

Feature Encoding of Spectral Signatures for 3D Non-Rigid Shape Retrieval

Frederico A. Limberger
<http://www.cs.york.ac.uk/~fal504>

University of York
York, UK

Richard C. Wilson
<http://www.cs.york.ac.uk/~wilson>

Abstract

As the Internet and 3D modelling tools have led to an increasingly growth in the number of available 3D models, it becomes necessary to have a proper and smaller representation for searching purposes that captures the most important information about shapes. A large number of encoding methods have been proposed in the literature to create shape signatures from local descriptors. Two encoding methods have been receiving most attention from researchers given its informative characteristics: Fisher Vector [9] and Super Vector [6]. We propose to use these encoding methods combined with spectral signatures to represent 3D shapes. Although spectral signatures have many desirable properties to describe 3D shapes, for instance being invariant under rigid transformations and stable against non-rigid transformations, they do not perform so well in recent benchmarks. We propose improvements to the Wave Kernel Signature by analysing its behaviour when combined to different encoding methods for the purpose of shape retrieval and classification. At the end, we show a comparison of our method in two recent benchmarks.

1 Introduction

A large number of local signatures have been created to represent local characteristics of geometric shapes for the purpose of many computer vision, geometry processing and shape analysis tasks. A local signature is a compact representation that characterizes a small region of a shape. They usually capture information about the neighbourhood of a vertex and so they can be directly applied to some important tasks like point correspondence and shape segmentation. For this purpose, it is desirable to compute signatures that are invariant under rigid, non-rigid and isometric deformations, the typical deformations that 3D models undergo. However, local descriptors cannot be immediately applied to the problem of shape retrieval, because this task is not addressed by comparing local signatures but by comparing global descriptors (signatures that represent the shape as a whole). Creating a global descriptor is not a simple task since shapes can have arbitrary number of vertices, edges and faces. To create a generic representation of a shape all important characteristics should be preserved during the encoding process. Even so, the global representation must compress local characteristics using the same basis for all shapes in order to facilitate comparisons.

With the popularization of the Bag-of-Features paradigm (BoF), Ovsjanikov *et al.* [18] and later Bronstein *et al.* [9] proposed to encode local descriptors to a global representation

of the shape, removing the local dependence of each descriptor. They use Bag-of-Features to obtain a visual word-based representation of the shape that can be compared efficiently. They formulate their signature as the normalized probability of associating each geometric word from the vocabulary with all vertices of the shape. The distance between two shapes is given by a Hamming metric between the bag of features of each shape in the Hamming space.

Recently, image retrieval and classification tasks have been improved [6, 7] by the use of Fisher Vectors (FV) [8], which have many advantages over BoF [30]. Concisely, Fisher Vector consists of representing a sample x by the concatenation of weighted deviations with respect to the parameters of a model, describing the direction which the parameters should be modified to fit the data. This model can be seen as a “probabilistic visual vocabulary” [30], which can be represented by a Gaussian Mixture Model (GMM). The final FV representation is very discriminative and can be used to describe the initial sample with applications in retrieval and classification. FV has shown to be a more discriminative probabilistic descriptor than the classical BoF, since it includes second-order statistics in its formulation [9].

1.1 Main Contribution

In this paper, we propose an efficient and discriminative encoding framework to address the problem of creating global signatures for 3D models from local descriptors based on the spectrum of the shape, for the purpose of shape retrieval and classification. In this way, we propose the use of Fisher Vector to describe the entire representation of a shape. Differently from [9, 13], our approach uses a Gaussian Mixture Model as a dictionary of probabilistic visual words, and encodes the global signature using three orders statistics (0-th, 1-st, 2-nd) rather than using only the first order. Further, while BoF generates a K -dimensional histogram, where K is the vocabulary size, Fisher Vector encoding generates a high-dimensional vector with $2KD$ dimensions, where D is the size of each local descriptor, being more discriminative but still simple to compare, as all shapes are encoded in the same basis.

We compare the use of three different spectral descriptors (the Heat Kernel Signature (HKS) [11], the Scale-invariant Heat Kernel Signature (SI-HKS) [12] and the Wave Kernel Signature (WKS) [10]) in recent benchmarks by encoding them using the Fisher Vector and Super Vector paradigms. From that, we propose some improvements to the WKS descriptor, since it overtakes HKS and SI-HKS in retrieval performance. A better scaling is proposed for the eigenvalues of the Laplace-Beltrami operator which captures more information about the shape and we also propose the use of principal curvatures to increase the efficiency of the encoding method.

The rest of this paper is laid out as follows. Sect. 2 discusses works related to spectral signatures, shape retrieval and encoding methods. Sect. 3 explains how to represent shapes using encoding processes. In Sect. 4 we evaluate our method in recent benchmarks showing the relevance of our improved method. Finally, Sect. 5 concludes the paper.

