Vis Comput
DOI 10.1007/s00371-014-0938-1

ORIGINAL ARTICLE

A sparse coding approach for local-to-global 3D shape description

Davide Boscaini · **Umberto Castellani**

Abstract The definition of reliable shape descriptors is an essential topic for 3D object retrieval. In general, two main approaches are considered: global, and local. Global approaches are effective in describing the whole object, while local ones are more suitable to characterize small parts of the shape. Recently some strategies to combine these two approaches have been proposed which are mainly concentrated to the so-called bag of words paradigm. With this paper we address this problem and propose an alternative strategy that goes beyond the bag of word approach. In particular, a sparse coding technique is exploited for the 3D domain: a set of local shape descriptors are collected from the shape, and then a dictionary is trained as generative model. In this fashion the dictionary is used as global shape descriptor for shape retrieval purposes. Several experiments are performed on standard databases in order to evaluate the proposed method in challenging situations like the case of 'SHREC 2011: robustness benchmark' where strong shape transformations are included, and the case of 'SHREC 2007: partial matching track' where composite models are considered in the query phase. A drastic improvement of the proposed method is observed by showing that sparse coding approach is particularly suitable for local-to-global description and outperforms other approaches such as the bag of words.

D. Boscaini
University of Lugano, Lugano, Switzerland
e-mail: davide.boscaini@usi.ch

U. Castellani (✉)
University of Verona, Verona, Italy
e-mail: umberto.castellani@univr.it

Keywords 3D object retrieval · Sparse coding · Bag of words · Partial shape matching

1 Introduction

3D shape retrieval methods are important to deal with the continuously increased availability of 3D models [9,14,18,20,34]. A general aim is to define a proper representation of the objects in order to improve the indexing phase for data retrieval purposes. In the literature, such representation is called shape *descriptor* or *signature*.

Roughly speaking, descriptors can be *global* or *local* [14,19,34]. The former consist of a set of features that effectively and compactly describe the whole 3D model [14]. The latter, instead, are collections of local features of relevant object subparts (i.e. single points or regions) [14,19,34].

In this paper, we address the problem of combining the two approaches by defining a global shape descriptor starting from a set of local point signatures [1,8,12,33]. In this fashion it is possible to obtain the twofold advantage: from one side we are able to compare global shapes rather than a set of single points. From the other side, we exploit local information which, in general, is more robust to noise and missing parts and more suitable to deal with partial objects. A popular method consists of introducing a counting procedure by collecting local characteristics into a histogram which provides a local-to-global signature. Examples of this paradigm are shape distance distributions (SDD) [14] or the *bag of words* approach [6,12,36].

With this work we propose to go beyond the bag of words approach by exploiting recently proposed dictionary learning methods employing sparse coding techniques [23,24]. Starting from a set of local signatures we learn a *dictionary* which is able to summarize the most relevant properties of such set.

Published online: 23 April 2014

This leads to a more sparse representation of the shape which is used for its description. We propose this approach in a shape retrieval context. In particular, we design and evaluate two matching strategies: (i) shape-to-shape, and (ii) shape-to-class. In shape-to-shape case we train a dictionary from the set of point-signatures extracted in a single shape (i.e., one dictionary per shape). In shape-to-class case we train a dictionary from all point-signatures of shape belonging to the same class (i.e., one dictionary per class). Once the dictionary have been learned, in the query phase, a given shape is generated by all available dictionaries and it is assigned to the class or shape with less generative error.

A well-defined shape retrieval pipeline is proposed by combining effectively the most promising local shape descriptors with the proposed *local-to-global* approach based on sparse coding. The main steps are:

- Local descriptors computation,
- Dictionary learning by sparse coding,
- Shape matching by best generative signature estimation.

In particular, as local descriptors we exploit two recently proposed approaches: (i) the Wave Kernel Signature [1] which adopts a diffusion geometry paradigm, and (ii) the local depth SIFT (LD-SIFT) [12] that extends the SIFT image descriptor [22] to meshes.

Then, dictionary learning method is applied by using the *Lasso* model [35]. Although such approach is quite popular in signal processing, to the best of our knowledge, only recently it has been proposed in computer vision for 2D image coding and very few work has been done in 3D domain (e.g., [28]).

Experiments are evaluated on two standard dataset, i.e. SHREC 2011 robustness benchmark [5] and SHREC 2007 partial matching track [25,37]. We show a drastic improvement over both state-of-the-art global shape descriptor, such as Shape DNA [29], and standard local-to-global approaches, such as shape distance distribution [14]. Furthermore, the proposed sparse coding approach is exhaustively compared with bag of words method [6,12,36] from both methodological and experimental aspects. A preliminary version of this work has appeared at the 3DOR 2013 conference [4].

The rest of the paper is organized as follows. Section 2 reports the state of the art by focusing mainly on the shape descriptors which are related to the proposed work. Section 3 introduces the two proposed local shape descriptors, namely WKS, and LD-SIFT. Section 4 describes the background of sparse coding and it introduces the proposed local-to-global approach. In Sect. 5 we discuss the main differences between the bag of words and the proposed sparse coding approach. Section 6 reports the experimental results by showing the performance of the proposed descriptor in comparison with other methods. Finally, in Sect. 7, conclusions are drawn and future work is envisaged.

2 Related work

Ideal shape descriptors should satisfy properties like discriminativeness, invariance to pose and shape deformation, robustness to noise, compactness and so on (see, e.g., [21,29]). Several work have been proposed in order to define reliable descriptors satisfying these properties. In the following we revise the main approaches for 3D shape description namely global, local, and local-to-global methods. For a more detailed overview we refer to [19].

2.1 Global methods

Regarding *global* methods we highlight two different paradigms: (i) methods based on the deformation-invariant representation of the shape, and (ii) methods based on spectral shape analysis. In the first approach, in order to deal with deforming objects, the non-rigid shape is modified to obtain a 'standard' common pose. To this aim Elad and Kimmel [13] proposed a multidimensional scaling (MDS) algorithm by embedding the geodesic distances among the mesh vertices in a 3D Euclidean space. The resulting *canonical form* can be treated as a rigid shape, and can be compared to other shapes by a standard rigid alignment method such as iterative closest points (ICP) [3].

In the second approach, rather then directly matching the shapes, the comparison is carried out between their descriptors. A largely employed class of methods is based on eigendecomposition of Laplace–Beltrami Operator (LBO) of the underlying shape [14,17,34]. For instance Shape DNA [29] computes the spectral decomposition of the LBO defined on the manifold represented by the shape and uses the truncated set of the computed eigenvalues as global signature. This leads to a very effective descriptor which was successfully employed on several applicative scenarios, such as shape retrieval and shape matching in medical domain [11,18,29]. A similar approach is proposed in [30] where the signature is defined as the collection of the eigenvectors of the LBO normalized by the corresponding eigenvalues.

2.2 Local methods

Local descriptors are employed for point-to-point correspondences [10,14,34]. A common approach is to collect local geometric properties on the point neighborhood and accumulate these values on a multidimensional histogram. Examples are *Spin Images* [15] or Shape Context [2]. In particular, Wuhrer and colleagues [38] proposed to deal with deformable shapes by combining Spin Images with shape transformation methods. Surfaces are first projected onto the canonical form [13] and then local descriptors are computed [15]. Darom and Keller [12] proposed an extension of the well known SIFT descriptor [22] from images to shapes, namely the Local Depht SIFT (LD-SIFT). A feature detection phase is carried out in order to select the most relevant surface points. Then a local descriptor is computed by collecting local surface gradients in the neighborhood of the feature point. Here the scale invariance is obtain by estimating automatically the size of the local support. Other approaches exploit probabilistic properties of the shape. For intance in [10] a Hidden Markov Models is adapted to work on 3D surfaces and it is used as local generative descriptor.

