Spectral Geometric Methods for Deformable 3D Shape Retrieval

Chunyuan Li

A Thesis

in

The Concordia Institute

for

Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science (Quality Systems Engineering) at

Concordia University

Montréal, Québec, Canada

February 2013

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Chunyuan Li

Entitled: Spectral Geometric Methods for Deformable 3D Shape Retrieval

Master of Applied Science (Quality Systems Engineering)

Complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. A. Awasthi

_____ External Examinar
Dr. W. F. Xie

_____ Internal Examiner
Dr. S. Li

_____ Supervisor
Dr. A. Ben Hamza

Approved by: _____
Dr. M. Debbabi, Director
Concordia Institute for Information Systems Engineering

_____
Dr. Robin A. L. Drew, Dean,
Faculty of Engineering and Computer Science

Date: _____

ABSTRACT

Spectral Geometric Methods for Deformable 3D Shape Retrieval

Chunyuan Li

As 3D applications ranging from medical imaging to industrial design continue to grow, so does the importance of developing robust 3D shape retrieval systems. A key issue in developing an accurate shape retrieval algorithm is to design an efficient shape descriptor for which an index can be built, and similarity queries can be answered efficiently. While the overwhelming majority of prior work on 3D shape analysis has concentrated primarily on rigid shape retrieval, many real objects such as articulated motions of humans are nonrigid and hence can exhibit a variety of poses and deformations.

In this thesis, we present novel spectral geometric methods for analyzing and distinguishing between deformable 3D shapes. First, we comprehensively review recent shape descriptors based on the spectral decomposition of the Laplace-Beltrami operator, which provides a rich set of eigenbases that are invariant to intrinsic isometries. Then we provide a general and flexible framework for the analysis and design of shape signatures from the spectral graph wavelet perspective. In a bid to capture the global and local geometry, we propose a multiresolution shape signature based on a cubic spline wavelet generating kernel. This signature delivers best-in-class shape retrieval performance. Second, we investigate the ambiguity modeling of codebook for the densely distributed low-level shape descriptors. Inspired by the ability of spatial cues to improve discrimination between shapes, we also propose to adopt the isocontours of the second eigenfunction of the Laplace-Beltrami operator to perform surface partition, which can significantly ameliorate the retrieval performance of the time-scaled local descriptors. To further enhance the shape retrieval accuracy, we introduce an intrinsic spatial pyramid matching approach. Extensive experiments are carried out on two 3D shape benchmarks to assess the performance of the proposed spectral geometric approaches in comparison with state-of-the-art methods.

Table of Contents

List of Tables

List of Figures

List of Acronyms

LB Laplace-Beltrami

SIFT Scale-Invariant Feature Transform

EVD Eigenvalue Descriptor

HKS Heat Kernel Signature

SIHKS Scale-Invariant Heat Kernel Signature

WKS Wave Kernel Signature

GPS Global Point Signature

HMS Heat Mean Signature

BoF Bag-of-Features

VLAD Vector of Locally Aggregated Descriptors

SPM Spatial Pyramid Matching

ISPM Intrinsic Spatial Pyramid Matching

DCG Discounted Cumulative Gain

CWT Continuous Wavelet Transform

SGWT Spectral Graph Wavelet Transform

SGWS Spectral Graph Wavelet Signature

CHAPTER 1

INTRODUCTION

1.1 FRAMEWORK AND MOTIVATION

Recent advances in 3D imaging and processing, graphics hardware and networks have led to a whopping increase in geometry models available freely or commercially on the Web. As a result, the task of efficiently measuring the 3D object similarity to find and retrieve relevant objects for a given query and categorize an object into one of a set of classes has become of paramount importance in a wide range applications, including computer-aided design, video gaming, special effects and film production, medicine, and archaeology. The main challenge in 3D object retrieval algorithms is to compute an invariant shape descriptor that captures well the geometric and topological properties of a shape [1, 2, 3, 4, 5].

A 3D shape is usually represented as a volume or surface/mesh. Other effective representations methods are based on medial [8] or multiple views [37]. Content-based shape retrieval based on the comparison of shape properties is complicated by the fact that many 3D objects manifest rich variability, and invariance to different classes of transformations and shape variations is often required. One of the most challenging settings addressed is the case of nonrigid or deformable shapes, in which the shapes undergoing changes that can be well approximated by intrinsic isometries, i.e. deformations that preserve geodesic distances between all pairs of points. This class of deformations is much richer than rigid motions and can approximate. Recently, various methods have been proposed to tackle nonrigid 3D shape recognition problem, particularly with the isometric invariant representation. These methods can be mainly categorized into two main classes: skeleton-based [6, 7, 8, 9, 10] and surface-

based [11, 12, 13, 14, 15, 17, 18]. The former approaches usually capture the global topological structure of the shape, and a dissimilarity is often determined as the cost function to match two or more shapes. The latter methods, on the other hand, often represent a shape as a frequency histogram of deformation invariant local distances or vertex signatures. Over the past decade, there has been a flurry of research activity on surface based shape recognition due largely to two key reasons: First, surface-based 3D models are more popular because of their highly-effective representation ability and less memory storage. Second, humans are taught to differentiate between shapes mainly by surface features, and in many shape applications only the surface is of interest. Therefore, in this paper, we focus on surface-based shape recognition with local vertex descriptors. Research efforts on spectral shape analysis have recently resulted in numerous spectral descriptors [11, 12, 13, 14, 15, 16], which are predominately based on the LB operator [19, 20]. However, to date, no comprehensive comparison has been conducted in the literature, which often results in intractable situation when choosing appropriate descriptors for certain applications.

In the field of image processing, an analogous problem is image retrieval which refers to finding images depicting similar scenes or objects. Like 3D shapes, images may manifest significant variability. The computer vision and pattern recognition communities have recently witnessed a wide adoption of feature-based methods in object recognition and image retrieval applications. One popular family of feature-based techniques is the Scale-Invariant Feature Transform (SIFT) [21], which has shown a good performance in various scenarios [33]. Feature-based methods also allow to represent images as collections of "visual words" and treat them using text search approaches, such as the codebook model paradigm. State-of-the-art recognition algorithms in image domain usually adopt this local patch based, multiple-layer pipeline to obtain good representations. In this thesis, we follow such a pipeline for 3D surface comparison. It can be viewed in two steps. The first step is to extract an informative local descriptor, and the second step is to design a favorable aggregating comparison method. This thesis solves the deformable 3D shape retrieval problem from both aspects within the popular framework of diffusion geometry: (1) We comprehensively review and compare the recent spectral signatures, and propose the Spectral Graph Wavelet Signature; (2) We investigate the ambiguity modeling of codebook for the densely distributed low-level shape descriptors, and introduce Intrinsic Spatial Pyramid Matching.

1.2 Literature Review

Since the introduction of SIFT and the codebook model, image classification has witnessed a rapid and fruitful development in recent years. By contrast, the overwhelming majority of 3D shape recognition

methods are ad-hoc, and the performance is usually limited due largely to two main reasons: (1) There is no excellent descriptor like SIFT that is available for 3D shapes; (2) an appropriate codebook model and its intrinsic spatial extension are challenging to find due to isometric shape transformation. In this section, we discuss some previous works and current developments pertinent to shape analysis and the codebook model.

1.2.1 Shape Analysis

In recent years, considerable research efforts on shape analysis have been conducted in an effort to design an appropriate shape descriptor aimed at finding the most relevant shapes. In the literature, there are several survey works [1, 2, 3, 4] that have keen interest in systematic shape retrieval and the taxonomy. In the sequel, we present some developments on 3D shape analysis from early general shape description to recent spectral shape analysis.

Shape Description

Early research works on 3D shape description have been centered primarily on invariance under global Euclidean transformations (i.e. rigid transformations). These works include the Shape Context [38, 39], Shape Distributions [40], and Spherical Harmonics [41]. Recently, significant efforts have been invested in exploring the invariance properties of shapes to nonrigid deformations. An intuitive approach is to replace the Euclidean distance with the geodesic one. The primary motivation is that unlike the Euclidean distance, which is basically a straight line between two points in 3D space, the geodesic distance captures the global nonlinear structure and the intrinsic geometry of the data. For example, Elad *et al.* [42] computed a bending invariant signature of a surface by applying the multidimensional scaling procedure to the geodesic distance matrix. In [43], an information-theoretic framework using the geodesic shape distributions was proposed. Also, Jain *et al.* constructed a shape descriptor for correspondence [44] and retrieval [5] in the spectral domain of the geodesic distance matrix. The main drawback of the geodesic distance is that it suffers from strong sensitivity to topological noise, which might heavily damage the shape invariants.

Spectral Shape Analysis

The recently emerging field of diffusion geometry provides a generic framework for many methods in the analysis of geometric shapes [45]. It formulates the heat diffusion processes on manifolds. Spectral shape analysis is a methodology that relies on the eigensystem (eigenvalues and/or eigenfunctions) of

the Laplace-Beltrami operator to compare and analyze geometric shapes. Levy [46] showed that these eigenfunctions can be well adapted to the geometry and the topology of an object. Coifman and Lafon [45] constructed diffusion distances as the L_2-norm difference of energy distribution between two points initialized with unit impulse functions after a given time. Through the statistic of the distribution, the spectral distances can also be used for nonrigid shape recognition [47]. Other similar spectral distances include the commute time distance [48] and the biharmonic distance [49]. Since the eigensystem of the LB operator is isometric invariant, it is well suited for the analysis and retrieval of nonrigid shapes, and it is more robust than the geodesic distance. By integrating the local distribution of features, the Intrinsic Shape Context was proposed in [39] as a natural extension of the 2D Shape Context to 3D nonrigid surfaces, and it was shown to outperform individual vertex descriptors in 3D shape matching.

A recent survey [50] on spectral mesh processing comprehensively reports the spectral methods derived from certain appropriately defined mesh operators and their applications. In this paper, however, we theoretically and experimentally review and compare spectral signatures based on the LB operator, including the Heat Kernel Signature (HKS) [11], Scale Invariant Heat Kernel Signature (SIHKS) [12], Heat Mean Signature [13], Wave Kernel Signature [14], Global Point Signature [15]. The details are provided in Section 2.2.

More recently, the authors in [34] explored analogous codebook model approaches applied to the problem of nonrigid 3D shape retrieval. They use the HKS and the SIHKS as "geometric words", and constructed shape descriptors by means of soft-assignment of visual words to the densely distributed vertex signatures. Low-level features may have a considerable effect on the recognition performance. Following the work in [34], we compare in Chapter 3 the recent spectral descriptors in the framework of the codebook model, and thus comprehensively analyze and recommend the descriptor which plays the same role as SIFT in the image domain. One inherent component of the codebook model is the assignment of discrete codewords to continuous low-level features. Despite the clear mismatch of this hard assignment with the nature of continuous features, the approach has been applied successfully to images. We explicitly model the codeword assignment ambiguity for the densely described 3D shape, which also provides an understanding of the different spectral descriptors.

1.2.2 Codebook Model

The past decade has witnessed the surge in popularity of the codebook model in the image domain. It was first introduced in text retrieval, and then later applied to image categorization in the seminal paper [51]. Subsequent research has focused on overcoming its two intrinsic limitations to improve dis-

crimination, namely (1) the information loss of the assignment of local features to visual words, and (2) the lack of information on the spatial layout of the local features.

Quantization Issues

Increasing the size of the dictionary is often reported to be able to improve the performance of the codebook model, but leads to a higher computational complexity for dictionary construction and feature assignment. On the other hand, when the vocabularies are more compact, the information lost in the quantization process becomes more important, in particular when using hard assignment [52]. By directly using the image-to-class distances without descriptor quantization, Boiman *et al.* [53] showed that the discrimination ability is considerably decreased due to the rough quantization of the feature space. But with the soft-assignment of descriptors to multiple visual words, the loss can be compensated as reported in [54, 55]. Inspired by compressive sensing methodology, other approaches for assignment were guided by sparsity constraints [56] and locality constraints [57].

Bag-of-Features (BoF) usually encodes the 0-order statistics of the distribution of descriptors. The Fisher vector extends the BoF by encoding high-order statistics (first and, optionally, second order). This description vector is the gradient of the sample's likelihood with respect to the parameters of this distribution, scaled by the inverse square root of the Fisher information matrix [58]. A simplified version of Fisher kernels, the Vector of Locally Aggregated Descriptors (VLAD) was also proposed in [59]. These three different ways of aggregating local image descriptors into a vector were evaluated by Jegou *et al.* in [60]. Also, Picard *et al.* [61] expanded the VLAD approach by adding an aggregation of the tensor product of descriptors.

In Chapter 3, the description of 3D shapes is obtained by densely sampling salient points on the surface of the shape. In other words, the spectral signatures on every mesh vertex are considered to obtain the codebook representation. To shed some light on the feature space, we use the Laplacian and Gaussian kernels. We also use different kinds of ambiguity modeling methods to help us understand the information loss in quantization.

Spatial Information

Similar to the image domain, the codebook model representation for 3D surfaces is a frequency histogram of quantized local geometric appearance, where the spatial layout of the geometric appearance is completely ignored [34]. Clearly, the spatial information may convey useful cues to improve the discrimination between 3D shapes. Before modeling the spatial layout on surfaces, it is necessary to review

the technique for images. In the literature, two different ways to encode spatial information have been explored, which are based on local relative positions of pairwise features, and on global absolute positions.

Relative Spatial Relation. Modeling pairwise spatial features into the codebook model is an intuitive way to incorporate spatial information. A spatially-sensitive affine-invariant image descriptor was constructed by Bronstein *et al.* [62] using canonical relation, in which both the features and their relation are affine-invariant. They also generalize the pairwise spatially-sensitive descriptors called "Expression" for 3D surface using the heat kernel as the relation [34]. Moreover, the relationship of visual words was also considered. Saverese *et al.* [63] used correlograms of visual words to model the spatial correlations between quantized local descriptors. Ling and Soatto [64] characterized the relative locations of visual words. Their proximity distribution representation is a 3D structure which records the number of times a visual word appears within a particular number of nearest neighbors of another word. Finally, besides pairwise relation, more complex relation such as the graph manner layout of groups of quantized local invariant descriptors was proposed by Behmo *et al.* [65], which can preserve translational relations between features. Liu *et al.* [66] calculated spatial histograms where the co-occurrences of local features are computed in circular regions of varying distances.

