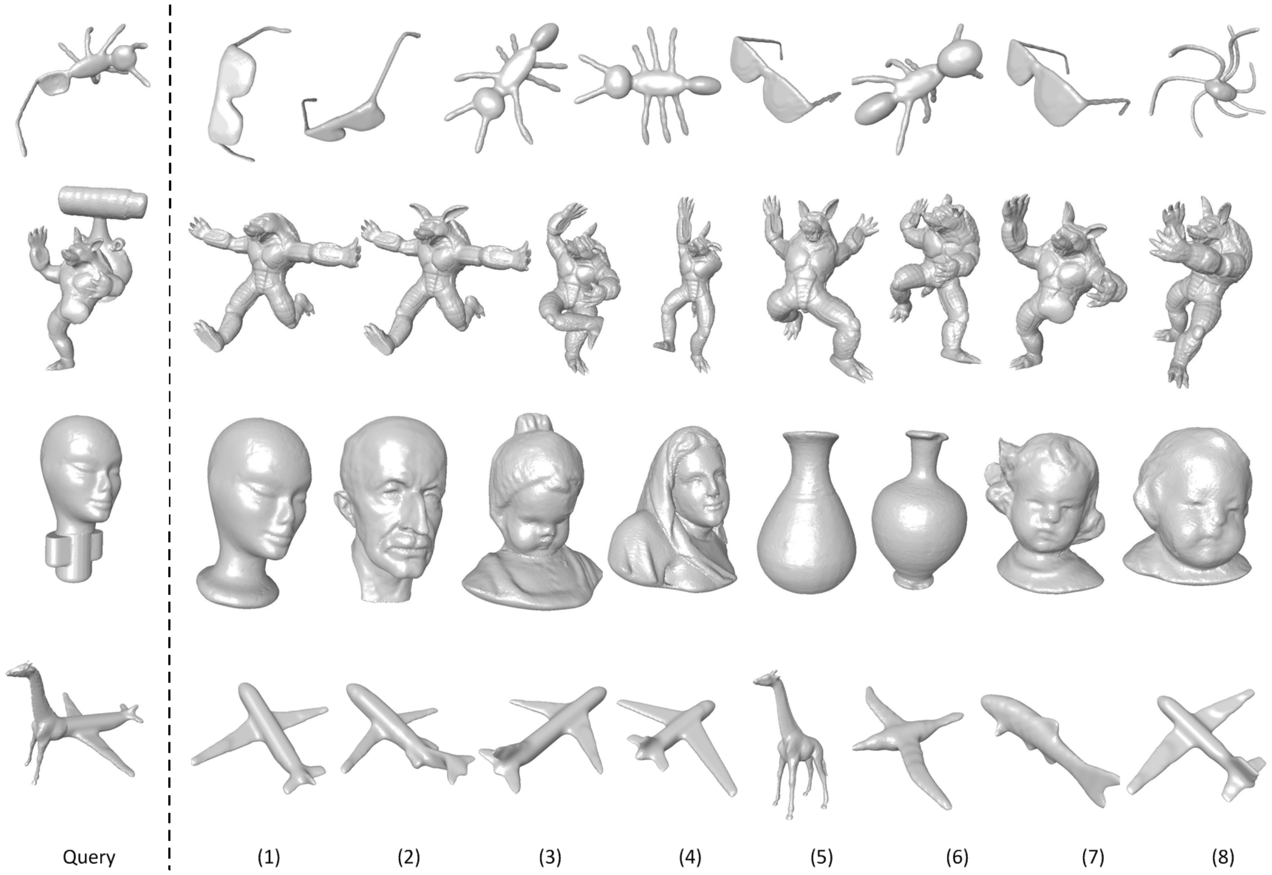

Fig. 9 Some examples of query objects from the SHREC 2007 Partial retrieval dataset and the top-8 retrieved models.

method provides clearly better results than the spatial one. This is consistent with the theory, indeed the query objects have been constructed by cropping and pasting parts of the original ones, hence several spatial constraints over the objects have been broken and/or modified. However these methods remain complementary since the hybrid method is still the best one.

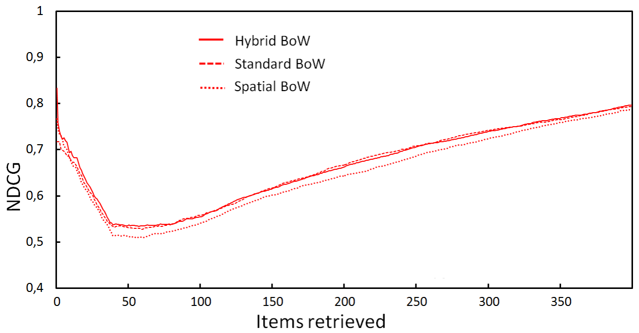

Fig. 10 NDCG curves of our methods, for the SHREC 2007 Partial retrieval dataset

Figure 11 illustrates the *NDCG* plot for our hybrid method and several methods from the state of the art:

- The BoW method from Toldo et al. [40]
- The graph-based technique from Tierny et al. [39]
- The best runs of the two methods from the SHREC 2007 Partial retrieval contest: the extended reeb graphs (ERG) [27] and the curve-skeleton based many-to-many matching (CORNEA) [7].