Another popular class of local descriptors are based on the property of the heat kernel associated to the manifold. All of these methods can be grouped under the name *diffusion geometry*. In [33] the so-called heat kernel signature (HKS) was introduced which exploits the local surface properties at different scales. Some extensions of HKS are proposed in [8] to deal with scale invariance. Recently, in [1], the so-called wave kernel signature (WKS) was proposed. WKS employs a different physical model and therefore it is related with oscillation rather than heat diffusion processes.

2.3 Local-to-global methods

Local-to-global approaches are therefore introduced to define a global signature from a collection of local descriptors [6,11,12,14,36]. A simple method consists of computing pairwise distances among points in the descriptor space and accumulate these distances into a histogram [14]. In [11] the authors proposed the Global-HKS descriptor by collecting all the HKS signatures and defining a histogram for each scale. More sophisticate techniques exploit probabilistic methods, such as in [26], where a probabilistic fingerprint is introduced. Being encouraged by feature-based methods developed in Computer Vision, several work employed the so-called bag of words paradigm [6,12,16,36]. In [6] the bag of word descriptor is computed from the set of local HKS signatures. In [36] a region-based approach is introduced where the *visual words* are defined by region properties computed after shape segmentation. In [12] authors extracted the bag of words from local LD-SIFT signature described above. In [16] a spatial constraint is introduced in the bag of words process. Moreover, the spectral transform of the LBO is computed over local patches rather than the whole shape.

In this paper, we propose to exploit dictionary learning and sparse coding approaches [23,24]. To the best of our knowledge, sparse coding approach is very few adopted for 3D shapes and only recently some methods have been proposed, such as in [28], for point-to-point correspondences of non-rigid or partial shapes.

3 Local descriptors

In this work, we exploit two different approaches to describe local shapes: (i) descriptors based on diffusion geometry, and (ii) descriptors based on scale-invariant features.

3.1 Descriptors based on diffusion geometry

In the context of diffusion geometry, the most popular local descriptor is the HKS [33] and its scale-invariant version, SI-HKS [8]. They are based on the properties of the heat diffusion process on the shape governed by the heat equation

$$\left(\frac{\partial}{\partial t} - \Delta\right) u(x,t) = 0. \tag{1}$$

From the signal processing perspective the solution $u(x,t)$ of differential equation (1) at time t is defined by the convolution between the *impulse response* $h_t(x,t)$ and the initial data $u_0(x)$,

$$u(x,t) = \int h_t(x,y)u_0(y)\, d\sigma(y).$$

The kernel of this integral operator is called *heat kernel* and it corresponds to the amount of heat transferred from point x to point y after time t. In particular, HKS represents the autodiffusion process $h_t(x,x)$ centered in a vertex x of the shape, at different time scales.

As described in [21], the heat kernel descriptor could be thought as a collection of low-pass filters. This emphasizes how low frequencies damage the ability of the descriptor to precisely localize shape features. A remedy to the poor feature localization of the heat kernel descriptor was proposed by the so-called WKS [1]. The authors proposed to replace the heat diffusion equation (1), by the Shrödinger equation:

$$\left(\frac{\partial}{\partial t} + i\Delta\right) v(x,t) = 0,$$

where $v(x,t)$ is the complex wave equation. Here the physical interpretation is different: it represents the average probability of measuring a quantum particle with a certain energy distribution at a specific location. That is, instead of representing diffusion, v has oscillatory behavior.

Letting vary the energy of the particle, the WKS encodes and separates information from various different frequencies. Similarly to the former signal processing interpretation of HKS, the WKS can be thought as a collection of *band-pass* filters [21]. As a result, the wave kernel descriptor exhibits superior feature localization properties.

3.2 Descriptors based on scale-invariant features.

Darom and Keller [12] proposed to extend the popular scale-invariant feature transform (SIFT) [22] from images to meshes. Their method consists of two steps: (i) detection of feature points, (ii) computation of the scale-invariant local descriptors.

In the image domain, the SIFT operator achieves the detection of interest points invariant to image scale and rotation, exploiting the property of the DoG operator. The main problem in the generalization of the DoG operator from a flat 2D images to a curved 3D shape is that, in the latter case, the vertices of the mesh are non-uniformly sampled. To overcome this difficulty, in [12] the authors proposed a *density-invariant* Gaussian filter on the mesh geometry. This gives rise to a set of filtered meshes. Feature points are then extracted as local maxima both in scale and location. The main characteristic of the method consists of providing each interest point with an adaptive support proportional to the filter width. Indeed, by controlling the local scale the tradeoff between locality and robustness to noise is carried out.

Once feature points are obtained the local scape descriptors are computed. More in details, Darom and Keller [12] compute a plane at each interest point using the two leading eigenvectors of the PCA around the interest point. In particular, the neighborhood of the interest point is defined by the adaptive local scale by providing *scale invariance*. The descriptor is the depth map computed by projecting the vertices around the interest point onto the estimated plane. In order to gain *rotation invariance*, the authors define a dominant angle in the local plane and cyclically shift the radial-polar histograms within the SIFT descriptor, such that the dominant angle is the zero angle.

4 Global shape descriptor by sparse coding

In this section we give a brief overview of general theoretical background on sparse coding. Then, we highlight our main contribution that is how to exploit sparse coding to propose a global signature from a set of local descriptors.

4.1 Background

A general problem in machine learning and pattern recognition can be formulated as follows (see, e.g., [31]). Given two classes of objects x_i, and a new previously unseen object x, how can we assign the unknown object to the right class? To distinguish the objects belonging to a class from the others, we assign them a *label* y_i, i.e.

$$(x_1, y_1), \ldots, (x_n, y_n) \in X \times \{\pm 1\},$$

where the labels are chosen as $+1$ and -1 for the sake of simplicity and X is some non-empty set containing the *patterns* x_i. Given some new pattern $x \in X$, we want to infer the corresponding label $y \in \{\pm 1\}$. To this end an interpolation function f on the given data, i.e.

$$\min_f \|y_i - f(x_i)\|_2^2$$

is not effective since it is not able to generalize well for unseen patterns. A possible approach to avoid this problem is suggested by Tychonoff regularization theory and consist of a restriction of the class of admissible solutions to, e.g., a compact set. Indeed, the previous problem can be reformulated as:

$$\min_f \|y_i - f(x_i)\|_2^2 + \lambda R(f),$$

where

- $\|y_i - f(x_i)\|_2^2$ is the data term,
- $R(f)$ is the *regularization term*,
- $\lambda > 0$ is the so-called *regularization* parameter which specifies the trade-off between fidelity to the data in the sense of ℓ^2 norm, as represented by the former term, and simplicity of the solution, enforced by $R(f)$.

An example of regularization operator is $R(f) = \|f^{(m)}\|_2$, for some $m \in \mathbb{N}$. This particular choice promotes the smoothness of the solution.

Let we now address a slightly different problem. Given a sentence s and a dictionary D, we want to explain the sentence s with words contained in D. This problem could be defined as

$$\min_\alpha \|s - \alpha D\|_2^2,$$

where the idea is that the vector α picks up only the words that describe the sentence s. In general, a dictionary is *overcomplete*: there are a lot of words with the same or similar meaning. For this reason we might be interested to consider the minimum number of words as possible. Again regularization theory help us. If we consider the problem

$$\min_\alpha \|s - \alpha D\|_2^2 + \lambda R(\alpha), \tag{2}$$

with the choice $R(\alpha) = \|\alpha\|_1$ we are promoting the sparsity of the solution. In this case we refer as *sparse coding* and the corresponding problem is known in the literature as *Lasso formulation* (see, e.g., [35]).