Absolute Spatial Relation. The Spatial Pyramid Matching (SPM), proposed by Lazebnik *et al.* [35], was one of the first works to address the lack of spatial information in the BoF representation. Their spatial pyramid representation was motivated by an earlier work, termed pyramid matching by Grauman and Darrell [67], on finding approximate correspondences between sets of points in high-dimensional feature spaces. The fundamental idea behind pyramid matching is to partition the feature space into a sequence of increasingly coarser grids and then compute a weighted sum over the number of matches that occur at each level of resolution. However, SPM and relative spatial relation modeling are still too weak. Recently, stronger spatially encoding methods include encoding geometric information of objects within the images. Local features of an image are projected onto different directions or points to generate a series of ordered BoF, based on which families of spatial partitions can guarantee the invariance of object translation, rotation, and scaling [68]. Additionally, there are some methods characterizing both the absolute and relative spatial layout of an image. Spatial Pyramid Co-occurrence [69] computes local co-occurrence with respect to spatial layout over a hierarchical spatial partitioning of an image. In addition to co-occurrences, Geometry-Preserving Visual Phrases [70] can encode more spatial information through capturing the local and long-range spatial layouts of the words. Unlike manually defined spatial regions for pooling, Jia *et al.* [71] proposed to learn more adaptive receptive fields to increase the

performance even with a significantly smaller codebook size at the coding layer. In [72], the Gaussian mixture model was encoded with spatial layout to improve the performance of Fisher kernel for image classification.

Local relative position was generalized to 3D surfaces as spatially-sensitive descriptor in [34], but modeling the Absolute Spatial Positions is not straightforward as in the case of 2D images because of the intrinsic deformation of nonrigid shapes. One contribution in this thesis is to introduce the intrinsic partition to capture the global absolute spatial position, thus significantly improving the performance.

1.3 Spectral Geometry

Spectral geometry is concerned with the eigenvalue spectrum of the Laplace-Beltrami (LB) operator on a compact Riemannian manifold, and aims at describing the relationships between such a spectrum and the geometric structure of the manifold.

1.3.1 Laplace-Beltrami Operator

Let $\mathbb{M}$ be a smooth orientable 2-manifold (surface) embedded in $\mathbb{R}^3$. A global parametric representation (embedding) of $\mathbb{M}$ is a smooth vector-valued map (also called surface patch) x defined from a connected open set (parametrization domain) $U \subset \mathbb{R}^2$ to $\mathbb{M} \subset \mathbb{R}^3$ such that

$$x(u) = \left(x^1(u), x^2(u), x^3(u)\right) \tag{1.1}$$

where $u = (u^1, u^2) \in U$, as shown in Figure 1.1.

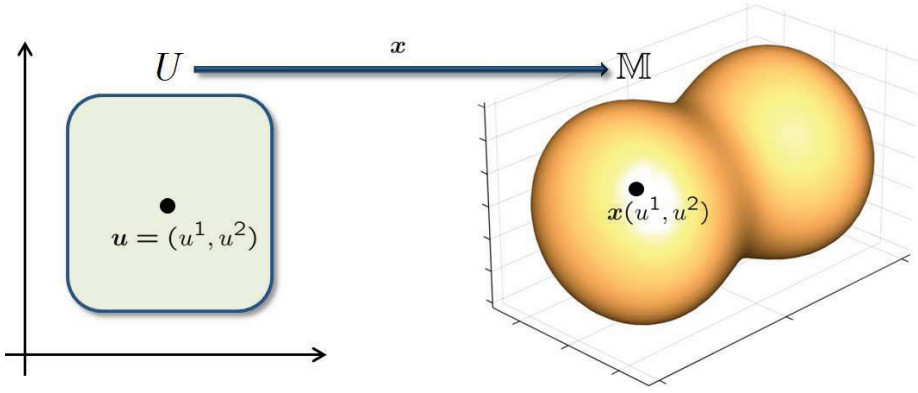


Figure 1.1: Parametric representation of a surface.

Given a twice-differentiable function $f : \mathbb{M} \to \mathbb{R}$, the LB operator [19] is a second-order partial differential operator defined as

$$\begin{aligned}\Delta_{\mathbb{M}} f &= -\frac{1}{\sqrt{|g|}} \sum_{i,j=1}^{2} \frac{\partial}{\partial u^j}\left(\sqrt{|g|}\, g^{ij} \frac{\partial f}{\partial u^i}\right) \\ &= -\sum_{i,j=1}^{2} g^{ij} \frac{\partial}{\partial u^j} \frac{\partial f}{\partial u^i} + (\text{lower order terms})\end{aligned} \tag{1.2}$$

where the matrix $g = (g_{ij})$ is referred to as a *Riemannian metric tensor* on $\mathbb{M}$, g^{ij} denote the elements of the inverse of the metric tensor g^{-1}, and $|g|$ is the determinant of g. The functions g_{ij} are sometimes referred to as the *metric coefficients*. The Riemannian metric g is an intrinsic quantity in the sense that it relates to measurements inside the surface. It is the analogous of the speed in the case of space curves, and determines all the intrinsic properties of the surface $\mathbb{M}$. These properties depend on the surface and do not depend on its embedding in space. Additionally, the tensor g is invariant to rotation of the surface in space because it is defined in terms of inner products that are rotation invariant.

1.3.2 Discretization

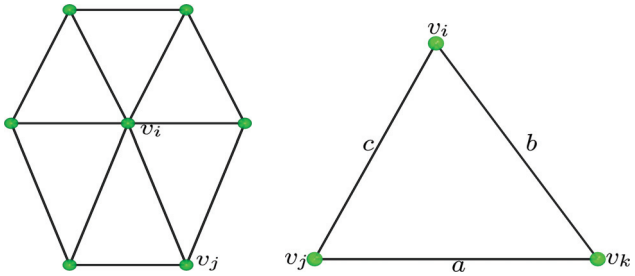
Assume that the surface $\mathbb{M}$ is approximated by a triangular mesh. A triangle mesh $\mathbb{M}$ may be defined as $\mathbb{M} = (\mathcal{V}, \mathcal{E})$ or $\mathbb{M} = (\mathcal{V}, \mathcal{T})$, where $\mathcal{V} = \{v_1, \ldots, v_m\}$ is the set of vertices, $\mathcal{E} = \{e_{ij}\}$ is the set of edges, and $\mathcal{T} = \{t_1, \ldots, t_n\}$ is the set of triangles. Each edge e_{ij} (denoted by $[v_i, v_j]$ or simply $[i, j]$) connects a pair of vertices $\{v_i, v_j\}$. Two distinct vertices $v_i, v_j \in \mathcal{V}$ are adjacent (denoted by $v_i \sim v_j$ or simply $i \sim j$) if they are connected by an edge, i.e. $e_{ij} \in \mathcal{E}$. The neighborhood (1-ring) of a vertex v_i is the set $v_i^\star = \{v_j \in \mathcal{V} : i \sim j\}$. The neighborhood (1-ring) of a vertex v_i is the set $v_i^\star = \{v_j \in \mathcal{V} : i \sim j\}$ as shown in Figure 1.2(a). Consider a triangle $t \in \mathcal{T}$ with vertices v_i, v_j and v_k, and sides a, b and c as illustrated in Figure 1.2(b). Then, according to Heron's formula, area(t) is equal to

$$\frac{1}{4}\sqrt{(a+(b+c))(a+(b-c))(c+(a-b))(c-(a-b))}, \tag{1.3}$$

where the length of the sides are arranged such that $a \geq b \geq c$.

Several discretizations of the LB operator are available in the literature [16, 93, 74, 75, 76]. In this thesis, we use the approach developed in [93], which employs a mixed finite element/finite volume method on triangle meshes. Hence, the value of $\Delta_{\mathbb{M}} f$ at a vertex v_i can be approximated using the cotangent weight scheme as follows

$$\Delta_{\mathbb{M}} f(v_i) = \frac{1}{a_i} \sum_{v_j \in v_i^\star} \frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} \left[f(v_j) - f(v_i)\right], \tag{1.4}$$



(a) (b)

FIGURE 1.2: (a) Vertex neighborhood $v_i^\star$. (b) Illustration of area(t).

where α_{ij} and β_{ij} are the angles $\angle(v_i v_{k_1} v_j)$ and $\angle(v_i v_{k_2} v_j)$ of two faces $t^\alpha = \{v_i, v_j, v_{k_1}\}$ and $t^\beta = \{v_i, v_j, v_{k_2}\}$ that are adjacent to the edge $[i, j]$, and a_i is the area of the voronoi cell (shaded polygon), as shown in Figure 1.3. It is worth pointing out that the cotangent weight scheme is numerically consistent and preserves several important properties of the continuous LB operator, including symmetry and positive-definiteness [74].

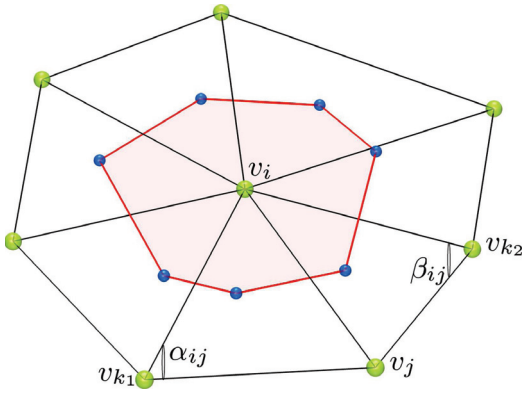


FIGURE 1.3: Cotangent weight scheme: illustration of the angles α_{ij} and β_{ij}.

Define the weight function $\omega : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ as

$$\omega_{ij} = \begin{cases} \dfrac{\cot \alpha_{ij} + \cot \beta_{ij}}{2a_i} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{1.5}$$

Then, for a function $f : \mathcal{V} \to \mathbb{R}$ that assigns to each vertex $v_i \in \mathcal{V}$ a real value $f(v_i)$ (we can view f as a column vector of length m), we may write the LB operator given by Eq. (1.4) as

$$Lf(v_i) = \sum_{v_j \in v_i^\star} \omega_{ij} \left(f(v_i) - f(v_j) \right), \tag{1.6}$$

where the matrix L is given by

$$L = \begin{cases} d_j & \text{if } i = j \\ -\omega_{ij} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{1.7}$$

and $d_j = \sum_{i=1}^{m} \omega_{ij}$ is the weighted degree of the vertex v_i.

1.3.3 EIGENANALYSIS

Note that $\omega_{ij} \neq \omega_{ji}$ implies L is not a symmetric matrix. Thus, the spectrum (set of eigenvalues) of the eigenvalue problem $L\varphi_i = \lambda_i \varphi_i$ may not be real [15]. Noting that $\omega_{ij} = c_{ij}/a_i$, where

$$c_{ij} = \begin{cases} \dfrac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{1.8}$$

we may factorize the matrix L as $L = A^{-1}C$, where $A = \text{diag}(a_i)$ is a positive-definite diagonal matrix and C is a sparse symmetric matrix given by

$$C = \begin{cases} \sum_{i=1}^{m} c_{ij} & \text{if } i = j \\ -c_{ij} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{1.9}$$

Therefore, we may write the eigenvalue problem $L\varphi_i = \lambda_i \varphi_i$ as a generalized eigenvalue problem $C\varphi_i = \lambda_i A \varphi_i$, which can be solved efficiently using the Arnoldi method of ARPACK. Figure 1.4 shows a 3D elephant model and the sparsity pattern of the cotangent matrix C. Recall that the sparsity pattern (or support) of a matrix $A = (a_{ij})$ is the set of indices ij with $a_{ij} \neq 0$.

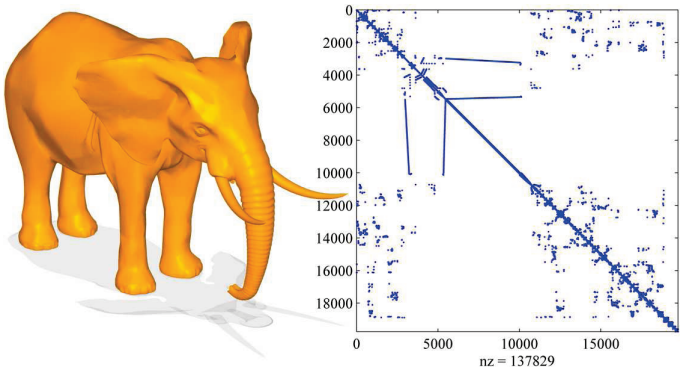


FIGURE 1.4: 3D elephant model and sparsity pattern plot of the associated cotangent matrix C.

1.4 Thesis Overview and Contributions

The organization of this thesis is as follows

- ❑ Chapter 1 contains a brief review of essential concepts and definitions which we refer to throughout the thesis, provides a literature review, and presents a short summary of material relevant to 3D shape retrieval in the spectral geometric framework.

- ❑ In Chapter 2, we present a comprehensive review and analysis of recent spectral shape descriptors for nonrigid 3D shape retrieval. More specifically, we compare the latest spectral descriptors based on the LB operator, including shapeDNA, heat kernel signature, scale invariant heat kernel signature, heat mean signature, wave kernel signature, and global point signature. We also include the eigenvalue descriptor (EVD), which is a geodesic distance based shape signature. The global descriptors ShapeDNA and EVD are compared via the chi-squared distance, while all local descriptors are compared using the codebook model. Moreover, we investigate the ambiguity modeling of codebook for the densely distributed low-level shape descriptors. Inspired by the ability of spatial cues to improve discrimination between shapes, we also propose to adopt the isocontours of the second eigenfunction of the LB operator to perform surface partition, which can significantly ameliorate the retrieval performance of the time-scaled local descriptors. In addition, we introduce an intrinsic spatial pyramid matching approach in a bid to further enhance the retrieval accuracy. Extensive experiments are carried out on two 3D shape benchmarks to assess the performance of the spectral descriptors.

- ❑ In Chapter 3, we a spectral graph wavelet framework for the analysis and design of efficient shape signatures for nonrigid 3D shape retrieval. Although this work focuses primarily on shape retrieval, our approach is, however, fairly general and can be used to address other 3D shape analysis problems. In a bid to capture the global and local geometry of 3D shapes, we propose a multiresolution signature via a cubic spline wavelet generating kernel. The parameters of the proposed signature can be easily determined as a tradeoff between effectiveness and compactness. Experimental results on two standard 3D shape benchmarks demonstrate the much better performance of the proposed shape retrieval approach in comparison with three state-of-the-art methods.

- ❑ In Chapter 4, we summarize the contributions of this thesis, and propose several future research directions that are directly or indirectly related to the ideas developed therein.

CHAPTER 2

INTRINSIC SPATIAL PARTITION MATCHING

This chapter presents a comprehensive review and analysis of recent spectral shape descriptors for non-rigid 3D shape retrieval. More specifically, we compare the latest spectral descriptors based on the Laplace-Beltrami (LB) operator, including shapeDNA, heat kernel signature, scale invariant heat kernel signature, heat mean signature, wave kernel signature, and global point signature. We also include the eigenvalue descriptor (EVD), which is a geodesic distance based shape signature. The global descriptors ShapeDNA and EVD are compared via the chi-squared distance, while all local descriptors are compared using the codebook model. Moreover, we investigate the ambiguity modeling of codebook for the densely distributed low-level shape descriptors. Inspired by the ability of spatial cues to improve discrimination between shapes, we also propose to adopt the isocontours of the second eigenfunction of the LB operator to perform surface partition, which can significantly ameliorate the retrieval performance of the time-scaled local descriptors. In addition, we introduce an intrinsic spatial pyramid matching approach in a bid to further enhance the retrieval accuracy. Extensive experiments are carried out on two 3D shape benchmarks to assess the performance of the spectral descriptors. Our proposed approach is shown to provide the best performance.