2 Related Works

Shape retrieval is a well-established research area that has many approaches and methods. We are interested in the family of spectral methods, which uses the eigenvalues and eigenvectors of the Laplace-Beltrami operator (LBO) defined on the shape, since they have many desirable properties to tackle 3D shapes as they are invariant under isometric deformations and stable against pose changes [11]. Moreover, the spectrum contains a considerable amount

of geometric and topological information about the shape. For a more detailed explanation about other methods of characterizing shapes we refer the reader to [8].

2.1 Spectral Methods

Creating signatures using the spectra of the shape was first addressed by Reuter *et al.* [23, 24]. In this seminal paper, the authors use a collection of eigenvalues of the LBO of a shape to represent local properties. Although Reuter’s signature ensures that it can recognize isometric shapes, there exist compact non-isometric shapes that have the same spectra therefore they cannot be fully distinguished by the eigenvalues of the LBO, and still it is not robust to noise and partial model matching. Rustamov [26] uses all the spectra (eigenvalues and eigenvectors) of a shape to create the Global Point Signature (GPS). He proposes an isometry-invariant shape embedding, where global signatures are computed by a histogram of pairwise distances [17] between points in the embedding space. This approach solves the previous problem of distinguishing non-isometric shapes with same spectra but introduces the sign correction problem.

Sun *et al.* [61] introduced a deformation-invariant signature based on the heat kernel, resolving the sign correction problem of the GPS. Named the Heat Kernel Signature (HKS), it is based on the behaviour of the heat diffusion over the surface of the shape governed by the heat equation. Although the HKS can describe a shape at multiple scales (at different times) it is not scale-invariant. Later, Bronstein and Kokkinos proposed a normalization to the HKS making it scale-invariant (SI-HKS) by translating and scaling the signature [9].

Recently, Aubry *et al.* [0] introduced the Wave Kernel Signature (WKS), which evaluates the probability of measuring a quantum mechanical particle at a specific location $x \in X$ by varying its energy e . Aubry *et al.* use the Schrodinger equation governed by the wave function $\Psi(x, t)$ to describe the quantum mechanical behaviour of particles over a object surface. Furthermore, improvements in the WKS descriptors by directly operating on their weights were previously tackled by [25, 35].

2.2 Shape Retrieval and Encoding Methods

The idea of creating shape signatures from local signatures is neither new nor straightforward. The usual way to describe a set of local descriptors into a shape-level signature for retrieval and classification purposes is building a Bag-of-Features model to remove the local dependency of each descriptor by writing local properties as a histogram of their occurrences. Although recent approaches mainly use the classical BoF, the use of other encoding methods can bring many advantages over the traditional method. For instance, the Fisher Vector combines the strengths of generative and discriminative models [19]. While the BoF characterizes a sample by the number of occurrences of visual words, FV is characterized by the deviation from a probabilistic vocabulary. Chatfield *et al.* [5] states that Fisher Vector and Super Vector encodings are better than other encoding methods, since they carry extra information about the displacement between descriptors and visual words. In the next paragraphs we review the principal methods that use encoding methods related to BoF paradigm.

Recently, Ovsjanikov *et al.* [33] and Bronstein *et al.* [3] have used BoF to combine spectral signatures (HKS and SI-HKS, respectively) to represent shapes, removing the local dependence of each local descriptor by describing the probability of occurrence of each visual word (local descriptor) in a geometric vocabulary. Later, Litman *et al.* [16] developed

a supervised learning approach to construct the dictionary of BoF model, showing significant improvements in performance over the compared methods.

There are other signatures that also make use of the BoF framework but do not use spectral local descriptors. In [10], Furuya *et al.* use BoF to encode SIFTs generated by depth images rendered in different views from a 3D model. In a recent benchmark of 3D shape retrieval [11], Furuya and Ohbuchi [8] has also applied BoF to their local features. In a benchmark of retrieval of non-rigid 3D human models [12], Tatsuma (Bag-of-Features approach with Augmented Point Feature Histograms), Bu (High-level Feature Learning for 3D Shapes) and Li [13], besides Litman *et al.* [16], have encoded local features using BoF. In other recent benchmarks [14, 15], the best results were obtained by techniques that have used the BoF or similar frameworks.

The use of Fisher Vectors to classify and retrieve images has been recently addressed by a number of researchers. Perronnin and Dance [19] applies FV to the problem of image categorization, Perronnin *et al.* [20] proposed a compressed form of FV to retrieve images in a large database, Sanchez *et al.* [50] and Csurka and Perronnin [6] show that FV framework is the state-of-the-art approach for classification and retrieval purposes since it has a more efficient representation of an image. Takeyoshi and Kikinis [53] use FV to classify patients with epilepsy, Schneider and Tuytelaars [27] applied FV in sketch classification and Simonyan *et al.* [49] create a face descriptor achieving state-of-the-art performance on a challenge benchmark. The use of Gaussian Mixture Models (basic concept of FV) to characterize 3D shapes was first experimentally addressed by [2].

On the other hand, Super Vector (SV) [56] has shown to be a good encoding to represent local features. In a recent benchmark [15], SV was used to aggregate local features achieving the best performance among other participants. In [13], Tatsuma used Super Vector to encode features extracted from rendered depth buffer images also performing the best on that track.