In general, as described in [28], s could be though as a generic signal and the interpretation of Lasso formulation could be the following: many families of signals can be represented as a sparse linear combination in an appropriate domain, usually referred to as *dictionary*, so that $s \approx \alpha D$. In other words, the signal s could be *generated* by αD. Finally, given the signal s and the dictionary D, the solution of the unconstrained convex minimization problem of equation (2) gives us the sparse vector α.

However, in general the dictionary D is not available. We therefore are interested in inferring both the vector α and the dictionary D from the signal s. The problem becomes:

$$\min_D \left(\min_\alpha \|s - \alpha D\|_2^2 + \lambda \|\alpha\|_1 \right). \tag{3}$$

In [23,24], problem (3) was solved employing an alternating minimization method between the variables D and α.

As a further step we should consider that in the more general case, instead of a single signal s, we have a collection of signals $\mathbf{s} = \{s^i\}_{i=1,\dots,N}$. Therefore, Equation (3) can be generalized as:

$$\min_D \frac{1}{N} \sum_{i=1}^{N} \min_{\alpha^i} \left(\|s^i - \alpha^i D\|_2^2 + \lambda \|\alpha^i\|_1 \right), \tag{4}$$

where $\alpha = \{\alpha^i\}_{i=1,\dots,N}$ is though as a collection of vectors.

4.2 Local-to-global descriptors

Since we want to apply sparse coding technique to the context of shape analysis, we consider as signals $\mathbf{s}$ the collection of local descriptors at each vertex of the considered shape. Then, learning techniques described in [23,24] are employed for solving problem (4), i.e. learn the dictionary D. In this paper we exploit two approaches: (i) shape-to-shape, and (ii) shape-to-class.

4.2.1 Shape-to-shape

In the shape-to-shape approach a dictionary is trained for each shape. In this fashion the estimated dictionary is a sort of global signature of the given shape. Then, in the query phase, given a new unseen object the retrieval is carried out by trying to generate it from all the available dictionaries and assign to it the shape associated to the dictionary which produces the less generative error. Note that in this approach the shapes are compared pairwise as usual for shape retrieval. More in details, given the set of n training shapes: $\{\mathbf{O}_1, \dots \mathbf{O}_n\}$, a set of n dictionaries are estimated $\{\mathbf{D}_1, \dots \mathbf{D}_n\}$. In particular a dictionary $\mathbf{D}_j$ is computed by solving problem (4) where the input $\mathbf{s}_j$ is the set of local signatures extracted from the shape $\mathbf{O}_j$.

In the query phase, given the query object $\mathbf{O}_q$ and its collection of local signatures $\mathbf{s}_q$, the retrieval is obtained by the following steps: (i) solve problem (2) for the case of multi-signals for each dictionary $\mathbf{D}_j$, which outputs is α_q^j, (ii) assign to the query object $\mathbf{O}_q$ the retrieved object $\mathbf{O}_{ret}$ such that:

$$\|\mathbf{s}_q - \alpha_q^{ret}\mathbf{D}_{ret}\| = \min_j \|\mathbf{s}_q - \alpha_q^j \mathbf{D}_j\|. \tag{5}$$

Note that in Eq. 5 a L_2 norm is computed which is very sensitive to outliers. This means that few wrong signals $s_q^i \in \mathbf{s}_q$ can affect the overall retrieval performance. In order to address this issue we introduce a robust approach. Let $\Delta^j = (\Delta_1^j, \dots, \Delta_N^j)$ the vector of generative errors of each signal, where $\Delta_i^j = (s_q^i - \alpha_q^j \mathbf{D}_j)$, and N is the number of signals in query shape $\mathbf{O}_q$. Let $\tilde{\Delta}^j$ be the ordered generative error vector aiming at organizing the error components in ascendent order. The robust overall generative error is therefore computed as:

$$\|\mathbf{s}_q - \alpha_q^{ret}\mathbf{D}_{ret}\| = \min_j \sum_{\tilde{i}=1}^{P} \tilde{\Delta}_{\tilde{i}}^j, \tag{6}$$

where $P < N$ represents the number of signal inliers. In this fashion, the highest error components, that are associated to ouliers, are excluded from the overall generative error computation.

Finally, if the query and the retrieved share the same label the retrieval phase is satisfied.

4.2.2 Shape-to-class

In this case we consider another matching strategy. It is worth noting that $\mathbf{s}$ could be considered also as the collection of local signatures of an entire class of shapes. For instance, several deformations of the same object or several objects that are instances of the same class can contribute in learning a joint dictionary. To this aim a single dictionary is trained for each class by therefore reducing the number of training procedures. In this fashion the matching is carried out at class level rather than at shape level. More precisely a *matching-by-recognition* strategy is employed by adopting a classification procedure based on the sparse coding framework. More in details:

- $\{\mathbf{O}_1^c, \dots \mathbf{O}_k^c\}$ are several instances of objects belonging to the same class c,
- k is the number of instances,
- $\mathbf{D}^c$ represents the dictionary of class c.

$\mathbf{D}^c$ is trained from all the signatures $\{\mathbf{s}_1^c, \dots \mathbf{s}_k^c\}$ of the instances of the class c. For example, if c represent a class of noise deformations of the same shape $\mathbf{O}$, then $\mathbf{O}_i^c$ represents a noise deformation of $\mathbf{O}$ and k the number of noise deformations of the shape $\mathbf{O}$ present in c.

Then, in the query phase, given a query shape $\mathbf{O}_q$ and its collection of local signatures $\mathbf{s}_q$, we solve (the multi-signals version of) problem (2) for each class-based dictionary $\mathbf{D}^{c_i}$ and we obtain the vectors $\alpha_q^{c_i}$. Finally, at the query shape $\mathbf{O}_q$ is then assigned the class c such that

$$\|\mathbf{s} - \alpha_q^c \mathbf{D}^c\| = \min_i \|\mathbf{s} - \alpha_q^{c_i} D^{c_i}\|. \tag{7}$$

Also in this case it is possible to compute a robust generative error as described in Sect. 4.2.1.

In retrieval applications, the principal advantage of this matching-by-recognition method is that it allows to compare the signatures of a query shape only with the dictionary of the classes of the shapes present in the database considered. Conversely, in standard shape-to-shape methods, the matching is done between the query shape and all the instances of the database. Another important advantage is that the dictionary can encode more instances of the same non-rigid deformation, making the proposed signature a descriptor of the entire class of deformations rather than a single shape descriptor.