2.1 INTRODUCTION

State-of-the-art recognition algorithms in image domain usually adopt a local patch based, multiple-layer pipeline to obtain a good representation. These methods start from local image patches using either normalized raw pixel density or descriptors such as SIFT [21] or HOG [22], and encode them into

an overcomplete representation using various algorithms such as K-means or sparse coding. After coding, global image representations are formed by spatially pooling the coded local descriptors. Methods following such a pipeline have achieved competitive performance on image classification tasks [23]. During the whole procedure, the spatial pooling step brings a notable performance improvement. One significant milestone in the construction of this arsenal of tools was the *spatial pyramid matching* (SPM) introduced in [35]. It partitions the image into increasingly fine sub-regions and computes histograms of local features found inside each sub-region. The empirical success of technique is the fact that the spatial cue is integrated, and an approximate geometric matching is actually performed when multiple resolutions are combined in a principled way.

Codebook model, as a simplified version of such pipeline without spatial pooling, has also been considered for 3D shapes and curved surfaces. Early research mainly deal with the global Euclidean transformations (rigid motion) [24] and multiple views [25]. Recent effort has also been invested in extending the invariance properties to non-rigid deformations. One of the common ways is to replace the Euclidean metric with its geodesic counterpart [26]. However, geodesic distances suffer from strong sensitivity to topological noise, which limits its usefulness in real applications.

The emerging field of *diffusion geometry* provides a generic framework for many intrinsic methods in the analysis of geometric shapes. It formulates the heat diffusion processes on manifolds. Based on the theoretical works by [27], later Lévy [46] showed that the eigenfunctions of the Laplace-Beltrami operator can be well adapted to the geometry and the topology of an object. Coifman and Lafon [45] introduced invariant metrics known as diffusion distances, which is the L_2-norm difference of energy distribution between two points initialized with unit impulse functions after a given time. Finally, *Shape Google* algorithm [34] was proposed as a classic method for deformable shape retrieval. It uses the multiscale diffusion heat kernels as "geometric words", and constructs compact and informative shape representation by means of the codebook approach.

Recently, there have been attempts to adapt 2D planar shape contexts [28], popular image feature detectors [29] and descriptors [30] to 3D surfaces. This line of works partially inspires our paper. Another inspiration is the great success of SPM in image domain. Spatially enhanced techniques for 3D shape recognition were explored earlier in [31, 32]; but these works are not intrinsic, i.e. shape deformations affect the descriptors. "Geometric expressions" [34] was an earlier work that explored the exploitation of intrinsic geometry, but there the authors only deal with local relative spatial position, by considering diffusion distance between pairwise vertices. Our approach models the global absolute spatial positions, which allows us retain and exploit the information contained in the whole shape.

For shape retrieval tasks, the codebook models, which represent a shape as an orderless collection of local features, have demonstrated impressive levels of performance [34]. However, because these models disregard all information about the spatial layout of the features, they have a limited descriptive ability. Several spatial extensions in the image domain have been proposed recently, the most widely-used one is spatial pyramid matching [35]. Unfortunately, overcoming the spatial limitations in 3D shape analysis to build effective structural object descriptions is quite challenging, especially when the recognition system must be designed to work in the presence of large deformation changes. The direct approaches can use the existing consistent shape segmentation methods and geometric correspondence, but they achieve a good performance at a relatively high computational cost in addition to the fact the number of subregions cannot be fixed among different shape classes.

Inspired by the fact that the Reeb graph extracted from the second eigenfunction of the LB operator is pose independent and captures the global profile of surface geometry [36], we propose to adopt the level sets of this eigenfunction to intrinsically partition the surface. Since the construction of the second eigenfunction is an inseparable step in calculating the spectral descriptors, the proposed partition method is a natural ingredient of the current framework. Extensive experimental results show that the intrinsic partition significantly improves the retrieval accuracy of all the time-scaled spectral descriptors with varying codebook models. Moreover, the intrinsic spatial pyramid matching on surfaces is shown to be robust and yields the best results. In addition, the intrinsic spatial partition framework offers further insight into the success of these recently proposed spectral shape descriptors.

2.1.1 Contributions

The contributions of this chapter may be summarized as follows:

(i) We present a comprehensive survey and analysis of recent spectral descriptors for nonrigid 3D shape retrieval.

(ii) We investigate the ambiguity modeling of codebook for the densely distributed low-level shape descriptors.

(iii) We introduce the intrinsic spatial partition, which yields a significant retrieval accuracy improvement.

The rest of this paper is organized as follows. Section 2.2 briefly reviews the graph embedding and spectral shape descriptors. In Section 2.3, we describe the codebook model with various ambiguity methods.

In Section 2.4, we propose the intrinsic spatial partition. Experimental results on two standard datasets are presented in Section 2.5.

2.2 Graph Embedding and Shape Descriptors

Based the the isometric invariant of the Laplace-Beltrami operator, several spectral shape descriptors have been proposed recently. The Laplace-Beltrami operator over a compact manifold is bounded and symmetric negative semi-definite. Hence it has an eigen-decomposition. The eigenvalues λ_i and associated eigenfunctions φ_i of the LB operator can be computed by solving the generalized eigenvalue problem:

$$C\varphi_i = \lambda_i A\varphi_i, \quad i = 1, 2, \ldots, m \tag{2.1}$$

where φ_i is the unknown eigenfunction evaluated at m mesh vertices. That is, φ_i is an m-dimensional vector. We may sort the eigenvalues in ascending order as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_m$ with corresponding eigenfunctions as $\varphi_1, \varphi_2, \ldots, \varphi_m$, where each eigenfunction $\varphi_i = (\varphi_i(v_1), \ldots, \varphi_i(v_m))'$ is an m-dimensional vector. Note that the eigensystem $\{\lambda_i, \varphi_i\}_i$ is intrinsic to the manifold and enjoys a nice property of being isometric invariant. It should also be noted that the meshes are assumed to be connected.

2.2.1 ShapeDNA and Eigenvalue Descriptors

The ShapeDNA [16] is one of the first spectral shape descriptors. It is a normalized sequence of the first eigenvalues of the LB operator. Its main advantages are the simple representation (a vector of numbers) and scale invariance. Despite its simplicity, the shapeDNA yields a very good performance for the shape retrieval of nonrigid shapes. However, the eigenvalues are a global descriptor, therefore the shapeDNA cannot be used for local or partial shape analysis. The Eigenvalue Descriptor (EVD) [5], on the other hand, is a sequence of the eigenvalues of the geodesic distance matrix. Both ShapeDNA and EVD can be normalized by the second eigenvalue.

2.2.2 Global Point Signature

The global point signature (GPS) [15] at a surface point is a vector of scaled eigenfunctions of the LB operator. The GPS is a global feature in the sense that it cannot be used for partial shape matching. It is defined in terms of the eigenvalues and eigenfunctions of $\Delta_{\mathbb{M}}$ as follows:

$$\mathrm{GPS}(x) = \left(\frac{\varphi_2(x)}{\sqrt{\lambda_2}}, \frac{\varphi_3(x)}{\sqrt{\lambda_3}}, \ldots, \frac{\varphi_i(x)}{\sqrt{\lambda_i}}, \ldots \right) \tag{2.2}$$

GPS is invariant under isometric deformations of the shape, but it suffers for the problem of eigenfunctions switching whenever the associated eigenvalues are close to each other.

2.2.3 Heat Kernel Signature

The heat kernel $\mathfrak{p}_t(x,y)$ is a fundamental solution to the heat equation [77] at point x at time t with initial distribution $u_0(x) = \delta(x-y)$ at point $y \in \mathbb{M}$, and it is defined in terms of the eigenvalues and eigenfunctions of $\Delta_{\mathbb{M}}$ as follows:

$$\mathfrak{p}_t(x,y) = \sum_{i=1}^{\infty} e^{-\lambda_i t}\varphi_i(x)\varphi_i(y) \tag{2.3}$$

Intuitively, $\mathfrak{p}_t(x,y)$ describes the amount of heat that is propagated or transferred from point x to point y in time t. Therefore, $\mathfrak{p}_t(x,x)$ describes the amount of heat that remains at point x after time t. For each point $x \in \mathbb{M}$, the Heat Kernel Signature (HKS) [11] is represented in the discrete temporal domain by a n-dimensional feature vector

$$\mathrm{HKS}(x) = \left(\mathfrak{p}_{t_1}(x,x), \mathfrak{p}_{t_2}(x,x), \ldots, \mathfrak{p}_{t_n}(x,x)\right) \tag{2.4}$$

where $t_1, t_2, \ldots, t_n$ are different time-scales.

2.2.4 Scale Invariant Heat Kernel Signature

Let $\mathbb{M}$ and $\mathbb{M}'$ be a shape and its uniformly scaled version by a factor of a, respectively. Denote by $\mathfrak{p}_{\alpha^\tau}(x,y)$ the heat kernel with scale logarithmically sampled using some basis α at each point x. Thus, the heat kernel of the scaled shape becomes $\mathfrak{p}'(\tau) = a^{-2}\mathfrak{p}(\tau + 2\log_\alpha a)$. In order to remove the dependence on the multiplicative constant a^{-2}, the logarithm of the signal is taken and then differentiated with respect to the scale variable [12]:

$$\begin{aligned}\frac{d}{d\tau}\log\mathfrak{p}'(\tau) &= \frac{d}{d\tau}(-2\log a + \log\mathfrak{p}(\tau + 2\log_\alpha a) \\ &= \frac{\frac{d}{d\tau}\mathfrak{p}(\tau + 2\log_\alpha a)}{\mathfrak{p}(\tau + 2\log_\alpha a)}.\end{aligned} \tag{2.5}$$

Let $\mathfrak{p}' = \frac{\frac{d}{d\tau}\mathfrak{p}(\tau)}{\mathfrak{p}(\tau)} = \frac{-\sum_{i\geq 0}\lambda_i\alpha^\tau\log\alpha e^{-\lambda_i\alpha^\tau}\varphi_i^2(x)}{-\sum_{i\geq 0}e^{-\lambda_i\alpha^\tau}\varphi_i^2(x)}$ then a new function $\tilde{\mathfrak{p}}$ which transforms $\tilde{\mathfrak{p}}'(\tau) = \tilde{\mathfrak{p}}(\tau + 2\log_\alpha a)$ as a result of scaling is obtained. The Fourier transform of $\tilde{\mathfrak{p}}$ and its absolute value are given by

$$\begin{aligned}F\left[\tilde{\mathfrak{p}}'\right](\omega) = \tilde{H}'(\omega) &= \tilde{H}(\omega)e^{-j\omega 2\log_\alpha a} \\ |\tilde{H}'(\omega)| &= |\tilde{H}(\omega)|.\end{aligned} \tag{2.6}$$

Thus, the Scale-Invariant Heat Kernel Signature (SIHKS) is defined as

$$\mathrm{SIHKS}(x) = \left(|\tilde{H}(\omega_1)|, |\tilde{H}(\omega_2)|, \ldots, |\tilde{H}(\omega_n)|\right). \tag{2.7}$$

2.2.5 Wave Kernel Signature

The fundamental idea of the Wave Kernel Signature (WKS) [14] is to to represent a point $x \in \mathbb{M}$ by the average probabilities of quantum particles of different energy levels to be measured in x. Assume a quantum particle with unknown position is on the surface. Then the wave function of the particle is the Schrödinger equation solution, which can expressed in the spectral domain as

$$\psi_E(x,t) = \sum_{k=1}^{\infty} e^{i\lambda_k t}\varphi_k(x) f_E(\lambda_k) \tag{2.8}$$

where E denotes the energy of the particle at time $t = 0$ and f_E its initial distribution.

Since $\left|\psi_E(x,t)\right|^2$ is the probability to measure the particle at a point x at time t, it follows that the average probability (over time) to measure a particle in x is given by

$$\mathtt{P}_E(x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \left|\psi_E(x,t)\right|^2 = \sum_{k=1}^{\infty} \varphi_k(x)^2 f_E(\lambda_k)^2 \tag{2.9}$$

Let $E_1, E_2, \ldots, E_n$ be n log-normal energy distributions. Then, each point x on the surface $\mathbb{M}$ is associated with a wave kernel signature, which can represented by a n-dimensional feature vector of average probabilities as follows:

$$\mathrm{WKS}(x) = \left(\mathtt{P}_{e_1}(x), \mathtt{P}_{e_2}(x), \cdots, \mathtt{P}_{e_n}(x)\right) \tag{2.10}$$

where $e_i = \log E_i$ is the logarithmic energy scale. The WKS represents the average probability of measuring a quantum particle at a specific surface point. Unlike the HKS, the WKS separates influences of different frequencies, treating all frequencies equally. In other words, HKS uses low-pass filters, while WKS uses band-pass filters.

2.2.6 Heat Mean Signature

The Heat Mean Signature (HMS) [13] quantitatively evaluate the temperature distribution resulting from the heat flow process

$$\mathrm{HMS}_t(x) = \frac{1}{m}\sum_{y\neq x} \mathfrak{p}_t(x,y) \tag{2.11}$$

which can be physically interpreted as the average temperature on the surface obtained by applying a unit amount of heat on the vertex x and after a certain amount of time of heat dissipation. A relatively smaller parameter t is often empirically chosen to preserve a higher resolution version of the original surface [**?**]. Fang *et al.* also proposed the temperature distribution descriptor [17], which is based on the distribution of the values of average temperature for all of the vertices on the mesh. We construct a multi-scale HMS to compare temperature distribution with multiple diffusion times as follows:

$$\mathrm{HMS}(x) = \left(\mathrm{HMS}_{t_1}, \mathrm{HMS}_{t_2}, \ldots, \mathrm{HMS}_{t_n}\right). \tag{2.12}$$

For the sake of notational simplicity, we use $s(x)$ to represent the types of the above spectral signatures evaluated at a surface point x, i.e. GPS, HKS, SIHKS, WKS, or HMS.

2.3 Aggregating Local Descriptors with Codebook Models

In this section, we focus on the methods for aggregating dense local spectral descriptors into a compact representation of the whole shape. Bronstein *et al.* initiated the study of the nonrigid 3D shape retrieval via BoF with soft-assignment [34]. In this paper, we comprehensively investigate some variants of the codebook model for aggregating these local spectral descriptors in a dense signature space. In particular, we propose an intrinsic spatial partition, which can be seen as the counterpart of the spatial extension for the codebook model in image recognition; thus further enhancing the results.