3 Encoding Spectral Signatures

Similarly to images, encoding methods can be applied to shapes. Although shapes have a complex structure in a 3-dimensional space, encodings can be applied effectively after calculating proper local descriptors, which must respect the following properties:

1. Isometry-invariant signature: It is essential that a shape can be described by isometry-invariant descriptors to avoid prior alignment (place the model in the same orientation).
2. Multi-level signature: It is necessary to describe a shape in multiple scales. For this, all descriptors must have the same size though the number of descriptors still depends on the shape (number of vertices).

The spectral signatures fulfil all these requirements. Each scale of the descriptors is seen as a layer that describes the entire shape. Encoding methods are applied to these layers to encode all the information in a high dimensional vector. Besides, there are other good properties that descriptors should hold to properly represent shapes, for instance being stable against non-rigid motions and stable against topology changes. Although these are very important, they are not essential for the encoding process.

3.1 Encoding Local Descriptors

In order to encode local descriptors we need to characterize them by their deviation from a generative model. Let $X = \{\mathbf{x}_t, \mathbf{x}_t \in \mathbb{R}^D, t = 1 \dots T\}$ be a set of local descriptors of a shape S ,

where T is the number of vertices and D the descriptor dimension, and $\lambda = \{w_k, \mu_k, \Sigma_k, k = 1 \dots K\}$ a set of parameters of a Gaussian Mixture Model p_λ (Eq. (1)), where w_k , μ_k and Σ_k are respectively the weight, mean vector and covariance vector of the k -th Gaussian of the GMM. We assume that covariances matrices are diagonal thus writing them as vectors. The distribution of descriptors $p_\lambda(x)$ is given by:

$$p_\lambda(x) = \sum_{k=1}^K w_k \mathcal{N}(x | \mu_k, \Sigma_k) : \sum_{k=1}^K w_k = 1 \quad (1)$$

The Gaussian Mixture Model parameters are estimated using the Expectation Minimization (EM) algorithm [50] in order to optimize a Maximum Likelihood criterion for the data X . Differently from K -means algorithm that assigns each sample to a cluster, EM-algorithm gives the probability of each sample belonging to each cluster.

3.1.1 Fisher Vector

The Fisher Vector encoding characterizes a large set of vectors by their three-order deviation from a vocabulary, generating a high-dimensional gradient vector representation. The gradient of the log-likelihood, also called *Fisher score*, describes the contribution of each parameter to the generation process and is given by [9]:

$$G_\lambda^X = \nabla_\lambda \log p_\lambda(X) \quad (2)$$

To compute the FV encoding, we write the local shape descriptors wrt. the probabilistic model, which means expressing X by its gradient in respect to p_λ . This is done by associating each vector \mathbf{x}_t to a mode k in the GMM. First, we compute the association strength (soft assignment) that is given by the posterior probability [9, 20]:

$$q_{tk} = \frac{\exp[-\frac{1}{2}(\mathbf{x}_t - \mu_k)^\top \Sigma_k^{-1}(\mathbf{x}_t - \mu_k)]}{\sum_{i=1}^K \exp[-\frac{1}{2}(\mathbf{x}_t - \mu_i)^\top \Sigma_i^{-1}(\mathbf{x}_t - \mu_i)]}. \quad (3)$$

Second, for each mode k and each descriptor dimension $j = 1..D$, we can compute the deviation vectors (gradient) with respect to the mean and covariance, respectively

$$u_{jk} = \frac{1}{T \sqrt{w_k}} \sum_{i=1}^T q_{ik} \frac{x_{ji} - \mu_{jk}}{\sigma_{jk}}, \quad (4)$$

$$v_{jk} = \frac{1}{T \sqrt{2w_k}} \sum_{i=1}^T q_{ik} \left[\left(\frac{x_{ji} - \mu_{jk}}{\sigma_{jk}} \right)^2 - 1 \right]. \quad (5)$$

where σ_{jk} are the square roots of the covariances Σ_k . Finally, the FV representation of a shape S is the concatenation of the vectorization of the matrices u_{jk} and v_{jk}

$$\Gamma_{FV} = [\dots \mathbf{u}_k^\top \dots, \dots \mathbf{v}_k^\top \dots]^\top \quad (6)$$

To properly compare shape signatures (Γ_{FV}) we have also applied L2 Normalization and Power Normalization to the Fisher Vector, also known as the Improved FV [20].