5 Sparse coding and bag of words

There are clear analogies between the sparse coding and bag of words approaches. Both methods define a dictionary composed of *atoms* or *words* which are in somehow representatives of the input signals. The advantage consists of avoiding a point-to-point matching between shapes. More precisely, the matching is intermediated by the words of the dictionary: instead of performing a direct comparison between local signatures of two shapes, each signature of a given shape is compared with only the words of the dictionary. The main differences between sparse coding and bag of words are in the way such dictionaries are estimated. In the bag of words approach a dictionary is defined by a clustering procedure on the feature space where the centroids of such clustering become the visual words. In sparse coding instead a *generative* approach is introduced in order to represent a signal as a linear combination of the atoms. Indeed, in the bag of words approach an input signal is associated to a single visual word (i.e., the nearest). Conversely in sparse coding method an input signal can be associated to several atoms and, in general, the number of involved atoms depends by the strongness of the sparsity constraint. Moreover, a relevant difference between the two methods involves also the matching phase. In fact, in the bag of word paradigm the dictionary enables the construction of the bag of word descriptor and then shape matching is carried out by a direct comparison between such descriptors (in general L_2 norms is used). In the sparse coding method instead the shape matching is fully intermediated by the dictionaries. In particular, the dictionary is not used to compute a further descriptor, but it is employed in the matching phase to compute directly the matching error. It is worth noting that in [28] authors also proposed sparse coding method for shape matching. They formulated the problem of dense shape matching as a permuted sparse coding approach. In particular, they solved simultaneously for an unknown permutation ordering the regions on two shapes and for an unknown correspondence in functional representation [27]. This leads to a dense point-to-point matching, which is the main objective of this work. In our paper we differ from [28] since we obtain a shape comparison without requiring an explicit point-to-point matching.

6 Results

In this section we evaluate the proposed method on challenging datasets. SHREC 11 Robustness benchmark is useful to test the performance against strong shape deformations. In SHREC 2007 the evaluation is more focused on the capability of methods in dealing with partial or composed objects.

6.1 Experimental setup

According to the pipeline proposed in Sect. 4, we extract the local descriptors for each vertex of each shapes.

For the WKS [1] we based on the Matlab code freely available.[1] In all our experiments the parameters were fixed accordingly with [1]. In particular, we considered $n = 200$ eigenvalues of the LBO, a variance $\sigma = 6(\phi_2 - \phi_1)$ and we evaluate at $M = 500$ values of energy e, where $e_{\min} = \log \phi_1 + 2\sigma$ and $e_{\max} = \log \phi_n - 2\sigma$ and ϕ_i denotes the ith eigenvalue of the Laplace–Beltrami operator.

For the LD-SIFT we used the code available from the Matlab File exchange repository.[2] In all our experiments we used default parameters as described in [12].

Once local shape descriptors are computed, sparse coding is employed for local-to-global descriptor. For the numerical solution of the optimization problem (2) and (4), we use SPArse Modeling Software (SPAMS), an open-source optimization toolbox based on [23,24]. In all our experiments we make the trivial choice $\lambda = 1/2$. For the case of shape-to-shape matching we fix a dictionary of $M = 60$ atoms, while in the case of shape-to-class we define $M = 500$ atoms since the dictionary should be more expressive.

Finally, we compare the performances of our method with a global signature, namely Shape DNA [29], and with two well-known quantization approaches [14]: signature distance distribution (SDD) and bag of words (BoW).

Shape DNA signature [29] consist of the truncated spectrum of the LBO. For the application of this popular global descriptor to retrieval scenarios, we follow the suggestions reported in [18]. More specifically, we consider only the first 13 eigenvalues and rescale the spectrum by the shape's area to obtain the scale invariance of the descriptor.

Signature distance distribution (SDD) is a simple example of local-to-global descriptor. The idea consists of computing the histogram of Euclidean distances between pairs of point signatures randomly sampled on the shape. In order to capture the underlying geometry, the random selection of points are repeated several times. In our case, the random selection was repeated 10 times, leading to about 10^4 distances between local descriptors. The output is a histogram ables to discriminate between different shapes, as reported in Fig. 1. For matching purposes, at each pair of shapes we compute the ℓ^2 error between vectors of histogram occurrences.

Bag of word approach is the state-of-the-art method for local-to-global shape matching. In this paper we implemented the standard version. Starting from the set of all point

[1] http://vision.in.tum.de/publications.

[2] http://www.mathworks.com/matlabcentral/fileexchange/.

signatures extracted from all shapes a quantization procedure is computed by k-means clustering [31]. The number of centroids is defined in accordance with the number of atoms computed for the sparse coding approach, i.e., $k = 60$.

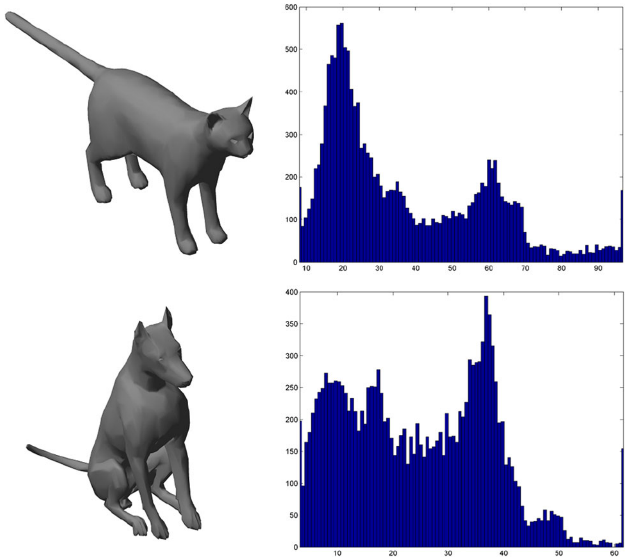


Fig. 1 Comparison between the SDD of the shapes of a cat and a dog. In the former there is a peak approximately around the 20th bin, in the latter around the 40th bin

6.2 SHREC 11 robustness benchmark

The database is composed of 12 different triangulated meshes from TOSCA [7] and Sumner [32] databases, that we consider as null shapes, and their non-rigid deformations. For each null shape reported in Fig. 2, transformations were split into 9 different types:

- *affine*,
- big *holes*,
- *micro holes*,
- *scale*,
- down *sampling* (less than 20 % of original points),
- additive Gaussian *noise*,
- *shot noise*,
- *topology* (welding of shapes vertices resulting in different triangulation), and
- *view*,

as reported in Fig. 3. Each triangular meshes has about 1,500 vertices.

Each type of transformation appeared in five different versions numbered from $1 - 5$. In all deformation types, the

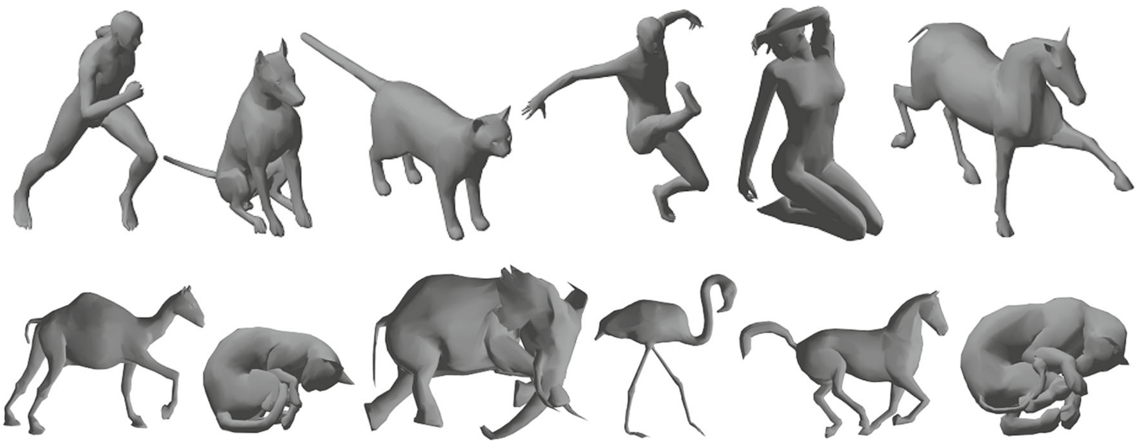


Fig. 2 Null shapes of our database taken from TOSCA [7] and Sumner [32] databases. From *left to right* we find man, dog, cat, man, woman, horse, camel, cat, elephant, flamingo, horse and lioness. In all our experiment we consider correct the matching between two men, cats or horses. The camel and elephant shapes are created by pose transfer from the galloping horse, and the lioness from pose transfer from the crouching cat