Our method clearly outperforms the other algorithms, even most recent ones [40,39]. This is probably due to the fact that our method discards most of the structural information, hence the topological changes due to the sub-part 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.

6.5 Robustness

The robustness of a shape retrieval method is a critical issue for its practical use, we evaluate in this section the robustness of our algorithm against noise addition

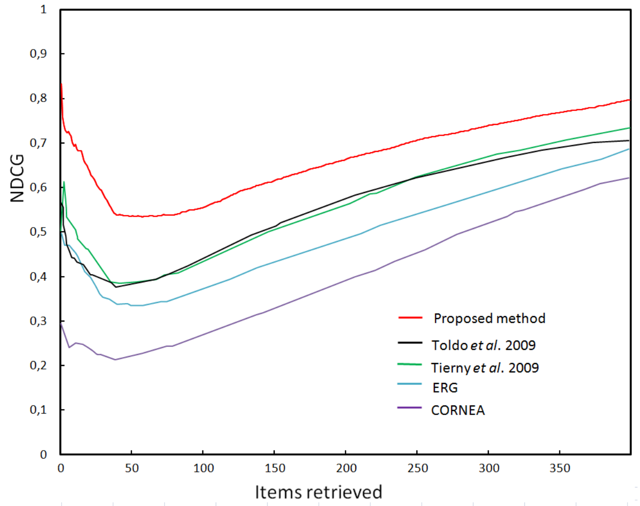

Fig. 11 NDCG curves of several methods, for the SHREC 2007 Partial retrieval dataset.

and simplification. For this purpose, we have distorted the query models and evaluated their partial shape retrieval performance (same protocol than the previous section). Figure 12 illustrates the distortions: Gaussian noise addition (two amplitudes) and canonical simplification (one to three iterations).

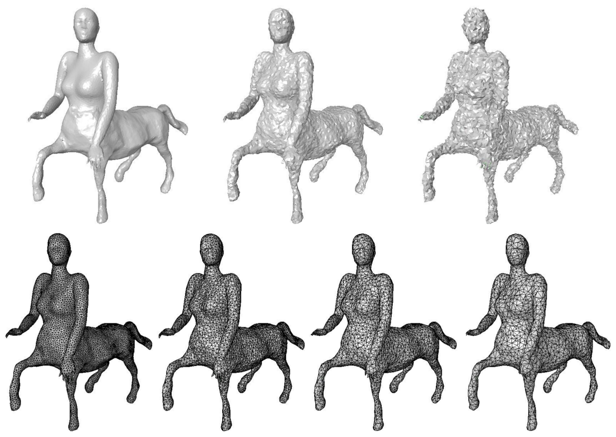

Fig. 12 Examples of distortions on a query model. *Top from left to right:* Original, results after 0.5% Gaussian noise and 1% Gaussian noise (in percentage of the bounding box). *Bottom from left to right:* Original (12728 vertices), results after one iteration (8435 vertices), two iterations (5464 vertices) and three iterations (3509 vertices).

Figures 13 and 14 illustrate the retrieval performances of our hybrid algorithm according to the distortions. We can observe that for smallest distortions (0.5% noise and 1 iteration of simplification) there are almost no performance loss; moreover a very interesting point is that even for strong distortions (1% noise and 3 iterations of simplification), although there is a significant loss of performance, results remain better than state-of-the-art methods (see figure 11). Providing such a high robustness against both connectivity and geometry distortions is a very good point, especially in such a partial retrieval scenario.

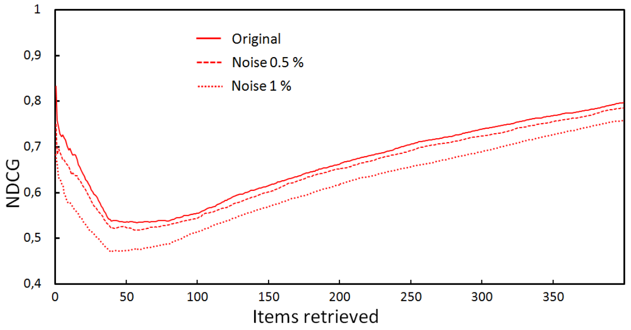

Fig. 13 NDCG curves of our hybrid method for different level of noise applied to the query objects.