3.1.2 Super Vector

Super Vector encoding is similar to the Fisher Vector encoding. In their framework Zhou *et al.* [6] perform a nonlinear map to create a high-dimensional sparse vector. Differently from FV, SV only considers the zero and first order differences between descriptors and Gaussian means. Instead of considering the second order differences, they add a component related to the mass of each cluster. Thus, the magnitude of their signature is $K(D + 1)$. The Super Vector encoding can be calculated by the following expressions [6]:

$$p_k = \frac{1}{N} \sum_{i=1}^N q_{ik} \quad s_k = s\sqrt{p_k} \quad \mathbf{u}_k = \frac{1}{\sqrt{p_k}} \sum_{i=1}^N q_{ik}(\mathbf{x}_i - \mu_k) \quad (7)$$

where s is a constant that balances s_k and \mathbf{u}_k numerically. As can be seen in Eq. (7), the SV encoding normalizes elements by the square root of the posterior probability ($\sqrt{p_k}$) instead of the prior probability ($\sqrt{w_k}$). The final descriptor is given by the following concatenation:

$$\Gamma_{SV} = [s_1, \mathbf{u}_1^\top, \dots, s_K, \mathbf{u}_K^\top]^\top \quad (8)$$

Using either the Fisher Vector or the Super Vector as shape descriptor, we define the distance between two shapes R and S as the L_1 distance between their encodings (Γ):

$$d = \sum \|\Gamma(R) - \Gamma(S)\|_1 \quad (9)$$

3.2 Improving WKS for Shape Retrieval

There are requirements so that a signature can perform well in shape retrieval benchmarks. These requirements concern the different shape transformations that a signature should be invariant to, for example, rigid-motions, noise, holes, etc. Therefore, we propose two improvements to the Wave Kernel Signature (informative scaling and curvature aggregation) so it can be more discriminative over the encoding process. We chose to improve the WKS since it achieves better performances for shape retrieval among other spectral signatures. Besides those two improvements, we consider the case of the shape being disconnected, which means that it may have more than one connected component. Following the Laplacian matrix property, which states that the eigenvalues will have as many zeros as the number of connected components in the graph, we remove all zeros from the eigenvalues rather than only one.

3.2.1 Informative Scaling

The Wave Kernel Signature uses the logarithmic scaled versions of the Laplace-Beltrami eigenvalues to compute its signature based on the fact that the variation of eigenvalues for articulated shapes are log-normally distributed. The WKS is computed by Eq. (10), where f_E is a distribution that properly characterizes shape properties at different scales and C_e normalizes the sum.

$$\text{WKS}(x, e) = C_e \sum_{k=1}^{\infty} \phi_k(x)^2 f_E(\Lambda_k)^2 \quad f_E(\Lambda_k)^2 = e^{-\frac{(e - \log(\Lambda_k))^2}{2\sigma^2}} \quad (10)$$

We have analysed the influence of this scaling to the computation of the signature and we concluded that it loses high frequency information about the spectrum of the shape (even

important high frequencies that do not originate from noise are blurred). Then, we analysed the distributions of the differences between eigenvalues of 24 articulated woman shapes, as can be seen in Image 1. By analysing the differences of the eigenvalues (left), we propose to use the cubic root scaling rather than the logarithmic scale since it fits much better a normal distribution, which is used to handle the differences between same-class shapes.

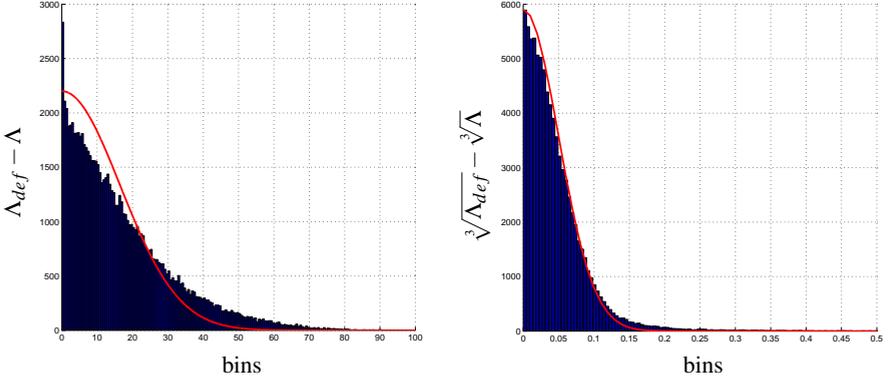


Figure 1: Histograms of the differences between the eigenvalues of the Laplace-Beltrami operator of 24 articulated woman shapes. (left) Histogram of the differences not scaled. (right) Histogram of the differences scaled by the cubic root. On each graph, the red line is a reference normal distribution fitted to the respective histogram.

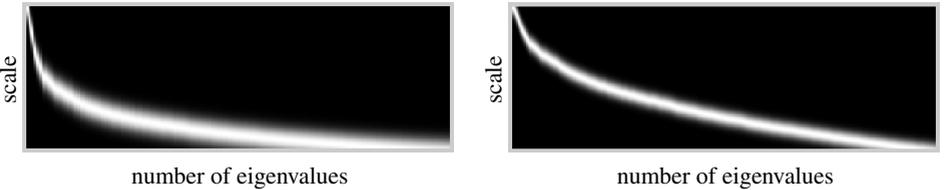


Figure 2: Weightings of the eigenfunctions of the LBO for a shape using the logarithmic scale (left) ($f_E(\Lambda_k)^2$) and the power scale (right) ($f_C(\Lambda_k)^2$). We use $\sigma = 3.75 * (\lambda_1 - \lambda_2)$, where λ_1 and λ_2 are the respective scaled versions of the first and second eigenvalues.