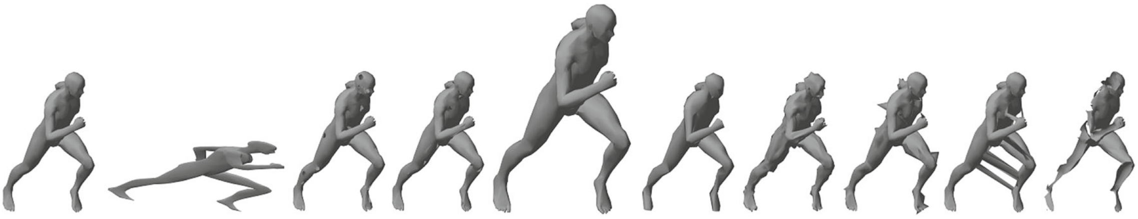


Fig. 3 Examples of deformations types considered in our database, taken from SHREC 11 robust benchmark. From *left to right* we find the null shape, affine, holes, micro holes, scale, sampling, noise, shot noise, topology and view

Table 1 Comparison between the nearest neighbor retrieved shape by SDD, shape DNA and sparse coding approach

Deformation	SDD		Shape DNA		BoW		Sparse coding	
	Corrects	%	Corrects	%	Corrects	%	Corrects	%
Affine	45/72	0.67	49/72	0.68	**60/72**	**0.83**	59/72	0.82
Holes	36/72	0.50	58/72	0.81	52/72	0.72	**65/72**	**0.90**
Microholes	**65/72**	**0.90**	64/72	0.89	**65/72**	**0.90**	**65/72**	**0.90**
Scale	**72/72**	**1.00**	71/72	0.99	71/72	0.99	**72/72**	**1.00**
Sampling	67/72	0.93	69/72	0.96	**72/72**	**1.00**	**72/72**	**1.00**
Noise	71/72	0.99	**72/72**	**1.00**	71/72	0.99	**72/72**	**1.00**
Shotnoise	70/72	0.97	**72/72**	**1.00**	**72/72**	**1.00**	**72/72**	**1.00**
Topology	58/72	0.80	55/72	0.76	68/72	0.94	**72/72**	**1.00**
View	11/72	0.15	16/72	0.22	37/72	0.51	**49/72**	**0.64**
Average	495/648	0.76	526/648	0.81	568/648	0.88	**598/648**	**0.92**

Best performance are highlighted in bold

version number correspond to the transformation strength level: the higher the number, the stronger the deformation (e.g., in noise transformation, the noise variance is proportional to the strength number). For scale transformation, the levels 1–5 correspond to scaling by the factor 0.5, 0.875, 1.25, 1.625 and 2.

For each class of deformations, we have 60 shapes, 5 for every null class. The entire database contains 552 shapes: the 540 deformed shapes and the 12 null shapes.

Comparison results are shown in Table 1. Note that we show shape-to-class strategy only since results are already quite satisfactory. In practice, the table shows the retrieval performance in terms of recognition accuracy obtained in a Nearest Neighbor principle. For this set of experiments we employed WKS signatures as local shape descriptor. The proposed sparse coding method clearly improves on the most of the deformation classes with respect to SDD, Shape DNA, or BoW. Only for affine transforms sparse coding performs slightly worse than BoW. In particular, it improves drastically in the class of big holes, topology and view deformations. It is worth noting that on view deformation the improvement with respect to Shape DNA was expected: a global descriptor fails to identify correctly partial views of a shape. The interesting facts with this kind of deformation is that our method evidently outperforms with respect that SDD which is a basic local-to-global approach. In noise and shot noise deformations it performs like Shape DNA although sparse coding approach consider significantly more eigenvalues. (it is a well known fact that the first eigenvalues are related to shape's lower-frequency contents, meanwhile higher eigenvalues is related to higher-frequency contents and manifest themselves as rough geometric features, i.e. shape details). Bag of words approach performs well as expected but overall the proposed sparse coding method shows better results, by evidencing a more stable and robust behavior.

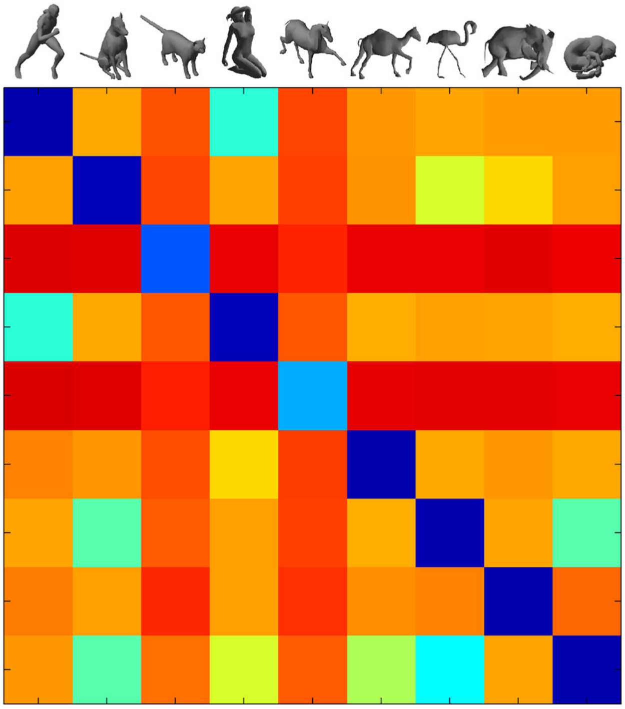


Fig. 4 Class-signature dissimilarity matrix for noise deformations. *Cold colors* represents lower values, *hot colors* represent higher value

Figure 4 reports the dissimilarity matrix of the class of noise deformations. Note that since we used shape-to-class strategy the matix is not a classical dissimilarity matrix between shape descriptors. Here, each column represents a class of shape while each row represents the mean error between the generative errors of shapes belonging to the same class (i.e., 5 non-rigid deformations and the underlying null shape). In this experiment we have omitted the second instances of repeated classes as man, cat and horse for a better visual result. Cold colors represent small error values, hot colors represent high error values. It is interesting to note that