Given a set of local point-wise signatures densely computed on each vertex on the mesh surface, we quantize the signature space to obtain a compact histogram representation of the shape using the codebook model approach. The visual word vocabulary in the codebook model may be constructed in various ways. We use the k-means algorithm to generate the visual vocabulary. This is computationally expensive, but as this step is performed off-line, it has no impact on the search time. Thus, the "geometric words" of a vocabulary $P = \{p_k,\ k = 1,2,\ldots,K\}$ are obtained as the K centroid of k-means clustering in the signature space. We can use various types of spectral descriptors. From any shape, we get a specific type of local spectral descriptors $S = \{s_t,\ t = 1,2,\ldots,T\}$ for comparison. By a certain vector coding technique, each shape will be described by a histogram H. Since the number of vertices is usually different among different meshed shapes, an appropriate normalization technique is also important for the codeword-cumulative histogram representation. We normalize P by the total number of vertices of each shape.

Modeling the codeword ambiguity plays a crucial role on the performance of the codebook model. In the literature, visual word ambiguity modeling is used occasionally, often ad-hoc motivated, and rarely evaluated. However, a formal summarized work was recently proposed by Gemert *et al.* [55], who motivated and evaluated several types of visual word ambiguity, and provided ample analysis. For completeness, we introduce these types in the scenario of 3D geometric shapes:

- L_p **norm Codebook:** Each local descriptor s_t is associated with its nearest visual word $NN(s_t)$ in the codebook. For each codeword p_k, the differences of vector s_t assigned to p_k are accumulated by L_p norm as follows:

$$q_i = \sum_{s_t : NN(s_t) = i} \left\| s_t - p_i \right\|_\mathrm{p}, \quad \mathrm{p} = 0,1,2 \tag{2.13}$$

Note that L_0 is the traditional codebook, which is the histogram of the number of local descriptors assigned to each visual word.

- **Kernel Codebook:** The histogram estimator of the code words may be replaced by a kernel density estimator. Moreover, a suitable kernel (such as the Gaussian) allows kernel density estimation to become part of the codewords, rather than the data samples. A symmetric kernel allows for transferring the kernel from the data samples to the codewords, yielding a kernel codebook

$$q_i = \sum_{t=1}^{T} K_\sigma(\varrho(s_t, p_i)), \tag{2.14}$$

  In order to make it clear which distance is more fit for the descriptors, we use both the L_1 and L_2 norms as distance functions. The Euclidean distance (L_2-norm) is paired with a Gaussian-shaped kernel, while the L_1-norm is paired with a Laplacian-shaped kernel. The latter assumes that the variation between a local feature and a codeword is described by a sharper distribution. Both distributions have a smoothing parameter σ which represents the size of the kernel. For simplicity, we denote the kernels as $K_\sigma(\varrho(\cdot,\cdot))$, where $\varrho(\cdot,\cdot)$ is the L_1-norm when it is the Laplacian kernel, and L_2-norm when it is the Gaussian kernel.

- **Codeword Uncertainty:** Codeword uncertainty indicates that one image region may distribute probability mass to more than one codeword. It is modeled to normalize the amount of probability mass to a total constant weight of 1 and is distributed over all relevant codewords. Relevancy is determined by the ratio of the kernel values for all codewords p_i in the vocabulary

$$q_i = \sum_{t=1}^{T} \frac{K_\sigma(\varrho(s_t, p_i))}{\sum_{k=1}^{K} K_\sigma(\varrho(s_t, p_k))}. \tag{2.15}$$

- **Codeword Plausibility:** Codeword plausibility means that an image feature may not be close enough to warrant representation by any relevant codeword in the vocabulary. For a given descriptor s_t, it selects the best fitting codeword p_i and assigns its probability mass proportional to the kernel value of that codeword. Hence, codeword plausibility will give a higher weight to more relevant data samples. However, it cannot select multiple codeword candidates

$$q_i = \sum_{s_t : NN(s_t) = i} K_\sigma(\varrho(s_t, p_i)). \tag{2.16}$$

The four types of ambiguity modeling methods use different numbers of geometric words in their construction. In the traditional codebook and codeword plausibility, the local descriptor only selects the

best candidate geometric word. On the other hand, the kernel codebook and codeword uncertainty divide the descriptor over multiple best fitting codewords. To formally compare the different ambiguity ways, we summarize all the codebook models in Table 2.1 and we categorize them in terms of the L_1- and L_2-norms.

TABLE 2.1: Codebook Model.

Clustering Norm	Vector Assignment	
L_1	Traditional Codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \|s_t - p_i\|_0$
	L_1-norm Codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \|s_t - p_i\|_1$
	L_2-norm Codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \|s_t - p_i\|_2$
	Laplace Kernel Codebook	$q_i = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2b} e^{-\frac{\|s_t-p_i\|_1}{b}}$
	Laplace Codeword Uncertainty	$q_i = \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{1}{2b} e^{-\frac{\|s_t-p_i\|_1}{b}}}{\sum_{k=1}^{K} \frac{1}{2b} e^{-\frac{\|s_t-p_k\|_1}{b}}}$
	Laplace Codeword Plausibility	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \frac{1}{2b} e^{-\frac{\|s_t-p_i\|_1}{b}}$
L_2	Traditional Codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \|s_t - p_i\|_0$
	L_1-norm Codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \|s_t - p_i\|_1$
	L_2-norm Codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \|s_t - p_i\|_2$
	Gaussian Kernel Codebook	$q_i = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\|s_t-p_i\|_2^2}{\sigma^2}}$
	Gaussian Codeword Uncertainty	$q_i = \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\|s_t-p_i\|_2^2}{\sigma^2}}}{\sum_{k=1}^{K} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\|s_t-p_k\|_2^2}{\sigma^2}}}$
	Gaussian Codeword Plausibility	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\|s_t-p_i\|_2^2}{\sigma^2}}$

The kernel size determines the degree of coherence to assign geometric word to a descriptor, and it is dependent on the descriptor dimensionality and the range of the descriptor values. Moreover, we only consider the kernel size that is fixed for all codewords. The case of constructing the variable kernel density estimator for different codewords can also be considered, but we adhere to a homogenous feature space by keeping the kernel size fixed for all codewords [55]. Note that we do not try to obtain the best fit of the data. In contrast, we aim at finding the kernel size that discriminates well between classes. In the experimental results section, we estimate the optimal kernel size in an interval inferred from the data distribution.

Besides directly modeling ambiguity on individual geometric words, ambiguity might be addressed by modeling geometric word co-occurrences. Co-occurrence modeling may address ambiguity because it

is likely that similar geometric words with high ambiguity co-occur frequently. When these ambiguous geometric words are grouped together, their intra-ambiguity is resolved. For 3D shapes, Bronstein *et al.* [34] introduced the spatially-sensitive Bag-of-Words description, which accounts not only for the frequency but also for the spatial relations between features. In this paper, since we are interested in analyzing different spectral descriptors and measuring ambiguity, we concentrate on single word ambiguity modeling.

2.4 Intrinsic Spatial Partition: Beyond Codebook Model

2.4.1 Isocontours

The eigenfunctions of the LB operator enjoy nice properties including isometry invariance and robustness to pose variations such as translation and rotation. These eigenfunctions are orthogonal $\langle \varphi_i, \varphi_j \rangle_A = 0, \quad \forall i \neq j$, where the orthogonality is defined in terms of the A-inner product. That is, $\langle \varphi_i, \varphi_j \rangle_A = \varphi_i' A \varphi_j$. Moreover, any function $f : \mathcal{V} \to \mathbb{R}$ (viewed as a column-vector of length m) on the triangle mesh $\mathbb{M}$ can be written in terms of the eigenfunctions as follows:

$$f = \sum_{i=1}^{m} \alpha_i \varphi_i, \quad \text{where} \quad \alpha_i = \langle f, \varphi_i \rangle. \tag{2.17}$$

Note that since the sum of each row in the matrix C equals zero, the first eigenvalue λ_1 is zero and the corresponding eigenfunction φ_1 is a constant m-dimensional vector. The top row of Figure 2.1 shows a 3D horse model colored by the second, third and fourth eigenfunctions, while the bottom row displays the isocontours of these eigenfunctions.

We can use the variational characterizations of the eigenvalues in terms of the Rayleigh-Ritz quotient. That is, the second eigenvalue is given by

$$\lambda_2 = \inf_{f \perp \varphi_1} \frac{f'Cf}{f'Af} = \inf_{f \perp \varphi_1} \frac{\sum_{i \sim j} c_{ij} (f(v_i) - f(v_j))^2}{\sum_i f(v_i)^2 a_i} \tag{2.18}$$

and $\varphi_2 = (\varphi_2(v_1), \ldots, \varphi_2(v_m))'$ is its corresponding eigenvector.

The eigenvalues and eigenfunctions have a nice physical interpretation: the square roots of the eigenvalues $\sqrt{\lambda_i}$ are the eigenfrequencies of the membrane, and $\varphi_i(x)$ are the corresponding amplitudes at x. In particular, the second eigenvalue λ_2 corresponds to the sound we hear the best. On the other hand, Uhlenbeck [78] showed that the eigenfunctions of the LB operator are Morse functions on the interior of the domain of the operator. Consequently, this generic property of the eigenfunctions gives rise to constructing the associated intrinsic isocurves.

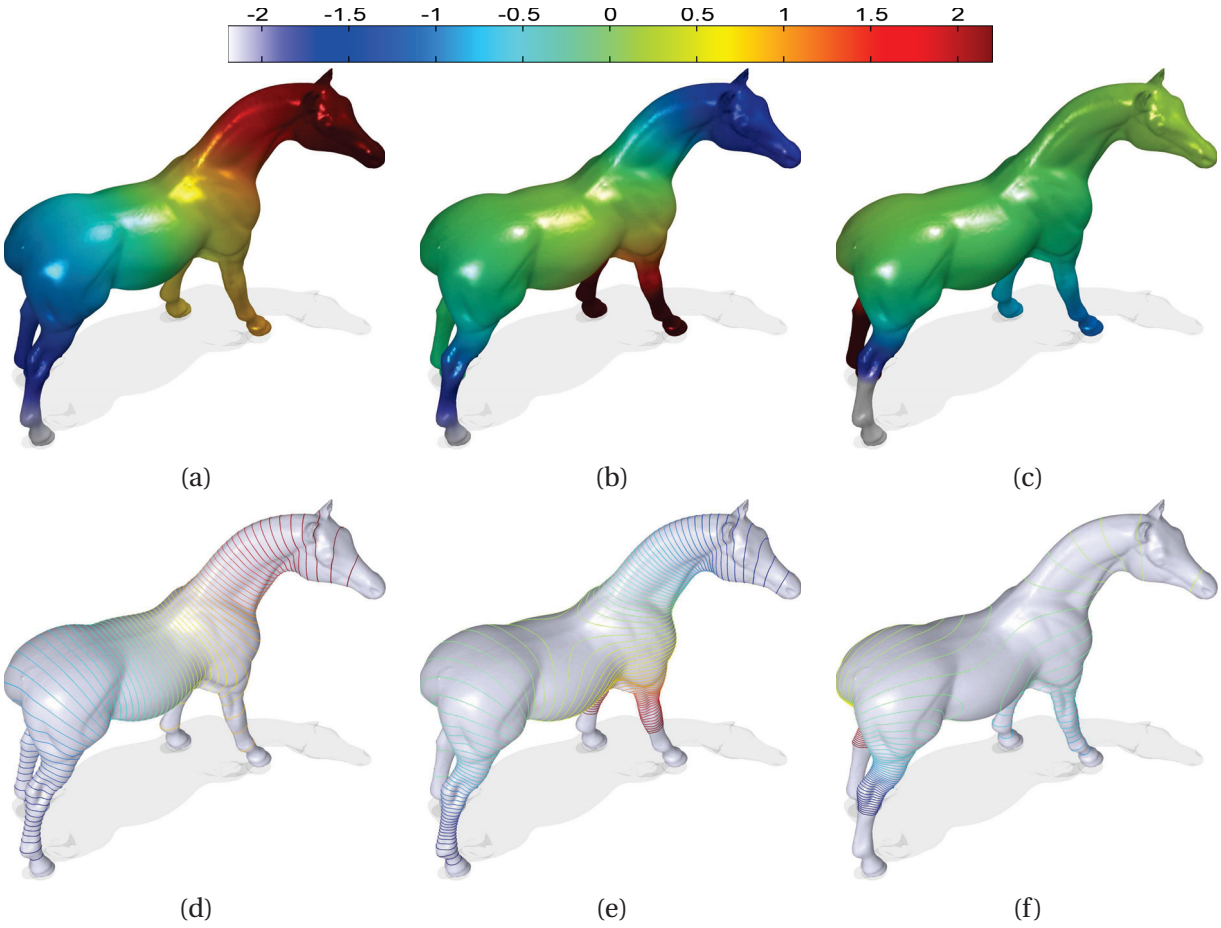


FIGURE 2.1: (a)-(c) 3D horse model colored by $\varphi_2, \varphi_3, \varphi_4$. (d)-(f) Level sets of $\varphi_2, \varphi_3, \varphi_4$.

2.4.2 INTRINSIC SPATIAL PARTITION

Motivated by the isometric invariance property of the second eigenfunction of the LB operator and also by its generic property as a Morse function, we propose to use the level sets (isocontours) of the second eigenfunction as cuts to partition the surfaces. In Figure 2.2(a)-(c), we show some examples of the level curves of φ_2. In Figure 2.2(a) we can observe that the isocontours are consistent with global large deformation (first column), local small bend (second column), and among the shapes from different classes, but share similar topological structure (third column). The correspondence of isocontours on the shapes from the same class are displayed in Figure 2.2(b), which shows models that include various topological structures. Finally, the consistency of isocontours on the shapes from the different class are displayed in Figure 2.2(c). Although the shapes are explicitly different, their isocontours can capture their intrinsic correspondence well.