The cubic root scaling also keeps more information at every energy scale than the logarithmic scale. Figure 2 shows the behaviour of the weightings when using logarithmic scale (left) (Eq. (10)) and our power scale (right) (Eq. (11)). We use power scale equal to $\frac{1}{3}$ (cubic root) in Figure 2 and in all our experiments. Note that the weightings (100×300) will be multiplied by the squared eigenfunctions of the LBO ($300 \times V$) to output the signature ($100 \times V$), where V is the number of vertices (using the first 300 eigenvalues and eigenfunctions). This way, every row of the weighting corresponds to the weighting of every energy scale and therefore it is possible to note that the last frequencies will be blurred by the WKS weighting since they are very similar. When using power scale, the weighting is more robust, capturing information about all the shape spectrum and not blurring any scale.

$$f_C(\Lambda_k)^2 = e^{\frac{-(e - \sqrt[3]{\Lambda_k})^2}{2\sigma^2}} \quad e \in [\sqrt[3]{\lambda_{min}}, \sqrt[3]{\lambda_{max}}] \quad (11)$$

3.2.2 Curvature Aggregation

Encoding methods are based on the differences between descriptor histograms and a probabilistic model. The more discriminative these histograms are the more the encoding process will be able to distinguish between shapes of different classes. We have noticed that when shapes undergo pose changes, the maximum principal curvature of the vertices remains stable. This is a local feature of the surface, which is isometry invariant and stable at most points under object articulation but not well coded in the WKS. Therefore, as another improvement to the shape retrieval task, we propose the use of principal curvatures to increase the separation of features leading to a more discriminative histogram. Practically, we shift the WKS individually by the maximum principal curvature c for each vertex x as stated in Equation (12). We smooth the maximum principal curvature by taking the mean of the respective neighbour vertices to diminish the influence of noise. In (12), α is a weight that normalizes c accordingly to the signature values. We use $\alpha = 0.015$ in all our experiments.

$$\text{IWKS}(x, e) = C_e \sum_{k=1}^{\infty} \phi_k(x)^2 f_C(\Lambda_k)^2 + c_x \alpha \quad (12)$$

4 Evaluation

In the following we present experimental results of our approach applied to the problem of shape retrieval. We compare two encoding methods: Fisher Vector and Super Vector when combined with HKS, SI-HKS, WKS, and IWKS (our Improved WKS). We first compare their performance on the most recent SHREC’15 benchmark [15]. In SHREC databases, the output of a shape retrieval task is a dissimilarity matrix $N \times N$, where N is the number of models and the entry (i, j) is the difference between models i and j . Using a classification file, which tells the class of each shape, we compute different standard retrieval measures: Nearest neighbour (NN), First-tier (FT), Second-tier (ST), e-Measure (E), Discounted Cumulative Gain (DCG). For a more detailed explanation about each measure we refer to reader to [15]. Table 1 compares these different retrieval measures. In bold are shown the best retrieval performances for each measure. As shown, FV-IWKS performs better than others. In Figure 3 is shown the Precision and Recall curves of the evaluated methods in the SHREC’15.

Method	NN	FT	ST	E	DCG
FV-HKS	0.9567	0.7489	0.8292	0.6661	0.9134
FV-SIHKS	0.9658	0.8104	0.8770	0.7102	0.9382
FV-WKS	0.9725	0.8628	0.9183	0.7511	0.9553
FV-IWKS	0.9975	0.9463	0.9801	0.8102	0.9884
SV-HKS	0.9217	0.6168	0.7061	0.5564	0.8539
SV-SIHKS	0.9642	0.7559	0.8371	0.6698	0.9222
SV-WKS	0.9600	0.7685	0.8520	0.6842	0.9243
SV-IWKS	0.9867	0.8748	0.9387	0.7649	0.9683

Table 1: Retrieval performance comparison of the different spectral signatures combined with FV and SV applied to the SHREC’15 benchmark.

We also compare our best method (FV-WKS) against the competitors of SHREC’11 [14] and SHREC’15 benchmarks [15]. We show a comparison using the same retrieval measures: NN, FT, ST, E and DCG. The SHREC’11 dataset contains 600 watertight 3D shapes, divided into 30 classes with 20 models per class. Watertight means that meshes do not have holes,