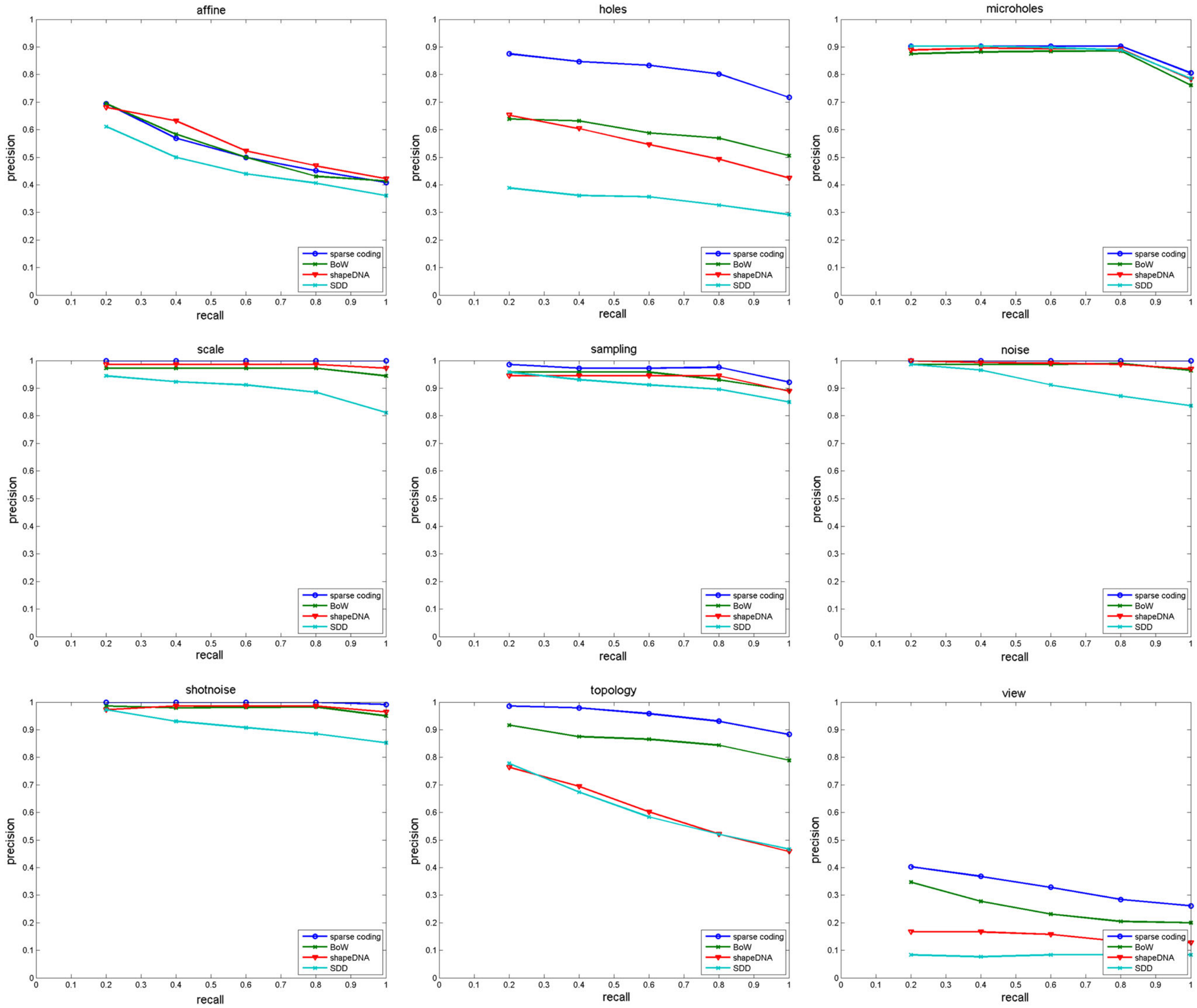


Fig. 5 Precision and Recall *curves for* each transformation

man and woman classes has small error with respect to other classes, this remark the *similarity* property of the proposed descriptor.

In order to further evaluate the retrieval performance we, show precision and recall curves for each transformation (Fig. 5). To compute this curve a shape-to-shape strategy is carried out. This performance confirmed the behaviour already observed in Table 1. Figure 6 reports the precision and recall curves for the overall experiment. As expected our method performs clearly better than others and overall BoW is superior than ShapeDNA and SDD.

6.3 SHREC 07 partial shape retrival

SHREC 2007 partial shape retrieval dataset is composed of a training set of 400 models grouped into 20 classes [37]. Every model is represented as a watertight manifold mesh. Figure 7 shows all 400 models of the 20 classes. Note that the dataset is challenging since there is a strong intra-class variability. Furthermore, there is a query set composed of 30 models each one obtained by combining subparts of the training set. In particular, for each query model a ground-truth classification is provided by defining which class is highly relevant, marginally relevant or non-relevant. Figure 8 shows the composite models of the query set.

In order to evaluate the retrieval performance the so-called Normalized Discounted Cumulative Gain curves are computed [37]. For this experiment we evaluated the shape-to-shape strategies since there is a strong intra-class variability and therefore the shape-to-class approach is likely to fail. Both the local shape signatures, namely WKS and LD-SIFT, are considered and combined with both sparse coding and Bag of Words quantization procedure. This leads to the following cases: (i) Sparse-WKS, (ii) Sparse-LD-SIFT,

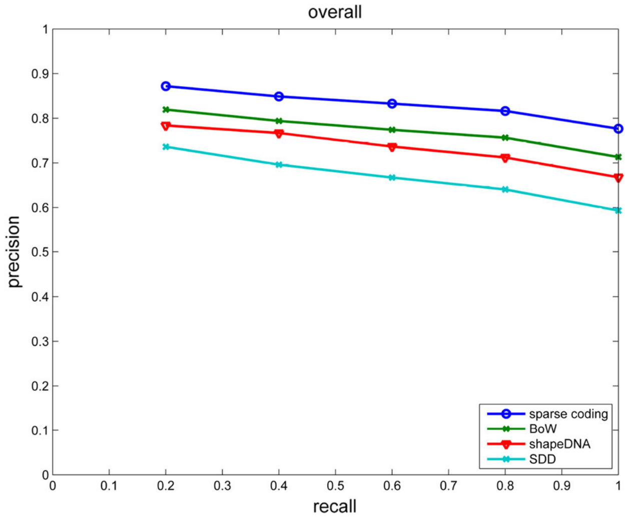


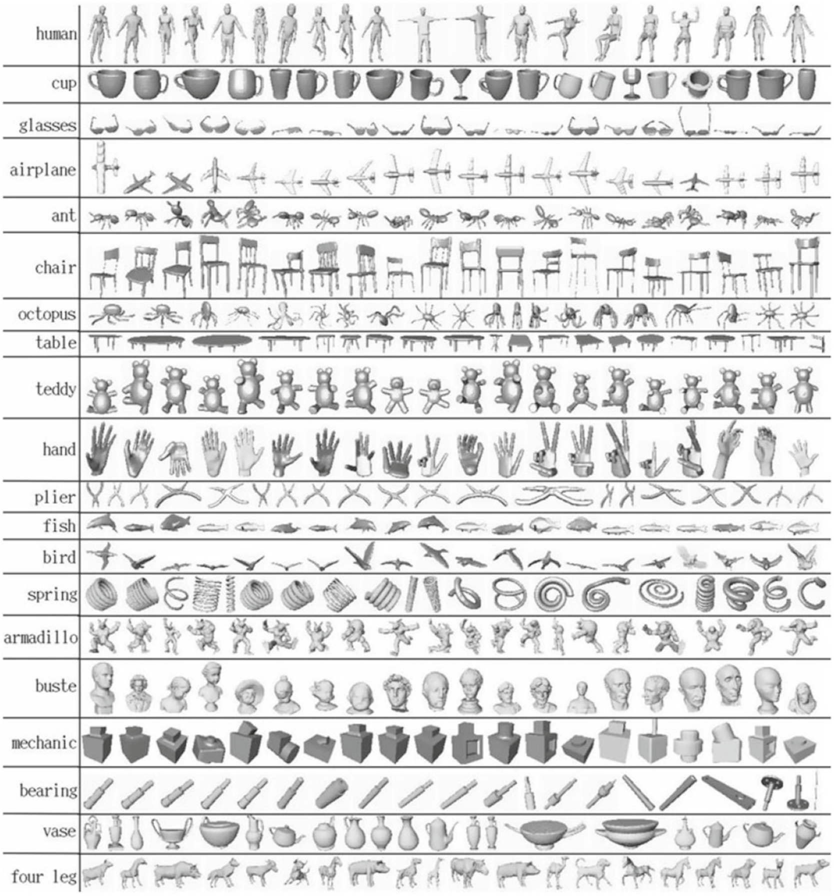
Fig. 6 Precision and recall *curves* of all transformations

(iii) BoW-WKS, and (iv) BoW-LD-SIFT. Figure 9 shows the performance of the evaluated methods. Results show a clear improvement of the sparse coding quantization method in comparison with the standard Bag of Words approach when the same local signatures are employed. Note that for these experiments the robust matching approach is carried out in order to reject the non-dominant part of the shape (i.e., outliers) from the overall error matching computation. In particular, a rejection rate of 50 % is employed by assuming that the dominant shape is at least the half of the whole shape.