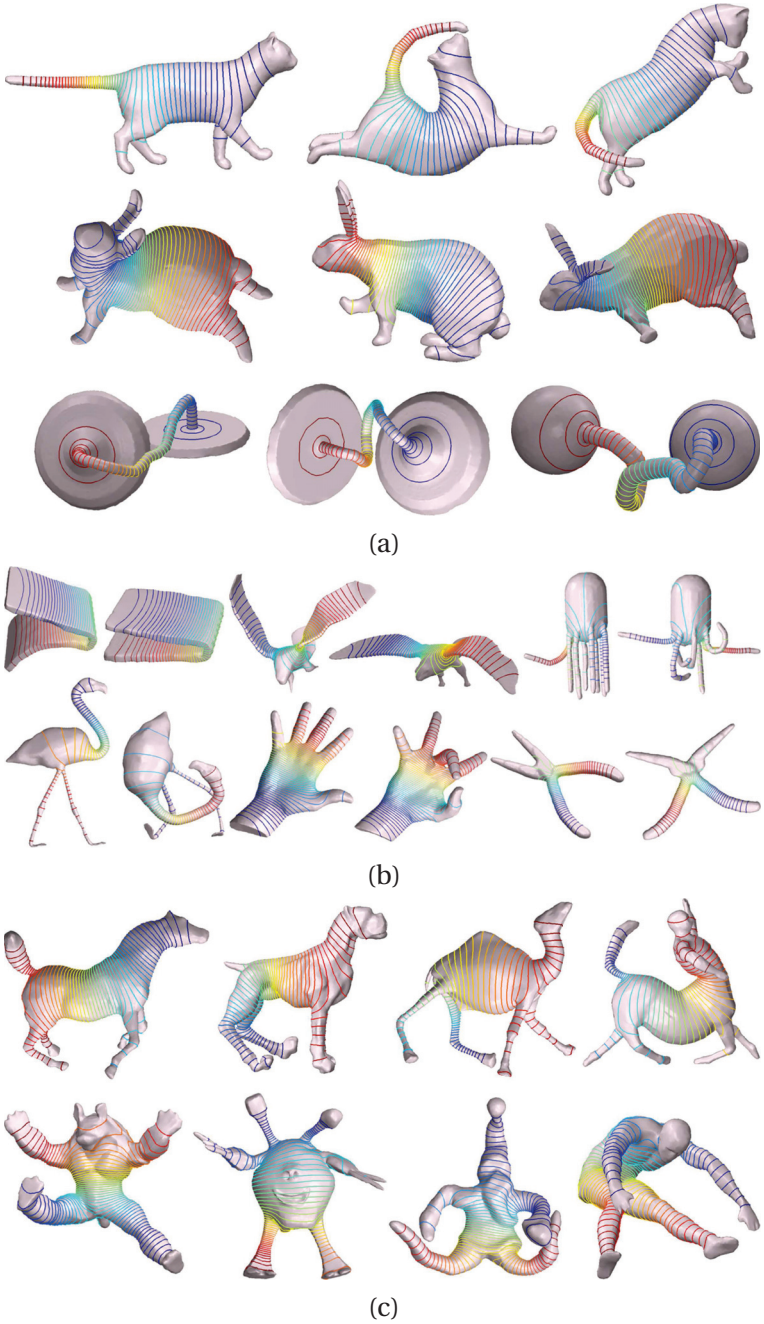


FIGURE 2.2: (a) Isocontours are invariant under both global and local deformation. (b) Proportionality correspondence of pairwise nonrigid shapes with varied topological structure. (c) Isocontours are consistent among different classes of shapes.

The level sets of the second eigenfunction have been used to extract curve skeletons of the nonrigid shapes [10, 36], which is a vivid clue that these isocontours capture the global topological structure of shapes. We visualize the procedure for extracting the curve skeleton in Figure 2.3.

(a) (b) (c)

FIGURE 2.3: (a) 3D horse model colored by φ_2; (b) level sets of φ_2; (c) spectral Reeb graph.

2.4.3 MATCHING BY INTRINSIC SPATIAL PARTITION

Instead of representing the whole shape by the codeword model without considering spatial layout of local descriptors, we enhance the discrimination by integrating the distribution of local descriptors in different spatial patches determined by the intrinsic spatial partition. For any shape cut by isocontours at resolution R, its description H is the concatenation of R sub-histograms:

$$H = [h^1, h^2, \ldots, h^i, \ldots, h^{\mathrm{R}}] \tag{2.19}$$

where h^i is the sub-histogram ordered in the ith position according to the intrinsic spatial partition from one end to the other. Note that the isocontour sequence might start from either end, and the situations are different from shape to shape. For example, in Figure 2.2(a), the heads of the first and third rabbit are colored in blue, but tail of the second is colored in red and head in blue, whose order is exactly the opposite. To guarantee that the semantic correspondent parts are matched in the comparison, we use an order-insensitive strategy comparison method. First, we get a new histogram T by making the order of the sub-histogram inverted in H:

$$T = [h^{\mathrm{R}}, h^{(\mathrm{R}-1)}, \cdots, h^i, \cdots, h^1]. \tag{2.20}$$

Secondly, to compare two shapes P and Q we define their dissimilarity under this feature as follows:

$$\mathscr{B}^{\mathrm{R}}(P,Q) = \min(\mathscr{A}^{\mathrm{R}}(H_P, H_Q), \mathscr{A}^{\mathrm{R}}(H_P, T_Q)) \tag{2.21}$$

where H_P and H_Q denote the histograms of P and Q, respectively. In other words, there are two possible matching schemes between the isocontour sequences of two shapes, head-to-head and head-to-end. We consider the schemes with the minimum cost to be better matched. For each scheme, the dissimilarity

measure $\mathscr{A}^{\mathtt{R}}(\cdot,\cdot)$ is defined as

$$\mathscr{A}^{\mathtt{R}}(H_P, H_Q) = \sum_{i=1}^{\mathtt{R}} \sum_{k=1}^{K} \Psi(h_P^i(k), h_Q^i(k)) \tag{2.22}$$

where $\Psi(\cdot,\cdot)$ can be any histogram comparison metric. In this paper we use the chi-squared kernel so that $h_P^i(k)$ and $h_Q^i(k)$ are the accumulations of the code of the local descriptors from P and Q that fall into the kth codeword cell/channel of the ith patch.

The degree of resolution would affect the performance of the spatial partition based method. To further improve the results, we proposed Intrinsic Spatial Pyramid Matching (ISPM) by extending the work in [35], which has been shown to yield excellent performance in image analysis, to nonrigid 3D shapes. The spatial pyramid divides an image into a multi-level pyramid of increasingly fine subregions and computes a codebook descriptor for each subregion. We construct a sequence of histograms at resolutions $\{\mathtt{R} = 2^\ell, \ell = 0,\ldots,L\}$ such that the surface at level ℓ has 2^ℓ patches, for a total of $2^L - 1$ patches. Thus, the final dissimilarity between the histograms of P and Q is given by

$$\begin{aligned} \mathscr{D}^L(P,Q) &= \mathscr{B}^L(P,Q) + \sum_{\ell=0}^{L-1} \frac{1}{2^{L-\ell}} (\mathscr{B}^\ell(P,Q) - \mathscr{B}^{\ell+1}(P,Q)) \\ &= \frac{1}{2^L}\mathscr{B}^0(P,Q) + \sum_{\ell=1}^{L} \frac{1}{2^{L-\ell+1}} \mathscr{B}^\ell(P,Q) \end{aligned} \tag{2.23}$$

Concerning the implementation, one issue that arises is that of normalization. To easily compare the methods of single level partition and intrinsic spatial pyramid matching, we normalize the histogram of each resolution using the L_1-norm.

2.5 Experiments

We experimentally compare different spectral descriptors and the codeword ambiguity modeling approaches on two standard datasets: SHREC 2010 [80] and SHREC 2011 [81]. We also show that the proposed intrinsic partition approach can significantly improve the performance of the spectral shape retrieval methods. We start our experiments with an in-depth analysis of the methods on a set of ten 3D nonrigid shape categories, after which we translate these findings into the experiments on a large dataset. In our experimental setup we closely follow the original works theoretically, and we select the optimal parameters that yield the best performance on these datasets. Thus, we do not bias any method in order to provide a fair comparison.

2.5.1 SETTINGS

Evaluation Measure: We evaluate the retrieval performance using the Discounted Cumulative Gain (DCG) [79]. DCG is a statistic that weights correct results near the front of the list more than correct results later in the ranked list, under the assumption that a user is less likely to consider elements near the end of the list. For a given query shape, DCG is defined as

$$DCG = \frac{DCG_N}{1 + \sum_{k=2}^{\tau} \frac{1}{\log_2(k)}} \tag{2.24}$$

where N is the total number of shapes in the database, τ is the number of shapes in the query's class (i.e. size of the class), and DCG_i is computed recursively as follows:

$$DCG_i = \begin{cases} G_i & \text{if } i = 1 \\ DCG_{i-1} + \frac{G_i}{\log_2(i)} & \text{Otherwise.} \end{cases} \tag{2.25}$$

where G_i is a gain value that depends on the relevance of the ith retrieved shape (1 if the shape that is the ith closest match to the query shape in the query shape's class, and 0 otherwise). Basically, DCG_i represents the relevance of the top-i results. Thus, the DCG score reflects the performance of the algorithm when correct results that are retrieved earlier are weighted higher than those retrieved later. All normalized DCG calculations are relative values in the interval $[0,1]$, and higher values imply better retrieval performance.

To gain further insight into the strengths and weaknesses of each descriptor, we compared the shape signatures in terms of four different properties: discriminative power, compactness, localization, and ambiguity. Discriminative power describes the signature's ability to distinguish between shapes belonging to different classes. Since we used watertight meshes from different classes in our experiments, the discounted cumulative gain (DCG) value reflects the discriminative power of the descriptors. Compactness refers to the dimension that a signature has to achieve for its maximum discriminative power. A lower dimension indicates higher compactness, and leads to simpler computation. We introduce localization as the ability of a signature to capture the local geometry around the mesh vertex. The DCG improvement after spatial partition reflects localization, since the partition separates the signatures into different patches. The finer is the partition, the more details can be captured if the DCG value further improves. To describe the distribution of the signature in the feature space, we observe the clustering centroid of the signature. The higher is the ambiguity between geometric words, the more kernel uncertainty is beyond the L_0-norm codebook. In Table 2.2, we rank the shape signatures in terms of different properties on a scale from 1 to 5. A value of 1 means the strongest, while a value of 5 means the weakest.

TABLE 2.2: Summary comparison of spectral signatures.

Property	SIHKS	HKS	WKS	HMS	GPS
Discriminative Power	1	2	5	3	4
Compactness	5	1	2	3	4
Localization	4	3	1	2	5
Ambiguity	3	5	2	4	1

Descriptors: To compare the aforementioned spectral descriptors, we design a reasonable strategy to choose the best parameters for each descriptor in our experiments. As pointed out in [34], the overall performance of the dense descriptor computed for each point of the shape is superior than a sparse descriptor computed for a set of points produced by a feature detection algorithm. So, we compute all spectral descriptors on every vertex of the shape. Additionally, due to the different ways of discretizing the LB operator, our implementation might generate slightly different results from the ones reported in [34]. However, for fair comparison, we use the same implementation for all the spectral methods in order to not favor any method.

Kernel Size: We choose the best kernel size according to the predefined range underlying the data, from *Local Kernel Size* (σ_{Local}) to *Global Kernel Size* (σ_{Global}). For each cluster, the distances from its centroid to other points of this cluster are computed. Then, the median absolute deviation is adopted to obtain the kernel size for each cluster. We use the median value as the local kernel size σ_{Local}. Since there is no explicit centroid for all the data points, we compute global "ghost centroid" in the same way that the centroid of each cluster is obtained during clustering. Then, the global kernel size σ_{Global} is estimated by treating the data space as a single cluster. σ_{Local} and σ_{Global} can help us roughly estimate the distribution of the data. Recall that our goal is to find the best smoothing parameters in order to maximize the discrimination, not to fit the data best, we use them as the size markers of the data space, and get the scale $S = \sigma_{Global}/\sigma_{Local}$. Multiplying by an augmenting coefficient A, we set an interval $[\sigma_{Local}/(AS), \sigma_{Local}AS]$ as the candidate space. With a sampling step $F_A \propto A$ to discretize the space, we select the best $\hat{\sigma}$ as the final kernel size. Note that in order to be consistent with the clustering stage, when estimating the kernel size, we use the L_1-norm for the dictionary learned by the cityblock distance, and each centroid is the component-wise median of the points in that cluster. We use the L_2-norm for the one learned by the Euclidean distance, and each centroid is computed as the mean of the points in that cluster. In Table 2.3, the parameters of different spectral descriptors are set for each dataset to get

the interval of the kernel size.

TABLE 2.3: Parameters' setting for kernel size interval estimation on SHREC 2010 and SHREC 2011 datasets.

	SHREC 2010					SHREC 2011				
	HKS	SIHKS	HMS	WKS	GPS	HKS	SIHKS	HMS	WKS	GPS
A	1	4	2	4	4	1	4	2	20	20
F_A	A	A	A	1	1	A	A	A	A/4	A/2

Complexity: The codes were implemented in MATLAB 7.14 (R2012a). The experiments were performed on a desktop computer with an Intel Core i3-2100 running at 3.1 GHz and 4 GB RAM. The computation of the vocabulary is performed offline in advance. It depends on the number of the descriptors (number of vertices), the dimension of the descriptor, and the vocabulary size (the number of clusters). Since we simplify our mesh to 2000 faces for each shape, we have a set of approximately 2×10^5 descriptors. Since the 3D models used in our experiments are watertight, simplifying a surface mesh by reducing its number of faces to 2000 would essentially preserve the shape semantically. Consequently, the discriminative power of the descriptors would not change drastically after mesh simplification.

To confirm getting optimal results, the clustering is repeated 3 times, and each by a new set of initial cluster centroid positions. The solution with the lowest value for the sum of distances is returned. In Table 2.4, we list the runtime in seconds for various descriptor dimensions and vocabulary sizes.

TABLE 2.4: Runtime (in seconds) with different descriptor dimensions and vocabulary sizes.

Runtime	Vocabulary Size							
Dimension	8	12	16	24	32	48	64	80
40	190	254	321	474	567	732	903	1351
100	522	617	774	1094	1461	1803	2054	2665
150	825	909	1193	1691	2002	2902	3645	4358
385	1725	2566	3347	4638	5702	8405	12285	15962

2.5.2 RESULTS ON SHREC 2010 DATASET

In-Depth Analysis of Descriptors and Ambiguity

The first dataset we consider is SHREC 2010 [80], which is a standard dataset of nonrigid 3D models used in the Shape Retrieval Contest, organized by National Institute of the Standards and Technology (NIST). The dataset consists of 200 shapes spread over 10 categories with 20 shapes each, and range from human body to man-made tools like glasses. Some of the deformations performed on the shapes are artificially generated, which would result in misleading recognition. In Figure 2.4 we show 4 models of each class in this dataset.

FIGURE 2.4: Sample shapes in SHREC 2010 dataset.