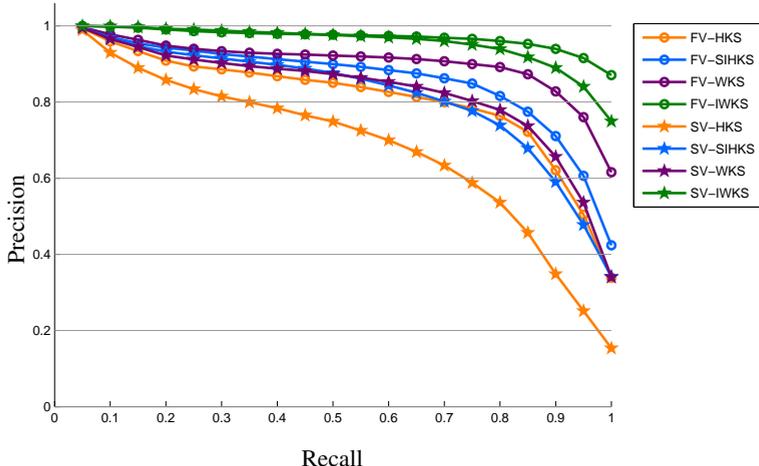


Figure 3: Precision and Recall plot of different spectral signatures (HKS, SI-HKS, WKS and WKS) tested with FV and SV encoding methods applied to the SHREC'15 benchmark. Colors represent different local descriptors. As shown, both the two best performances in this benchmark use our proposed descriptor (IWKS).

cracks or missing features. In Table 2 we compare our retrieval performance in this dataset against the best run of each group using five standard measures. The SHREC'15 dataset contains 1200 watertight 3D shapes, divided into 50 classes with 24 models per class. The models were selected from other public repositories and then were generated 23 deformed versions for each model. Table 3 summarizes the retrieval performances of the top groups from the contest in comparison to our method.

Method	NN	FT	ST	E	DCG
Our method (FV-IWKS)	0.9983	0.9591	0.9860	0.7318	0.9937
SD-GDM-meshSIFT	1.0000	0.9720	0.9901	0.7358	0.9955
MDS-CM-BOF	0.9950	0.9127	0.9691	0.7166	0.9822
OrigM-n12-normA	0.9917	0.9153	0.9569	0.7047	0.9783
FOG+MRR	0.9600	0.8810	0.9461	0.6958	0.9586
BOGH	0.9933	0.8111	0.8839	0.6469	0.9493
LSF	0.9950	0.7988	0.8631	0.6327	0.9432

Table 2: Retrieval performance comparison of the best runs of the six groups that performed better in SHREC'11 [14] against ours using five standard measures. In bold are highlighted the best performances for each retrieval measure.

Settings. We use the same dictionary settings to compute the FV and SV encodings for every spectral signature. We compute a probabilistic dictionary using the first 29 signature models for each database and estimate a GMM with 38 Gaussians for each signature scale. To calculate the signatures we compute the first 300 eigenvalues of the LBO for each model. We evaluate the HKS and SI-HKS at the time interval $[4\ln(10)/\lambda_{300}, 4\ln(10)/\lambda_2]$ logarithmic scaled, the WKS at the energy interval $[\log(\lambda_2), \log(\lambda_{300})]$, and the IWKS at every interval $[\sqrt[3]{\lambda_f}, \sqrt[3]{\lambda_{300}}]$, where λ_i corresponds to the i th eigenvalue of the LBO and λ_f the first nonzero

Method	NN	FT	ST	E	DCG
Our method (FV-IWKS)	0.9975	0.9463	0.9801	0.8102	0.9884
SV-LSF_kpaca50	1.0000	0.9972	0.9997	0.8357	0.9997
HAPT_run1	0.9983	0.9657	0.9821	0.8150	0.9919
SPH_SparseCoding_1024	0.9975	0.9568	0.9696	0.8047	0.9885
CompactBoHHKS10D	0.9842	0.8714	0.9082	0.7465	0.9582
EDBCF_NW	0.9775	0.7931	0.8839	0.7076	0.9431
SG_L1	0.9725	0.7596	0.8143	0.6597	0.9192

Table 3: Retrieval performance comparison of the best runs of the six groups that performed better in SHREC’15 [15] against ours using five standard measures. In bold are highlighted the best performances for each retrieval measure.

eigenvalue. We sample 100 points in the interval of the HKS, WKS and IWKS. In the SI-HKS, we use the first 6 lowest frequencies after scale normalization. After computing the encodings, we reduce the data dimensionality to 50 by performing Gaussian KPCA [52].

By analyzing our performance in the SHREC’11 track it is clear that our approach improves spectral descriptors to tackle shape retrieval problems since it achieves excellent retrieval measures (DCG very close to 1) and comparable results with best retrieval methods (very similar e-Measure to SD_GDM-meshSIFT).

Although our method is very close to the best performing groups of SHREC’15, it is not very clear why SV-LSF_kpaca50 performs relatively better than ours (Table 3) and at the time of writing, full details of this algorithm have not been published. Nevertheless, our method presents a much better performance when compared to other spectral descriptors, showing the potential of these methods.

5 Conclusion

We have presented a detailed comparison of different spectral methods combined with informative unsupervised Bag-of-Features models: Fisher Vector and Super Vector. We proposed some improvements that can be used with any spectral signature and a new scaling to be used with the Wave Kernel Signature. Although our method does not beat all other groups in the SHREC’11 and SHREC’15 benchmarks it is close to the best performances and achieves excellent results in most benchmark classes. The results of this paper have shown that spectral methods are a good choice to retrieve shapes when combined with informative encoding methods but it still suggests further research.