Overall, the LD-SIFT descriptor performed better than WKS. There are two important differences between these methods. LD-SIFT is a fully local method since the descriptor is computed by collecting information from only the neighborhood of the feature point. Conversely, WKS is based on the spectral shape computation which is influenced by the whole shape. Moreover, LD-SIFT adopot a feature-based approach, i.e., only few interesting points are considered for the global signature construction. WKS instead uses all the shape vertices. Therefore, for this scenarios, since the composite shapes can change drastically the global shape with respect to their single shape components, it is reasonable that a fully local method like LD-SIFT performed at best.

Finally, we compared results with other methods proposed in literature for SHREC 07 partial shape retrival. Other than our sparse coding approaches we evaluated the method introduced by Toldo et al. [36], and methods proposed for

Fig. 7 SHREC 2007 partial shape retrieval dataset



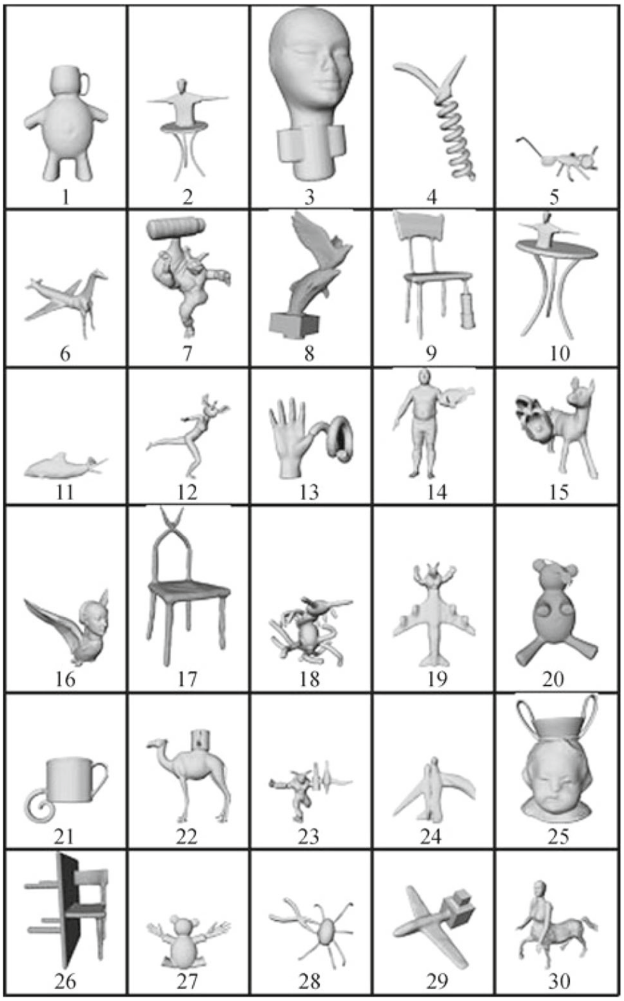


Fig. 8 SHREC 2007 partial shape retrieval query set

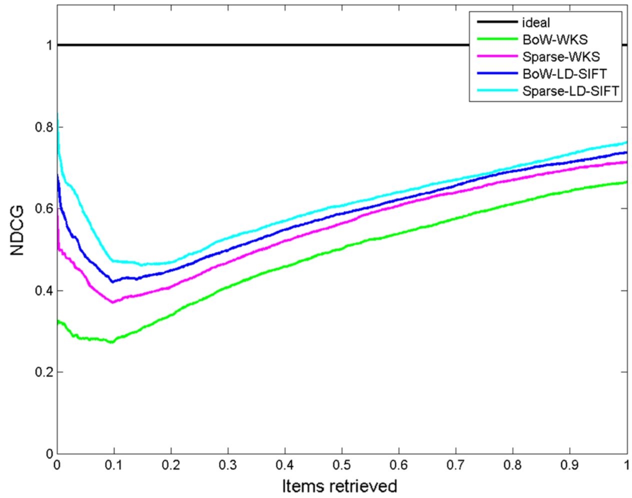


Fig. 9 NDCG *curves* on SHREC 2007 partial shape retrieval data. Comparison with different local-to-global approaches

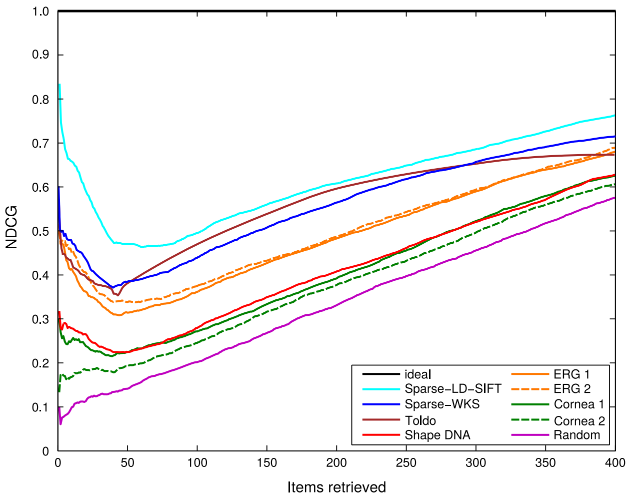


Fig. 10 NDCG *curves* on SHREC 2007 partial shape retrieval data. Comparison with other methods

the original SHREC 2007 contest [37], namely Cornea 1 and 2 [25], and ERG 1 and 2 [25]. Finally, we show the performance of the ShapeDNA method in order to evaluate a fully global method. In [36], a more sophisticated bag of words techniques is introduced considering a hierarchical approach. Moreover, the bag of words were computed from region descriptors after a shape segmentation procedure. In Cornea 1 and 2 [25] a skeleton-based approach is introduced, while in ERG 1 and 2 [25] the methods are based on the Extended Reeb Graph in order to decompose the whole shape in topologically connected subparts.

Figure 10 shows the performance of the experiments. The Sparse-LD-SIFT method clearly outperformed other methods. Sparse-WKS is comparable with Toldo's method. In general, global methods like ShapeDNA are not suitable for this context. Also perfomance obtained with skeleton-based methods are not very convincing. Methods based on Reeb Graph performed better but still are not comparable with more recently proposed approaches.

7 Conclusions

In this paper a new approach for local-to-global shape description is proposed. We have shown that sparse coding methods are particular suitable to compactly describe a large set of point-based descriptors. In particular we compared our method with the bag of words approach which represents the state of the art for extending local descriptors to the whole shape. We have evaluated our approach on 3D shape retrieval on some standard datasets in order to evaluate the proposed approach on different challenging scenarios. We have shown that our sparse coding method is quite robust to strong shape deformations due to noise, holes and so on. Moreover our method was successfully able to deal with partial and composite objects aiming at obtaining the benefit from both local and global approaches.

In particular, we evaluated our local-to-global approach starting from two different local shape descriptors. In both the cases our sparse based method outperformed other quantization methods. We expect that, given a new local descriptor that improves the local encoding, we can employ the same sparse coding approach and therefore improves the local-to-global description.