For the SHREC 2010 dataset, we analyze the types of spectral descriptors, vocabulary size and codeword ambiguity. The vocabulary sizes we consider are $8, 12, 16, 24, 32, 48, 64$ and 80. To gain further insight into the performance variation between the various types of spectral descriptors, we show the retrieval results for different parameters in Tables 2.14-2.17. The L_2-norm traditional codebook is used in all the descriptors. The main goal of our parameter determination is to select the appropriate heat diffusion time for each descriptor in order to maximally discriminate between the shapes. For HKS, we formulate the diffusion time as $t = t_0\alpha^\tau$, where τ is sampled from 0 to a given *scale* with a resolution $1/4$ in our case. We highlight the best result for each factor t_0, and notice that the largest diffusion times are

$t_{max} = t_0\alpha^{\max\tau} = \{343.44, 131.84, 74.08, 238.15, 139.57\}$. These times are obviously different from the best parameters $t = \{1024, 1351, 1783, 2353, 3104, 4096\}$ used in [34]. Since we use a different dataset, the difference of best parameters is reasonable. If the confusion time is larger than t_{max}, i.e. the heat diffuses for enough long time, then the heat distribution of the whole shape will be very similar, which tends to degrade the discrimination. On the contrary, if the confusion time is smaller than t_{max}, then only local patches of the shape are considered in the description, thus the global description of the whole shape is deficient. In order to construct the SIHKS, we use $t = \alpha^\tau$, where τ ranges from 1 to a given scale with finer increments of 1/16. After applying the logarithm, derivative, and Fourier transform, the first several discrete lowest frequencies are used as the local descriptor. In [34], the first 6 lowest frequencies are adopted, which yield a satisfactory result on SHREC 2010 dataset as shown in Table 2.15. To guarantee not to favor any method in our experiments, we test with various frequencies and find that the best result can always be obtained when all the frequencies are used. One-dimensional HMS is used to construct the temperature distribution descriptor [17]. For mesh segmentation, HMS prefers small time because high resolution details are preserved at a small scale. Our goal is to discriminate between shapes, so we construct a multi-scale HMS to compare the temperature distribution with multiple diffusion times. By defining a universal time unit δ as in the computation of the TD descriptor, we use $t = \tau\delta$ with τ ranging from 1 to a given scale to compute the descriptor. In Table 2.16, $\delta = 1$ and $t = [\delta, 2\delta, \cdots, 100\delta]$ give the best result for the multi-scale HMS. The WKS is associated with each vertex on the surface, and it describes the energy distributions over a family of fixed energy scales. There are two parameters in this descriptor, namely the increment δ and the variance σ. Since the energy range $[\lambda_{\min}, \lambda_{\max}]$ is determined by the eigenvalues and it is fixed, we seek the best increment by dividing the interval by a scalar M. The variance σ is also inferred as certain percentage of the interval. As shown in Table 2.17, the best parameters are $M = 50$ and percentage equal to 0.2. Finally, the best parameters for each spectral descriptor are summarized in Table 2.18 in the Appendix. SIHKS performs the best, and is slightly better than HKS since the variance of the scale transformation is small in SHREC 2010. As expected, the worst performance is obtained by WKS, which characterizes only local information of the small patch. We also considered a 200 vocabulary size for each signature, and the results remain virtually unchanged or go down slightly compared to a smaller vocabulary size.

In Table 2.5, we list the DCG values for different spectral signatures and classes on SHREC 2010. It should be noted that all of the signatures are good at retrieving the Teddy model. SIHKS yields the highest accuracy on most of the classes, such as Ant, Crab, Hand, Human, Octopus, Plier, Spectacle, and Spider. HKS is slightly lower than SIHKS. Surprisingly, HMS is the best one in retrieving the Teddy shape. Also,

the comparison between WKS, HMS and GPS is inspiring. WKS performs well on Octopus, Snake and Spectacle. These classes have a high percentage of thin-branch-like components. GPS gets a relatively high accuracy on Ant, Hand, and Human.

TABLE 2.5: DCG values for different spectral signatures and classes on SHREC 2010.

Signature	Ant	Crab	Hand	Human	Octopus	Plier	Snake	Spectacle	Spider	Teddy
SIHKS	0.951	0.901	0.897	0.884	0.783	0.935	0.714	0.773	0.899	0.990
HKS	0.925	0.877	0.804	0.805	0.726	0.979	0.707	0.732	0.886	0.990
WKS	0.793	0.813	0.652	0.704	0.731	0.784	0.713	0.744	0.723	0.913
HMS	0.668	0.785	0.709	0.687	0.646	0.880	0.696	0.633	0.731	0.998
GPS	0.846	0.675	0.713	0.744	0.576	0.881	0.622	0.675	0.727	0.821

To study the influence of the vocabulary size, we use various local descriptors and change the vocabulary size from 8 to 80 geometric words. Tables 2.7-2.10 show the resulting performance. As can be observed, the overall performance improves with the increase of the vocabulary size, but at the expense of the representation size (length of the BoF vector). However, for HKS and GPS, the best performance is obtained when the vocabulary size is 32 and 12, respectively.

Tables 2.6-2.10 also show the retrieval results with the various types of codeword ambiguity. The results show that codeword uncertainty consistently outperforms other types of ambiguity for all kinds of descriptors and all vocabulary sizes. Besides the 4 types of ambiguity modeling methods in [55], we also include the accumulation of L_1 and L_2 norms from descriptor to codeword. The reason is two-fold: (1) To keep coherent with the clustering stage, and (2) VLAD, which is the accumulation of vector from descriptor to codeword, achieves better results as reported in the literature. Indeed, in our experiment, SIHKS, HMS and WKS with distance accumulation outperform the traditional codebook, and even are superior than the codebook uncertainty in certain cases. In addition, the L_1 and L_2-norms in the clustering stage have a considerable effect on the performance. For GPS, the L_1-norm based methods are much worse than L_2-norm based ones. But for WKS, the L_1-norm based method is better than L_2-norm based one. As a result, we may conclude that the WKS descriptor is more discriminative than the L_1-norm.

To examine the influence of the kernel size, we show the kernel size found as the statistic among the five spectral descriptors. Following the technique of optimal kernel size estimation described in the experimental setting (Subsection 3.5.1), σ_{Local} is calculated first. Then, the optimal kernel size $\hat{\sigma}$ is determined experimentally inside the interval for each vocabulary size. To better visualize the results, we obtain the relative $\hat{\sigma}$ as the relative position in the interval. Thus, the relative σ_{Local} is 0.5. Our goal is

to compare $\hat{\sigma}$ for different ambiguity types in both Laplacian and Gaussian kernels. So the five number summary is computed on the set composed of the relative $\hat{\sigma}$ of five descriptors in the same vocabulary setting, and the results are shown in Figure 2.5. For both Laplacian and Gaussian kernels, the codeword plausibility has the largest kernel size, followed by the kernel codebook; the smallest kernel size is held by the codeword uncertainty. The Laplacian kernel has a larger kernel size than the Gaussian kernel for each type of ambiguity methods. In our experiment, we also found that increasing the kernel size of the codeword plausibility beyond a sufficiently large value does not significantly change the accuracy. The kernel size of the Laplacian codeword plausibility remains the largest in the interval, but the kernel size of the Gaussian codeword plausibility among different descriptors oscillates the most. Note that since we plot the relative position of $\hat{\sigma}$ compared to σ_{Local}, the value of σ_{Local} becomes smaller over the number of vocabulary elements. In fact, it shows that a larger vocabulary leads to slightly smaller kernels. This phenomenon is consistent with the image domain [55]. This may be expected since a larger vocabulary is formed by a smaller radius between codewords.

TABLE 2.6: Performance (DCG) using different codebook models of varying size based on SIHKS local descriptor.

Codebook Model		**Vocabulary Size**								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional Codebook	0.801	0.789	0.777	0.779	0.798	0.795	0.802	0.795	0.793
	L_1-norm Codebook	0.801	0.796	0.784	0.789	0.806	0.801	0.813	0.806	0.803
	L_2-norm Codebook	0.776	0.768	0.761	0.770	0.791	0.785	0.804	0.792	0.792
	Laplace Kernel Codebook	0.797	0.795	0.796	0.796	0.798	0.801	0.803	0.810	0.812
	Laplace Codeword Uncertainty	0.809	0.803	0.805	0.802	0.808	0.810	0.811	0.812	0.806
	Laplace Codeword Plausibility	0.798	0.782	0.774	0.774	0.792	0.792	0.798	0.793	0.783
L_2	Traditional Codebook	0.856	0.863	0.861	0.862	0.872	0.865	0.865	0.866	0.849
	L_1-norm Codebook	0.839	0.852	0.853	0.855	0.865	0.865	0.863	0.865	0.854
	L_2-norm Codebook	0.841	0.846	0.852	0.849	0.854	0.861	0.858	0.861	0.846
	Gaussian Kernel Codebook	0.847	0.839	0.840	0.842	0.850	0.846	0.845	0.847	0.831
	Gaussian Codeword Uncertainty	0.857	0.863	0.864	0.867	0.874	0.867	0.869	0.867	0.827
	Gaussian Codeword Plausibility	0.856	0.863	0.861	0.863	0.872	0.865	0.866	0.866	0.841

Improvement with Intrinsic Partition

First, we examine the effect of integrating spatial cues on surfaces via the intrinsic partition. Figure 2.6 shows the performance improvement of the retrieval experiments by matching shapes directly using intrinsic partitions on SHREC 2010 dataset. With the increase of the number of intrinsic partitions, all the

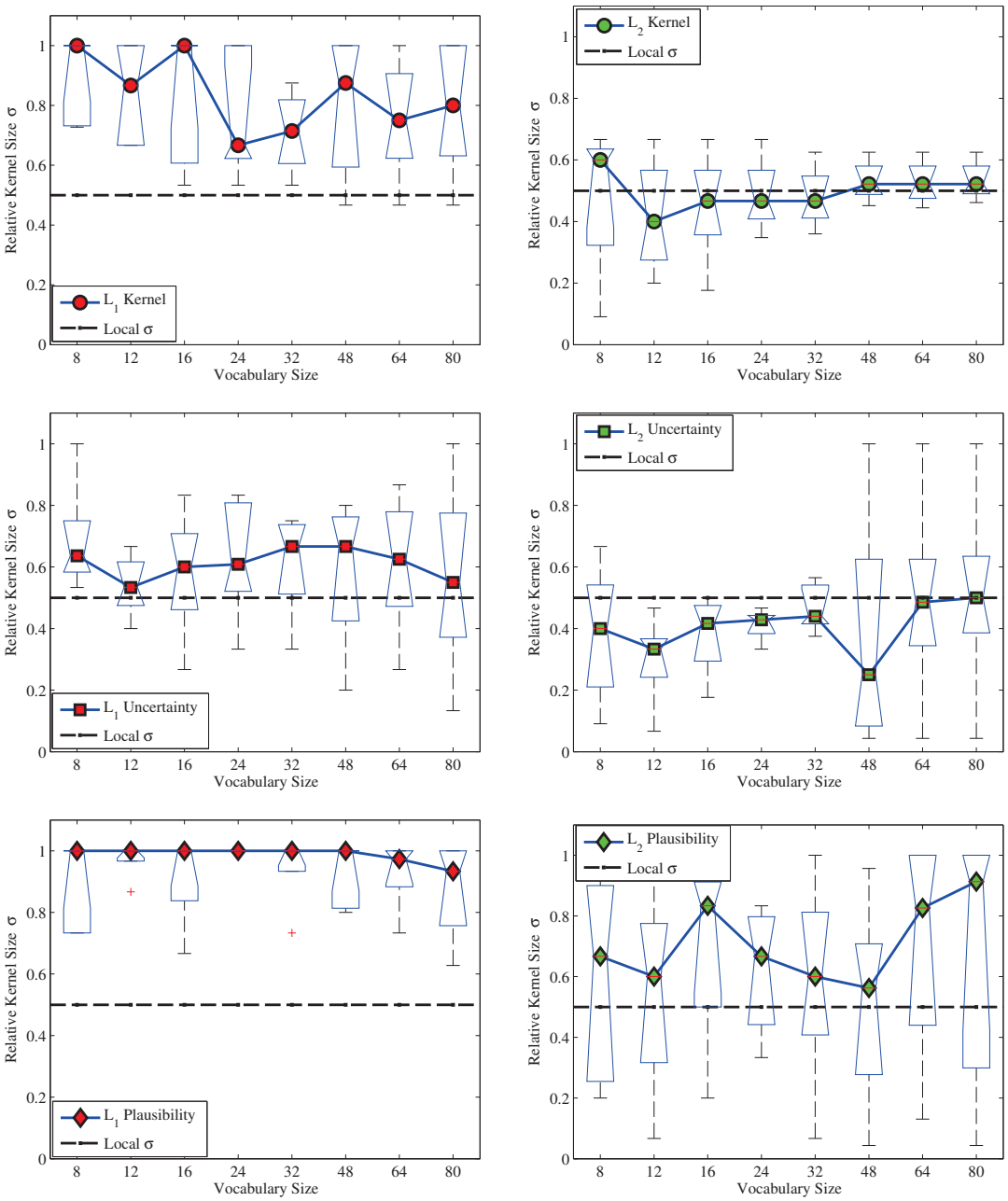


FIGURE 2.5: Illustration of relative kernel size of different ambiguity modeling methods. Top: kernel codebook. Middle: codeword uncertainty. Bottom: codeword plausibility.

spectral descriptors, except GPS, are improved substantially in a global fashion, even though the performance drops down in certain numbers. For ShapeGoogle [34], its performance is plotted as points whose partition number is one. Obviously, intrinsic spatial cues on shape surface proposed in our framework significantly outperform ShapeGoogle. We conjecture that GPS is degraded because of its global nature.

TABLE 2.7: Performance (DCG) using different codebook models of varying size based on HKS local descriptor.

Codebook Model		Vocabulary Size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional Codebook	0.827	0.841	0.852	0.846	0.854	0.836	0.828	0.837	0.834
	L_1-norm Codebook	0.812	0.825	0.839	0.808	0.848	0.794	0.776	0.823	0.813
	L_2-norm Codebook	0.784	0.793	0.812	0.765	0.808	0.746	0.726	0.774	0.755
	Laplace Kernel Codebook	0.820	0.832	0.840	0.852	0.841	0.851	0.847	0.845	0.831
	Laplace Codeword Uncertainty	0.829	0.843	0.853	0.849	0.857	0.840	0.839	0.848	0.834
	Laplace Codeword Plausibility	0.829	0.840	0.849	0.851	0.852	0.844	0.843	0.836	0.834
L_2	Traditional Codebook	0.826	0.851	0.843	0.845	0.845	0.834	0.831	0.828	0.834
	L_1-norm Codebook	0.804	0.838	0.823	0.837	0.825	0.791	0.806	0.785	0.820
	L_2-norm Codebook	0.776	0.807	0.788	0.813	0.794	0.748	0.770	0.741	0.789
	Gaussian Kernel Codebook	0.826	0.846	0.847	0.845	0.851	0.852	0.848	0.849	0.829
	Gaussian Codeword Uncertainty	0.828	0.845	0.845	0.852	0.851	0.836	0.834	0.833	0.829
	Gaussian Codeword Plausibility	0.831	0.852	0.846	0.845	0.846	0.842	0.837	0.840	0.845

TABLE 2.8: Performance (DCG) using different codebook models of varying size based on HMS local descriptor.