We observed that the worst retrieval performances happen when shapes undergo huge topology changes. In these cases, spectral signatures still need to be improved. In future works, we plan to create a spectral signature that is less variant to major topology changes.

Acknowledgements

This work was sponsored by CAPES - Brazil (Grant No.: 11892-13-7).

References

- [1] Mathieu Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *Computer Vision Workshops (ICCV*

- Workshops*), 2011 IEEE International Conference on, pages 1626–1633, Nov 2011.
- [2] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. Pose-consistent 3d shape segmentation based on a quantum mechanical feature descriptor. In *Proceedings of the 33rd International Conference on Pattern Recognition, DAGM'11*, pages 122–131, Berlin, Heidelberg, 2011. Springer-Verlag.
- [3] Alexander M. Bronstein, Michael M. Bronstein, Leonidas J. Guibas, and Maks Ovsjanikov. Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Trans. Graph.*, 30(1):1:1–1:20, February 2011.
- [4] M.M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Proc. CVPR*, pages 1704–1711, June 2010.
- [5] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman. The devil is in the details: an evaluation of recent feature encoding methods. In *BMVC*, 2011.
- [6] Gabriela Csurka and Florent Perronnin. Fisher vectors: Beyond bag-of-visual-words image representations. In Paul Richard and Josãl' Braz, editors, *Computer Vision, Imaging and Computer Graphics. Theory and Applications*, volume 229 of *Communications in Computer and Information Science*, pages 28–42. 2011.
- [7] Takahiko Furuya and Ryutarou Ohbuchi. Dense sampling and fast encoding for 3d model retrieval using bag-of-visual features. In *Proceedings of CIVR '09*, pages 26:1–26:8, New York, NY, USA, 2009. ACM.
- [8] Takahiko Furuya and Ryutarou Ohbuchi. Ranking on cross-domain manifold for sketch-based 3d model retrieval. In *2013 International Conference on Cyberworlds, Yokohama, Japan, October, 2013*, pages 274–281, 2013.
- [9] Tommi S. Jaakkola and David Haussler. Exploiting generative models in discriminative classifiers. In *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II*, pages 487–493. MIT Press, 1998.
- [10] Bo Li, Afzal Godil, Masaki Aono, X. Bai, Takahiko Furuya, L. Li, Roberto Javier López-Sastre, Henry Johan, Ryutarou Ohbuchi, Carolina Redondo-Cabrera, Atsushi Tatsuma, Tomohiro Yanagimachi, and S. Zhang. Shrec'12 track: Generic 3d shape retrieval. In *Eurographics Workshop on 3D Object Retrieval 2012, Cagliari, Italy, May 13, 2012. Proceedings*, pages 119–126, 2012.
- [11] Bo Li, Afzal Godil, and Henry Johan. Hybrid shape descriptor and meta similarity generation for non-rigid and partial 3d model retrieval. *Multimedia Tools and Applications*, 72(2):1531–1560, 2014.
- [12] Bo Li, Y. Lu, C. Li, Afzal Godil, Tobias Schreck, Masaki Aono, Martin Burtscher, Hongbo Fu, Takahiko Furuya, Henry Johan, Jianzhuang Liu, Ryutarou Ohbuchi, Atsushi Tatsuma, and Changqing Zou. Extended large scale sketch-based 3d shape retrieval. In *Eurographics Workshop on 3D Object Retrieval, Strasbourg, France, 2014. Proceedings*, pages 121–130, 2014.