Future work will address the evaluation of more advanced sparse coding methods to further improves the local-to-global encoding. In particular our aim is to exploit discriminative learning in the training of the dictionaries in order to introduce further constrains that can improve the shape retrieval performance. Moreover, other constraints can be introduced into the sparse coding problem to take into account of spatial relationships of local parts.

Acknowledgments We would like to thank Alex and Michael Bronstein for useful suggestions and fruitful discussions.

References

1. Aubry, M., Schlickewei, U., Cremens, D.: The wave kernel signature: a quantum mechanical approach to shape analysis. In: Proc. of ICCV Workshop Dyn. Shape Capture Anal. (4DMOD) (2011)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. **24**(24), 509–522 (2002)
3. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
4. Boscaini, D., Castellani, U.: Local signature quantization by sparse coding. In: Eurographics Workshop on 3D Object Retr. (2013)
5. Boyer, E., Bronstein, A.M., Bronstein, M.M., Bustos, B., Darom, T., Horaud, R.: SHREC 2011: robust feature detection and description benchmark. Proc. of Eurographics Workshop 3D Object Retr. (3DOR) (2011)
6. Bronstein, A.M., Bronstein, M.M., Guibas, L.J., Ovsjanikov, M.: Shape google: geometric words and expressions for invariant shape retrieval. ACM Trans. Graph. (TOG) **30**(1), 1–20 (2011)
7. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Numerical Geometry of Non-rigid Shapes, Monographs in Computer Science. Springer, New York (2008)
8. Bronstein, M.M., Kokkinos, I.: Scale-invariant heat kernel signature for non-rigid shape recognition. In: Proc. Comput. Vis. Pattern Recognit. (CVPR), pp. 1704–1711 (2010)
9. Castellani, U., Bartoli, A.: 3D shape registration. 3D Imaging, Analysis, and Applications. Springer, Berlin (2012)
10. Castellani, U., Cristani, M., Murino, V.: Statistical 3D shape analysis by local generative descriptors. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **33**, 2555–2560 (2011)
11. Castellani, U., Mirtuono, P., Murino, V., Bellani, M., Rambaldelli, G., Tansella, M., Brambilla, P.: A new shape diffusion descriptor for brain classification. . In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Computer Science, vol 6892. Springer, Berlin, pp 426–433 (2011)
12. Darom, T., Keller, Y.: Scale-invariant features for 3-D mesh models. IEEE Trans. Image Process. **21**(5), 2758–2769 (2012)
13. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. Trans. Pattern Anal. Mach. Intell. **25**(10), 1285–1295 (2003)
14. Funkhouser, T., Kazhdan, M., Min, P., Shilane, P.: Shape-based retrieval and analysis of 3D models. Commun. ACM **48**, 58–64 (2005)
15. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3-D scenes. IEEE Trans. Pattern Anal. Mach. Intell. **21**(5), 433–449 (1999)
16. Lavoue, G.: Combination of bag-of-words descriptors for robust partial shape retrieval. Vis. Comput. **26**, 1257–1268 (2012)
17. Lévy, B.: Laplace–Beltrami eigenfunctions: towards an algorithm that “understands” geometry. In: IEEE Int. Conf. on Shape Model. Appl. (2006)
18. Lian, Z., Godil, A., Bustos, B., Daoudi, M., et al.: SHREC 2011 track: shape retrieval on non-rigid 3D watertight meshes. In: Proceedings of the Eurographics Workshop on 3D Object Retrieval, pp. 79–88 (2011)
19. Lian, Z., Godil, A., Bustos, B., et al.: A comparison of methods for non-rigid 3D shape retrieval. Pattern Recognit. **46**(1), 449–461 (2013)
20. Lian, Z., Godil, A., Fabry, T., T., F., et al.: SHREC 2010: Non-rigid 3D shape retrieval. In: Proc. Eurographics Workshop 3D Object Retr. (3DOR) (2010)
21. Litman, R., Bronstein, A.M.: Learning spectral descriptors for deformable shape correspondence. IEEE Trans. Pattern Anal. Mach. Intell. **36**(1), 171–180 (2014)
22. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
23. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Proc. Int. Conf. Mach. Learn. (ICML), pp. 689–696 (2009)
24. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. **11**, 19–60 (2010)
25. Marini, S., Paraboschi, L., Biasotti, S.: Shape retrieval contest 2007 (SHREC07): Partial matching track. technical report 10/07, IMATI (2007)
26. Mitra, N.J., Guibas, L., Giesen, J., Pauly, M.: Probabilistic fingerprints for shapes. In: Symposium on Geometry Processing, pp. 121–130 (2006)
27. Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., Guibas, L.: Functional Maps: A flexible representation of maps between shapes. ACM Trans. Graph. **31**(4), 30:1–30:11 (2012)
28. Pokrass, J., Bronstein, A.M., Bronstein, M.M., Sprechmann, P., Sapiro, G.: Sparse modeling of intrinsic correspondences. Comput. Graph. Forum **32**(2), 459G–468 (2013)
29. Reuter, M., Wolter, F.E., Peinecke, N.: Laplace–Beltrami spectra as ‘shape-DNA’ of surfaces and solids. Comput.-Aided Des. **38**, 342–366 (2006)
30. Rustamov, R.M.: Laplace–Beltrami eigenfunctions for deformation invariant shape representation. In: Eurographics Symp. Geom. Process., pp. 225–233 (2007)
31. Schölkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. The MIT Press, Cambridge (2002)
32. Sumner, R.W., Popović, J.: Deformation transfer for triangle meshes. ACM Trans. Graph. (TOG) **23**, 399–405 (2004)
33. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. In: Proc. Symp. Geom. Process., pp. 1383–1392 (2009)

34. Tangelder, J.W., Veltkamp, R.C.: A survey of content based 3D shape retrieval methods. In: Int. Conf. Shape Modell. Appl., pp. 145–156 (2004)
35. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. pp. 267–288 (1996)
36. Toldo, R., Castellani, U., Fusiello, A.: The bag of words approach for retrieval and categorization of 3D objects. Vis. Comput. **26**, 1257–1268 (2010)
37. Veltkamp, R.C., Haar, F.B.: Shrec 2007: 3D shape retrieval contest. Tech. Rep. UU-CS-2007-015, Department of Information and Computing Sciences, Utrecht University (2007)
38. Wuhrer, S., Azouz, Z.B., Shu, C.: Posture invariant surface description and feature extraction. In: IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 374–381 (2010)

Davide Boscaini is Ph.D. student at the University of Lugano, Switzerland. He received both his BSc and MSc in Applied Mathematics at the University of Verona, Italy. During his last year of master, he got interested in Spectral and iffusion Geometry topics joining a stage in the Vision, Image Processing and Sound (VIPS) Lab of the Department of Computer Science in Verona. His current research interests are physically-based methods for dimensionality reduction, spectral approaches to multimodal similarity and functional methods for shape matching.

Umberto Castellani is Ricercatore (i.e., Research Assistant) of the Department of Computer Science at the University of Verona. He received his Dottorato di Ricerca (Ph.D.) in Computer Science from the University of Verona in 2003 working on 3D data modelling and reconstruction. During his Ph.D., he had been a Visiting Research Fellow at the Machine Vision Unit of the Edinburgh University, in 2001. In 2007, he was an Invited Professor at the LASMEA Laboratory in Clermont-Ferrand, France. In 2008, he was a Visiting Researcher at the PRIP Laboratory at Michigan State University (USA). His research is focused on 3D data processing, statistical learning, and medical image analysis. He has coauthored several papers which were published in leading conference proceedings and journals. He is a member of Eurographics and IEEE.