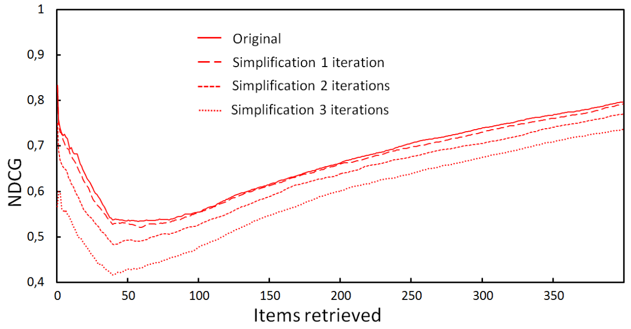

Fig. 14 NDCG curves of our hybrid method for different level of simplification applied to the query objects.

Figure 15 illustrates the retrieved results from distorted queries. It can be observed that even if the retrieved models vary according to the distortion of the query, in each case they are relevant.

6.6 Timing

Our is 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.

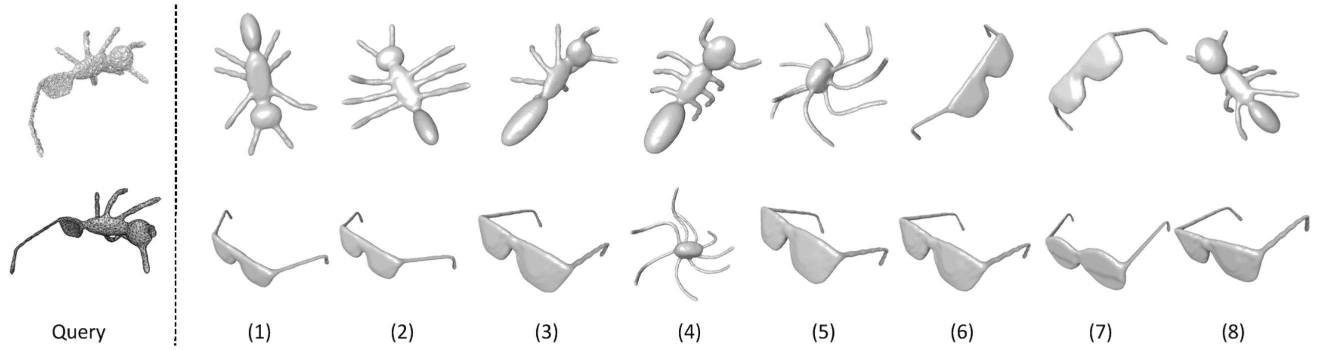

Fig. 15 Distorted query objects (highest noise and simplification) and top-8 retrieved models.

7 Conclusion

We have presented a new robust 3D shape retrieval method which combines standard and spatially-sensitive Bag of Words; 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. We have also shown that standard and spatially-sensitive methods are complementary since their combination provides a significant gain with regards to their individual performances.
It would be interesting to conduct a quantitative evaluation of our local Fourier descriptor on the TOSCA database, particularly on the SHREC 2010/2011 correspondence benchmarks [4].
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 thank the anonymous reviewers for helping us to greatly improve this paper.

References

1. 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 3D Object Retrieval (2009)
2. Ben-Chen, M., Gotsman, C.: Characterizing shape using conformal factors. In: Eurographics Workshop on 3D Object Retrieval (2008)
3. Bowers, J., Wang, R., Wei, L.y., Maletz, D.: Parallel Poisson Disk Sampling with Spectrum Analysis on Surfaces. In: SIGGRAPH Asia (2010)
4. 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), 1–20 (2011)
5. 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), 266–286 (2009)
6. Chen, D., Tian, X., Shen, Y.T., Ming Ouhyoung: On visual similarity based 3D model retrieval. Computer Graphics Forum **22**(3), 223–232 (2003)
7. Cornea, N., Demirci, M.: 3D Object Retrieval using Many-to-many Matching of Reconstruction-based Curve Skeletons. In: R.C. Veltkamp, Frank B. ter Haar (eds.) SHREC2007 3D Shape Retrieval Contest, pp. 50–52 (2007)
8. Csurka, G., Dance, C., Fan, L., Willamowski, J., C: Visual categorization with bags of keypoints. ECCV International Workshop on Statistical Learning in Computer Vision (2004)
9. 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), 1545–1554 (2010)
10. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. Computer Vision and Pattern Recognition pp. 524–531 (2005)
11. Ferreira, A., Marini, S., Attene, M., Fonseca, M.J., Spagnuolo, M., Jorge, J.a., Falcidieno, B.: Thesaurus-based 3D Object Retrieval with Part-in-Whole Matching. International Journal of Computer Vision **89**(2-3), 327–347 (2009)
12. Fu, Y., Zhou, B.: Direct sampling on surfaces for high quality remeshing. Computer Aided Geometric Design **26**(6), 711–723 (2009)
13. 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), 83 (2003)
14. Funkhouser, T., Shilane, P.: Partial matching of 3D shapes with priority-driven search. In: Eurographics symposium on Geometry processing, p. 142 (2006)
15. Gal, R., Shamir, A., Cohen-Or, D.: Pose-oblivious shape signature. IEEE transactions on visualization and computer graphics **13**(2), 261–71 (2007)