Codebook Model		Vocabulary Size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional Codebook	0.683	0.717	0.729	0.749	0.747	0.733	0.745	0.746	0.725
	L_1-norm Codebook	0.693	0.721	0.752	0.760	0.759	0.711	0.766	0.751	0.723
	L_2-norm Codebook	0.680	0.695	0.695	0.698	0.702	0.675	0.726	0.728	0.673
	Laplace Kernel Codebook	0.689	0.705	0.711	0.720	0.718	0.726	0.719	0.724	0.740
	Laplace Codeword Uncertainty	0.691	0.722	0.736	0.751	0.754	0.735	0.755	0.745	0.741
	Laplace Codeword Plausibility	0.680	0.709	0.719	0.738	0.740	0.733	0.734	0.740	0.724
L_2	Traditional Codebook	0.696	0.726	0.737	0.746	0.743	0.743	0.749	0.753	0.748
	L_1-norm Codebook	0.713	0.743	0.763	0.766	0.771	0.740	0.773	0.780	0.740
	L_2-norm Codebook	0.712	0.730	0.752	0.763	0.761	0.718	0.765	0.759	0.727
	Gaussian Kernel Codebook	0.699	0.721	0.731	0.732	0.732	0.733	0.732	0.724	0.722
	Gaussian Codeword Uncertainty	0.707	0.740	0.759	0.768	0.771	0.753	0.767	0.771	0.733
	Gaussian Codeword Plausibility	0.696	0.726	0.737	0.746	0.743	0.744	0.749	0.753	0.748

A crucial parameter is the number of partitions. Experimentally, we find that the accuracy remains stable after 16 partitions for all the spectral descriptors.

Next, let us examine the behavior of intrinsic spatial pyramid matching. For completeness, Table 3.1 lists the performance achieved using just the highest level of the pyramid (the "single" columns) as well

TABLE 2.9: Performance (DCG) using different codebook models of varying size based on WKS local descriptor.

Codebook Model		Vocabulary Size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional Codebook	0.691	0.711	0.720	0.727	0.726	0.741	0.737	0.737	0.748
	L_1-norm Codebook	0.697	0.718	0.726	0.733	0.730	0.742	0.737	0.737	0.751
	L_2-norm Codebook	0.694	0.711	0.712	0.721	0.719	0.731	0.725	0.725	0.733
	Laplace Kernel Codebook	0.694	0.701	0.707	0.713	0.714	0.725	0.724	0.726	0.738
	Laplace Codeword Uncertainty	0.697	0.716	0.721	0.725	0.727	0.740	0.736	0.737	0.741
	Laplace Codeword Plausibility	0.695	0.706	0.718	0.724	0.723	0.740	0.736	0.737	0.739
L_2	Traditional Codebook	0.659	0.689	0.703	0.718	0.728	0.722	0.716	0.718	0.743
	L_1-norm Codebook	0.666	0.704	0.710	0.724	0.730	0.728	0.721	0.722	0.738
	L_2-norm Codebook	0.672	0.701	0.704	0.716	0.727	0.727	0.720	0.719	0.731
	Gaussian Kernel Codebook	0.666	0.667	0.667	0.666	0.666	0.666	0.667	0.667	0.651
	Gaussian Codeword Uncertainty	0.667	0.666	0.665	0.666	0.665	0.666	0.667	0.667	0.651
	Gaussian Codeword Plausibility	0.662	0.689	0.703	0.718	0.728	0.722	0.716	0.718	0.736

as the performance of the complete matching scheme using multiple levels (the "pyramid" columns). For all kinds of features, except GPS, the results improve considerably as we go from $L = 1$ to a multi-level setup. We do not display the results for $L = 0$ because its highest single level is the same as with its pyramid. Although matching at the highest pyramid level seems to account for most of the improve-

TABLE 2.10: Performance (DCG) using different codebook models of varying size based on GPS local descriptor.

Codebook Model		Vocabulary Size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional Codebook	0.719	0.757	0.733	0.723	0.720	0.737	0.714	0.726	0.739
	L_1-norm Codebook	0.716	0.752	0.729	0.720	0.717	0.732	0.708	0.723	0.738
	L_2-norm Codebook	0.689	0.734	0.705	0.699	0.692	0.717	0.691	0.704	0.711
	Laplace Kernel Codebook	0.764	0.789	0.785	0.796	0.799	0.811	0.817	0.822	0.738
	Laplace Codeword Uncertainty	0.773	0.794	0.790	0.801	0.801	0.812	0.818	0.824	0.776
	Laplace Codeword Plausibility	0.724	0.759	0.740	0.726	0.725	0.737	0.718	0.731	0.696
L_2	Traditional Codebook	0.704	0.785	0.748	0.768	0.757	0.734	0.724	0.727	0.743
	L_1-norm Codebook	0.698	0.780	0.744	0.763	0.753	0.729	0.721	0.726	0.741
	L_2-norm Codebook	0.687	0.733	0.717	0.739	0.744	0.729	0.722	0.724	0.729
	Gaussian Kernel Codebook	0.793	0.808	0.775	0.804	0.785	0.790	0.791	0.783	0.759
	Gaussian Codeword Uncertainty	0.755	0.799	0.762	0.773	0.767	0.772	0.764	0.765	0.792
	Gaussian Codeword Plausibility	0.705	0.805	0.764	0.776	0.771	0.735	0.722	0.727	0.746

ment, using all the levels together yields a statistically significant benefit. For strong features, single-level performance actually drops as we go from $L = 3$ to $L = 4$. This means that the highest level of the $L = 3$ pyramid is too finely subdivided, with individual bins yielding few matches. Despite the diminished discriminative power of the highest level, the performance of the entire $L = 4$ pyramid remains essentially identical to that of the $L = 3$ pyramid. Thus, the main advantage of the intrinsic spatial pyramid representation stems from combining multiple resolutions in a principled fashion, and it is robust to failures at individual levels.

It is also important to compare the performance of different spectral descriptor sets. As expected, weak descriptors do not perform as well as strong descriptors, though in combination with the spatial pyramid, they can also achieve acceptable levels of accuracy. Note that only descriptors with a much higher density and much smaller spatial extent will continue to improve their performance as we increase L from 3 to 4. Such kinds of descriptors in Table 3.1 include SIHKS and WKS. In this respect, ISPM provides us with a way to analyze the spectral descriptors. On the other hand, the performances of HKS and HMS drop when the pyramid level increases from 3 to 4. Moreover, the performance of GPS decreases immediately when the descriptors are aggregated in terms of local patches instead of the whole shape. Increasing the visual vocabulary size from 8 to 80 might result in a small performance increase at $L = 0$, but this difference is eliminated at higher pyramid levels. Thus, we may conclude that the coarse-grained geometric cues provided by the pyramid have more discriminative power than an enlarged geometric vocabulary. Another explanation for the improvement is that the geometric cues eliminate the word ambiguity in a spatial context, and its ambiguity modeling ability is also more discriminative than the ambiguity codebook models. For example, HMS with L_2-norm, the traditional codeword, achieves 0.743 on SHREC 2010 dataset, and the improvement brought by the codeword uncertainty is $0.771 - 0.743 = 0.028$. However, the improvement brought by ISPM is $0.792 - 0.743 = 0.049$.

In Figure 2.7, we show two examples of top 9 retrieval results for different methods. There are plenty of examples to demonstrate that our proposed intrinsic spatial pyramid matching (ISPM) method improves the performance of the original codebook models with varied spectral descriptors. However, in order to illustrate the merits of different descriptors, we choose the two exemplar queries that bring consistent results with the conclusion we get for the spectral descriptors. For both queries, the ISPM method enhances the results by helping rank more relevant shapes higher as expected, and SIHKS combined with ISPM achieves a more satisfactory result. For the first query spectacle, the primitive descriptors always confuse it with lines, pliers, human body and octopus. This is because these objects also have several long, thin pipe-like parts and flat globular parts, and the proportions are similar. The spatial partition

integrates the intrinsic symmetry cue, and separates pipe-like parts and globular parts into different sub-histograms, thus resulting in a more descriptive representation. The pipe-like parts attached to fringe of shapes are almost similar among these classes, but the globular parts in the middle have different degree and amount of flat, which is the key to distinguish the spectacles. For the second query octopus, multiple legs objects like spiders and ants are usually retrieved mistakenly by the descriptors SIHKS, HKS and GPS. This is in fact understandable since even humans would consider these shapes as similar from their appearance. HMS ranks some hands higher, and WKS ranks some pliers higher. This phenomenon shows that the different spectral descriptors characterize different aspects of shape features.

It is worth pointing out that the intrinsic spatial partition helps measure the localization ability of the signatures, meaning the ability of a signature to capture the local geometry around the mesh vertex. The DCG improvement after applying spatial partition reflects localization due largely to the fact that the partition separates the signature into different patches. The finer is the partition, the more details can be captured if the DCG improves.

2.5.3 Results on SHREC 2011

Our second set of experiments is performed on a larger dataset, SHREC 2011 [81], which contains 600 watertight triangle meshes that are equally classified into 30 categories. SHREC 2011 is the most diverse nonrigid 3D shape database available today in terms of object classes and deformations. In Figure 2.8 we show 2 models of each class in the dataset.

We follow the same experimental procedure on SHREC 2010 dataset. First, we find the best performance of each spectral descriptor with its best parameters under the traditional codebook model, as shown in Table 2.12. Then, using these parameters we conduct further experiments on various ambiguity modeling models. In this way, we get the baseline result, which is actually the algorithm of ShapeGoogle [34]. Finally, our spatially aggregating approach is tested against each type of descriptors. In Figure 2.9, we display the performance improvement with the increasing number of partitions. For both L_1 and L_2-norms, all types of descriptors, except GPS, show improved accuracy when spatially aggregated. HKS and SIHKS are improved throughout all the number of partitions till 512 in our experiment. But we only show the results from 1 to 20 for the sake of visualization. HMS and WKS remain unchanged or become worse when 10 or more partitions are adopted. Note that an interesting phenomenon is that GPS with traditional and plausibility ambiguity has a lower accuracy than kernel and uncertainty ambiguity when no global spatial is integrated. However, the results are completely the opposite after a certain degree of partition is included. The traditional and plausibility ambiguity modeling methods go up, but

kernel and uncertainty ambiguity modeling methods go down, even lower than the former approaches.

Table 2.13 gives a breakdown of retrieval accuracies for different pyramid levels and different spectral descriptors with vocabulary size 32. On this large dataset, a major advantage of ISPM is shown to provide a robust and stable performance by weightily combining multiple spatial levels, especially, for HMS and WKS, and it consistently outperforms the single level approach. Although ISPM is not superior in all cases, it is, however, practical since determining the level of partition is much easier than determining the number of partitions to obtain satisfactory results.

In summary, our proposed shape retrieval approach is shown to outperform state-of-the-art orderless descriptor aggregating methods. Because of the geometric stability and lack of damage of shapes in SHREC 2011, dense descriptors combined with global spatial relations seem to capture more discriminative information about the objects by providing an approximate correspondence.

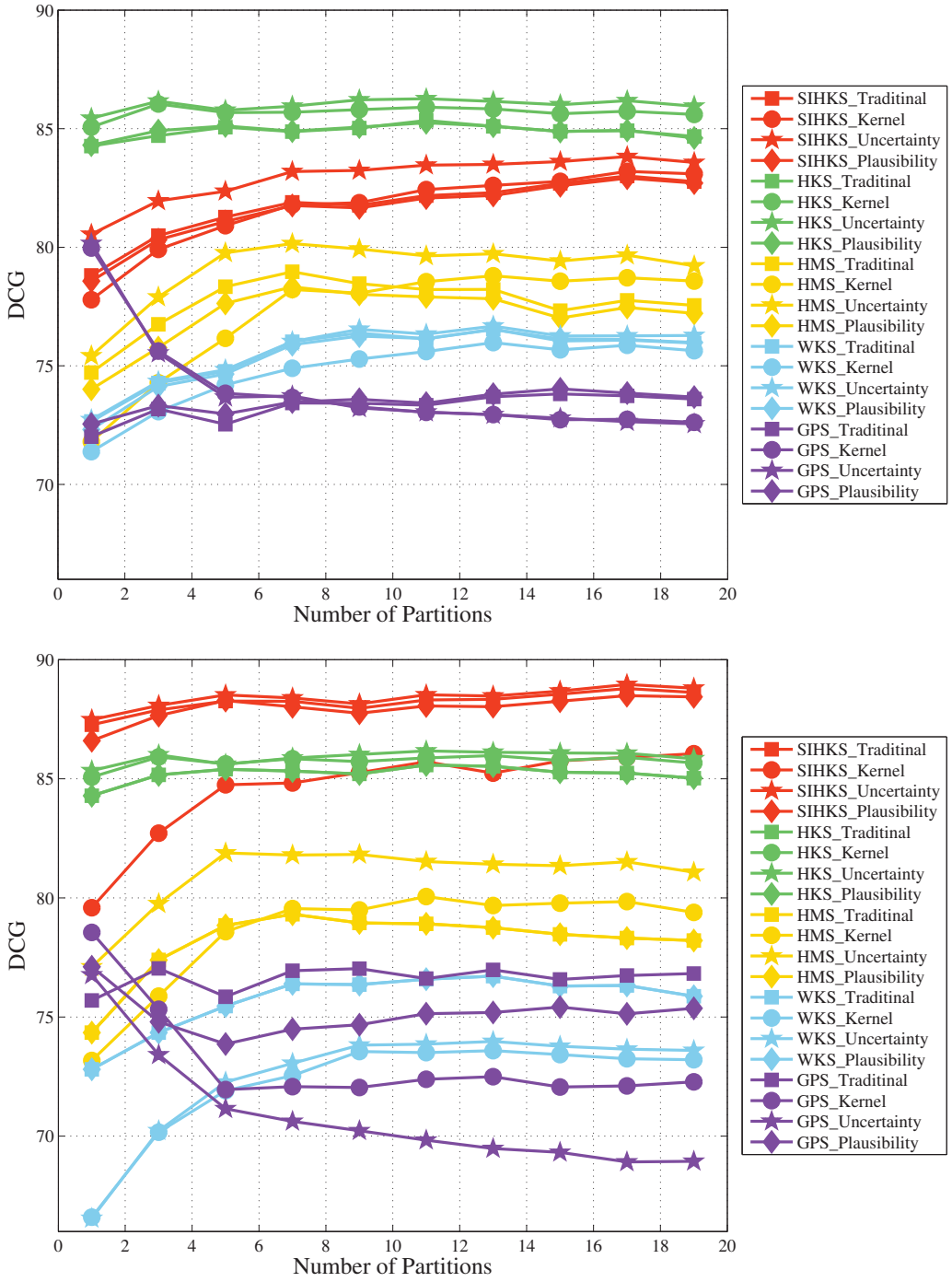
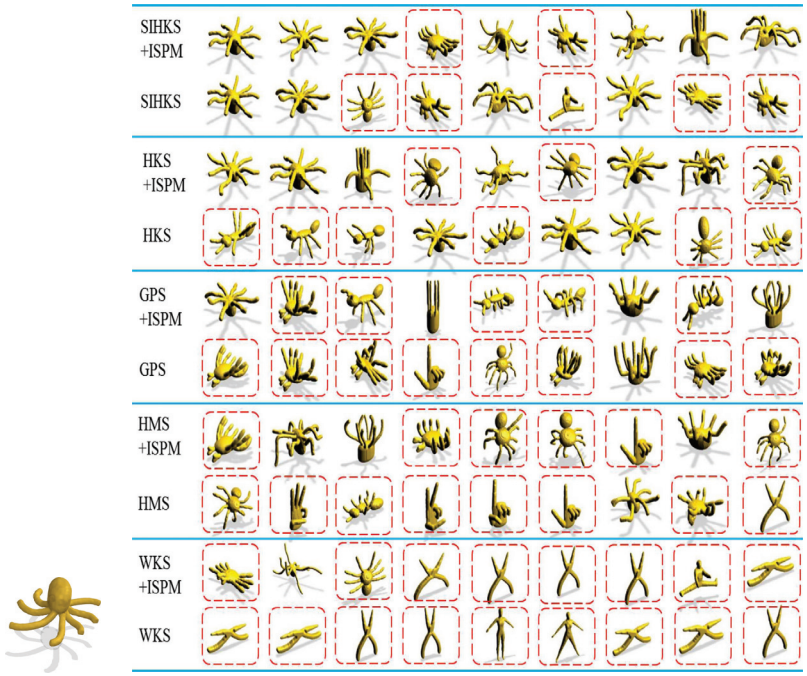


FIGURE 2.6: Performance improvement by increasing the number of intrinsic partitions on SHREC 2010 dataset. Top: L_1-norm. Bottom: L_2-norm.