- [13] Bo Li, Y. Lu, C. Li, Afzal Godil, Tobias Schreck, Masaki Aono, Qiang Chen, Nihad Karim Chowdhury, Bin Fang, Takahiko Furuya, Henry Johan, Ryuichi Kosaka, Hitoshi Koyanagi, Ryutarou Ohbuchi, and Atsushi Tatsuma. Large scale comprehensive 3d shape retrieval. In *Eurographics Workshop on 3D Object Retrieval, Strasbourg, France, 2014. Proceedings*, pages 131–140, 2014.
- [14] Zhouhui Lian, Afzal Godil, Benjamin Bustos, Mohamed Daoudi, J. Hermans, Shun Kawamura, Y. Kurita, Guillaume Lavouer, H.V. Nguyen, Ryutarou Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, Martin Reuter, Ivan Sipiran, Dirk Smeets, Paul Suetens, Hedi Tabia, and Dirk Vandermeulen. SHREC’11 track: shape retrieval on non-rigid 3D watertight meshes. In *Proc. Eurographics 2011 Workshop on 3D Object Retrieval (3DOR’11)*, pages 79–88. Eurographics Association, 2011.
- [15] Zhouhui Lian, J. Zhang, S. Choi, H. ElNaghy, Jihad El-Sana, Takahiko Furuya, Andrea Giachetti, Riza Alp Guler, L. Lai, C. Li, H. Li, F. A. Limberger, Ralph R. Martin, R. U. Nakanishi, A. P. Neto, Luis Gustavo Nonato, Ryutarou Ohbuchi, Kirill Pevzner, David Pickup, Paul L. Rosin, A. Sharf, L. Sun, Xianfang Sun, Sibel Tari, Gozde B. Unal, and R. C. Wilson. Non-rigid 3d shape retrieval. In *Eurographics Workshop on 3D Object Retrieval, Zurich, Switzerland, May 2-3, 2015.*, pages 107–120, 2015.
- [16] Roe Litman, Alex Bronstein, Michael Bronstein, and Umberto Castellani. Supervised learning of bag-of-features shape descriptors using sparse coding. In *Symposium on Geometry Processing (SGP)*, 2014.
- [17] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Trans. Graph.*, 21(4):807–832, October 2002.
- [18] M. Ovsjanikov, A.M. Bronstein, M.M. Bronstein, and L.J. Guibas. Shape google: a computer vision approach to isometry invariant shape retrieval. In *ICCV Workshops*, pages 320–327, Sept 2009.
- [19] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Proc. CVPR*, pages 1–8, June 2007.
- [20] F. Perronnin, Yan Liu, J. Sanchez, and H. Poirier. Large-scale image retrieval with compressed fisher vectors. In *Proc. CVPR*, pages 3384–3391, June 2010.
- [21] Florent Perronnin, Jorge Sanchez, and Thomas Mensink. Improving the fisher kernel for large-scale image classification. In *Proc. ECCV*, pages 143–156, Berlin, 2010.
- [22] D. Pickup, X. Sun, P. L. Rosin, R. R. Martin, Z. Cheng, Z. Lian, M. Aono, A. Ben Hamza, A. Bronstein, M. Bronstein, S. Bu, U. Castellani, S. Cheng, V. Garro, A. Giachetti, A. Godil, J. Han, H. Johan, L. Lai, B. Li, C. Li, H. Li, R. Litman, X. Liu, Z. Liu, Y. Lu, A. Tatsuma, and J. Ye. SHREC’14 track: Shape retrieval of non-rigid 3d human models. In *Proceedings of the 7th Eurographics workshop on 3D Object Retrieval, EG 3DOR’14*. Eurographics Association, 2014.
- [23] Martin Reuter, Franz-Erich Wolter, and Niklas Peinecke. Laplace-spectra as fingerprints for shape matching. In *Proceedings of the 2005 ACM Symposium on Solid and Physical Modeling, SPM ’05*, pages 101–106, New York, NY, USA, 2005. ACM.

- [24] Martin Reuter, Franz-Erich Wolter, and Niklas Peinecke. Laplace-beltrami spectra as 'shape-dna' of surfaces and solids. *Comput. Aided Des.*, 38(4):342–366, April 2006.
- [25] E. Rodola, S. Rota Bulò, T. Windheuser, M. Vestner, and D. Cremers. Dense non-rigid shape correspondence using random forests. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 4177–4184, June 2014.
- [26] Raif M. Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, pages 225–233, Switzerland, 2007. Eurographics Association.
- [27] Rosália G. Schneider and Tinne Tuytelaars. Sketch classification and classification-driven analysis using fisher vectors. *ACM Trans. Graph.*, 33(6):174:1–174:9, 2014.
- [28] Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser. The princeton shape benchmark. In *Proceedings of the Shape Modeling International 2004, SMI '04*, pages 167–178, Washington, DC, USA, 2004. IEEE Computer Society.
- [29] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher Vector Faces in the Wild. In *British Machine Vision Conference*, 2013.
- [30] Jorge SÁnchez, Florent Perronnin, Thomas Mensink, and Jakob Verbeek. Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision*, 105(3):222–245, 2013.
- [31] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Proceedings of the Symposium on Geometry Processing*, pages 1383–1392, Switzerland, 2009. Eurographics Association.
- [32] Johan W. Tangelder and Remco C. Veltkamp. A survey of content based 3d shape retrieval methods. *Multimedia Tools Appl.*, 39(3):441–471, September 2008.
- [33] N. Vohra, B.C. Vemuri, A. Rangarajan, R.L. Gilmore, S.N. Roper, and C.M. Leonard. Kernel fisher for shape based classification in epilepsy. In Takeyoshi Dohi and Ron Kikinis, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002*, Lecture Notes in Computer Science, pages 436–443. 2002.
- [34] Quan Wang. Kernel principal component analysis and its applications in face recognition and active shape models. *CoRR*, 2012.
- [35] Thomas Windheuser, Matthias Vestner, Emanuele Rodola, Rudolph Triebel, and Daniel Cremers. Optimal intrinsic descriptors for non-rigid shape analysis. In *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014.
- [36] Xi Zhou, Kai Yu, Tong Zhang, and Thomas S. Huang. Image classification using super-vector coding of local image descriptors. In *Proc ECCV*, pages 141–154, 2010.