[4] http://tosca.cs.technion.ac.il/book/shrec_correspondence.html

16. Itskovich, A., Tal, A.: Surface partial matching & application to archaeology. In: Computer Graphics Forum (2011)
17. Jain, V., Zhang, H.: A spectral approach to shape-based retrieval of articulated 3D models. Computer-Aided Design **39**(5) (2007)
18. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS) **20**(4), 422 (2002)
19. Lavoué, G.: Bag of words and local spectral descriptor for 3d partial shape retrieval. In: Eurographics Workshop on 3D Object Retrieval (2011)
20. Lévy, B., Zhang, H.: Spectral mesh processing. Siggraph 2010 Course (2010)
21. Li, X., Godil, A.: Exploring the Bag-of-Words method for 3D shape retrieval. IEEE International Conference on Image Processing pp. 437–440 (2009)
22. Lian, Z., Godil, A., Sun, X.: Visual Similarity based 3D Shape Retrieval Using Bag-of-Features. Shape Modeling International (2010)
23. Liu, Y., Zha, H., Qin, H.: Shape Topics: A Compact Representation and New Algorithms for 3D Partial Shape Retrieval. Computer Vision and Pattern Recognition pp. 2025–2032 (2006)
24. Lloyd, S.: Least squares quantization in PCM. IEEE Transactions on Information Theory **28**(2), 129–137 (1982)
25. Lowe, D.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60**(2), 91–110 (2004)
26. Marini, S., Patané, G., Spagnuolo, M., Falcidieno, B.: Feature Selection for Enhanced Spectral Shape Comparison. Eurographics Workshop on 3D Object Retrieval (2010)
27. Marini, S., S. Biasotti, Spagnuolo, M., Falcidieno, B.: Sub-part Correspondence by Structural Descriptors of 3D Shapes. In: Remco C. Veltkamp, Frank B. ter Haar (eds.) SHREC2007 3D Shape Retrieval Contest, pp. 53–55 (2007)
28. Mikolajczyk, K., Schmid, C.: Performance evaluation of local descriptors. IEEE transactions on pattern analysis and machine intelligence **27**(10), 1615–30 (2005)
29. Ohbuchi, R., Osada, K., Furuya, T., Banno, T., Others: Salient local visual features for shape-based 3D model retrieval. Shape Modeling International (2008)
30. 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)
31. Reuter, M., Wolter, F., Peinecke, N.: Laplace-Beltrami spectra as Shape-DNA of surfaces and solids. Computer-Aided Design **38**(4), 342–366 (2006)
32. 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), 248–265 (2009)
33. Rustamov, R.: Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In: Eurographics symposium on Geometry processing, pp. 225 – 233 (2007)
34. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The princeton shape benchmark. Shape Modeling International pp. 167–388 (2004)
35. Sipiran, I., Bustos, B.: Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes. The Visual Computer (2011)
36. Sun, J., Chen, X., Funkhouser, T.: Fuzzy Geodesics and Consistent Sparse Correspondences For Deformable Shapes. Computer Graphics Forum **29**(5), 1535–1544 (2010)
37. Sun, J., Ovsjanikov, M., Guibas, L.: A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. Computer Graphics Forum **28**(5), 1383–1392 (2009)
38. Tabia, H., Colot, O., Daoudi, M., Vandeborre, J.P.: 3D-Shape Retrieval using Curves and HMM. IEEE International Conference on Pattern Recognition (2010)
39. Tierny, J., Vandeborre, J.P., Daoudi, M.: Partial 3D shape retrieval by Reeb pattern unfolding. Computer Graphics Forum **28**(1), 41–55 (2009)
40. Toldo, R., Castellani, U., Fusiello, A.: Visual vocabulary signature for 3D object retrieval and partial matching. Eurographics Workshop on 3D Object Retrieval (2009)
41. Vallet, B., Lévy, B.: Spectral geometry processing with manifold harmonics. Computer Graphics Forum **27**(2), 251–260 (2008)
42. Wang, K.: Quantization-Based Blind Watermarking of Three-Dimensional Meshes. Phd, Institut National des Sciences Appliquées de Lyon (2009)