TABLE 2.11: Performance (DCG) of different spectral descriptors on SHREC 2010 dataset using intrinsic spatial pyramid matching with various codebook models.

			Codebook Models							
Spectral	**Clustering**	**Level L**	Traditional		Kernel		Uncertainty		Plausibility	
Descriptor	**Norm**	**(Partitions)**	Single	Pyramid	Single	Pyramid	Single	Pyramid	Single	Pyramid
HKS	L_1	1 (2)	0.851	0.855	0.829	0.835	0.848	0.853	0.847	0.852
		2 (4)	0.856	0.863	0.843	0.849	0.858	0.862	0.855	0.860
		3 (8)	0.863	**0.867**	0.8554	**0.858**	0.865	0.867	0.861	**0.865**
		4 (16)	0.856	0.862	0.849	0.856	0.861	**0.867**	0.855	0.860
	L_2	1 (2)	0.837	0.842	0.845	0.849	0.845	0.849	0.847	0.842
		2 (4)	0.848	0.851	0.8616	0.866	0.859	0.860	0.850	0.853
		3 (8)	0.850	**0.853**	0.862	**0.866**	0.860	0.863	0.851	**0.855**
		4 (16)	0.847	0.852	0.854	0.862	0.859	**0.864**	0.847	0.852
SIHKS	L_1	1 (2)	0.799	0.798	0.7989	0.7992	0.8113	0.8107	0.7952	0.7951
		2 (4)	0.813	0.800	0.817	0.814	0.829	0.821	0.810	0.807
		3 (8)	0.818	0.810	0.823	0.821	0.829	0.828	0.815	0.812
		4 (16)	**0.831**	0.8262	**0.827**	0.825	**0.832**	0.8313	**0.830**	0.824
	L_2	1 (2)	0.872	0.873	0.853	0.854	0.874	0.876	0.873	0.874
		2 (4)	0.877	0.879	0.871	0.869	0.878	0.882	0.877	0.880
		3 (8)	0.879	0.881	0.872	0.874	0.878	0.8830	0.879	0.882
		4 (16)	0.883	**0.884**	0.876	**0.877**	0.884	**0.885**	0.883	**0.884**
HMS	L_1	1 (2)	0.748	0.747	0.722	0.725	0.756	0.759	0.739	0.743
		2 (4)	0.783	0.750	0.757	0.751	0.795	0.792	0.774	0.773
		3 (8)	0.787	0.781	0.778	0.778	0.802	**0.804**	0.780	**0.783**
		4 (16)	0.774	**0.790**	0.783	**0.786**	0.792	0.800	0.771	0.780
	L_2	1 (2)	0.749	0.752	0.737	0.740	0.776	0.780	0.749	0.752
		2 (4)	0.787	0.786	0.777	0.775	0.813	0.812	0.787	0.786
		3 (8)	0.793	**0.7971**	0.797	0.798	0.821	**0.824**	0.793	**0.797**
		4 (16)	0.783	0.792	0.795	**0.799**	0.811	0.820	0.783	0.792
WKS	L_1	1 (2)	0.725	0.726	0.711	0.711	0.725	0.727	0.724	0.725
		2 (4)	0.750	0.747	0.739	0.734	0.751	0.745	0.749	0.746
		3 (8)	0.759	0.757	0.748	0.745	**0.759**	0.756	0.758	0.757
		4 (16)	0.759	**0.760**	0.754	0.751	0.760	**0.761**	0.758	**0.760**
	L_2	1 (2)	0.728	0.729	0.671	0.670	0.672	0.670	0.728	0.729
		2 (4)	0.754	0.751	0.713	0.703	0.715	0.706	0.754	0.751
		3 (8)	0.762	0.761	0.713	0.721	0.731	0.725	0.762	0.761
		4 (16)	0.762	**0.763**	**0.732**	0.727	**0.735**	0.732	0.762	**0.763**
GPS	L_1	1 (2)	0.722	0.722	0.778	**0.789**	0.780	**0.790**	0.726	0.726
		2 (4)	0.717	0.720	0.735	0.747	0.735	0.745	0.717	0.721
		3 (8)	0.732	0.729	0.734	0.739	0.733	0.739	0.733	0.730
		4 (16)	**0.736**	0.734	0.726	0.731	0.726	0.731	0.737	**0.736**
	L_2	1 (2)	0.759	0.760	0.767	**0.778**	0.734	**0.745**	0.761	**0.772**
		2 (4)	0.759	0.762	0.723	0.744	0.722	0.729	0.728	0.753
		3 (8)	0.765	0.769	0.716	0.729	0.704	0.712	0.748	0.759
		4 (16)	0.768	**0.770**	0.720	0.725	0.691	0.697	0.751	0.758

(a) Query shape spectacle and its retrieval results



(b) Query shape octopus and its retrieval results

FIGURE 2.7: Retrieval results using different spectral descriptors and its spatial enhanced version. Error result is marked in the red dashed box. (a) In the left is the query shape spectacle, and the ten rows in the right its top 9 retrieval results; (b) In the left is the query shape octopus, and the ten rows in the right its top 9 retrieval results.

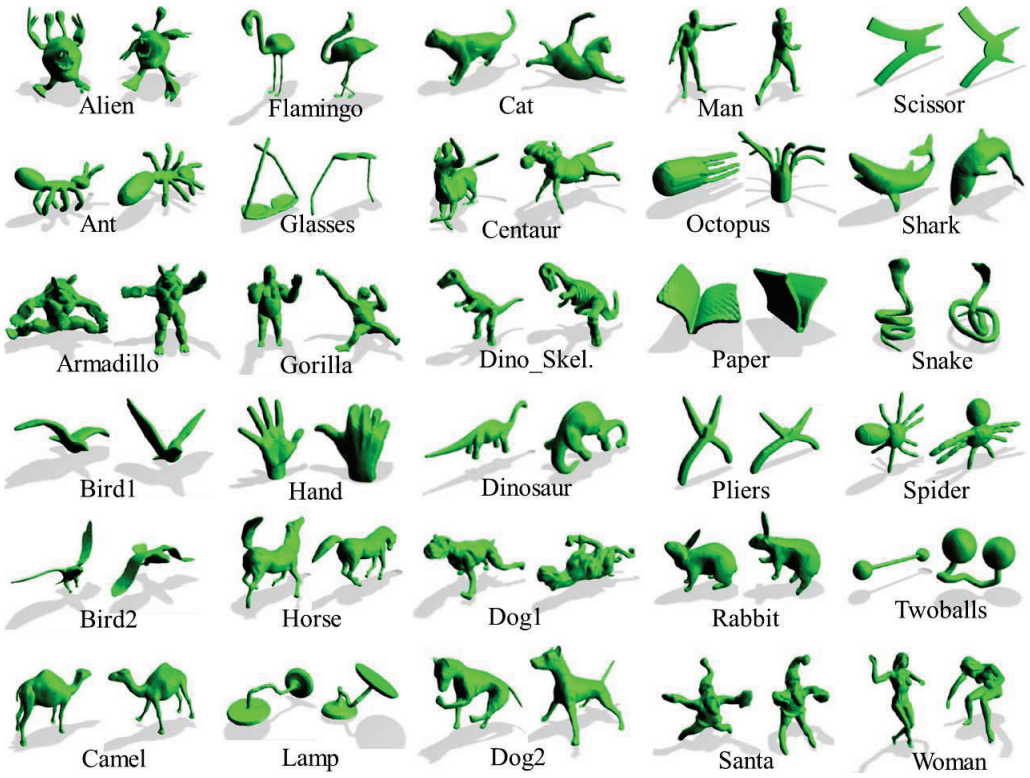


FIGURE 2.8: Sample shapes in SHREC 2011 dataset.

TABLE 2.12: Performance comparison of descriptors and their optimal parameters on SHREC 2011 dataset.

	Spectral Descriptors						
	HKS	SIHKS	HMS	WKS	GPS	ShapeDNA	EVD
Parameters	$\tau = 1/4$ $T = 5$ $t_0 = 0.01$ $\alpha = 4$	$\tau = 1/16$ $T = 25$ $F = 193$ $\alpha = 2$	$\tau = 2$ $T = 40$ $t_0 = 4$ $\alpha = 0.8$	$M = 100$ $\sigma = 0.05$			
DCG	0.811	0.826	0.773	0.680	0.709	0.782	0.560

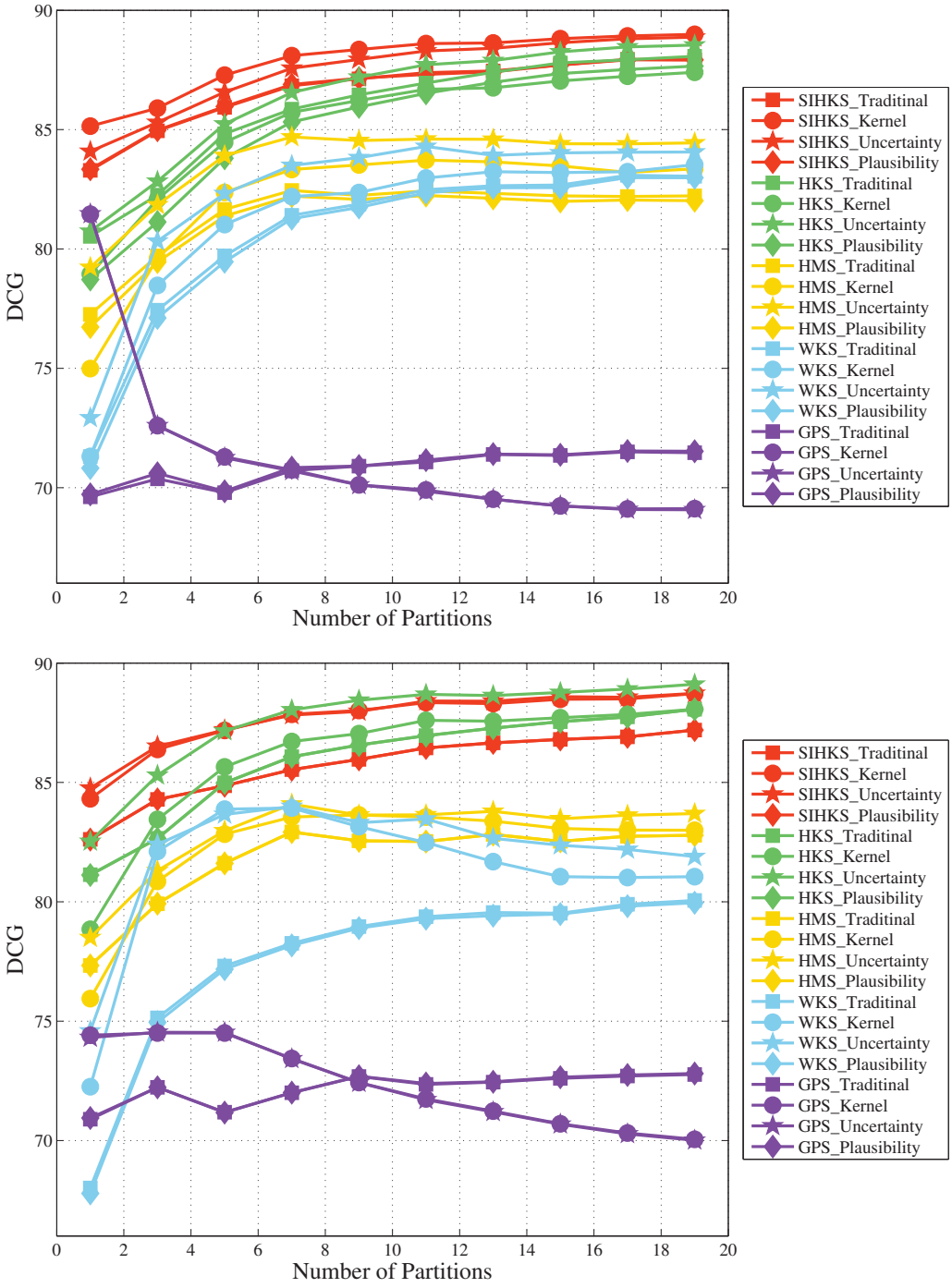


FIGURE 2.9: Performance improvement by increasing the number of intrinsic partitions on SHREC 2011 dataset. Top: L_1-norm. Bottom: L_2-norm.

TABLE 2.13: Performance (DCG) of different spectral descriptors on SHREC 2011 dataset using intrinsic spatial pyramid matching with varying codebook models.

			Codebook Models							
Spectral	**Clustering**	**Level L**	Traditional		Kernel		Uncertainty		Plausibility	
Descriptor	**Norm**	**(Partitions)**	Single	Pyramid	Single	Pyramid	Single	Pyramid	Single	Pyramid
HKS	L_1	1 (2)	0.829	0.822	0.811	0.806	0.831	0.824	0.808	0.802
		4 (16)	0.878	0.872	0.871	0.866	0.883	0.878	0.874	0.867
		7 (128)	0.890	0.888	0.880	0.879	0.892	0.891	0.887	0.885
		9 (512)	**0.893**	0.892	0.879	**0.880**	**0.894**	0.893	**0.890**	0.889
	L_2	1 (2)	0.834	0.827	0.812	0.807	0.849	0.843	0.834	0.827
		4 (16)	0.876	0.872	0.879	0.875	0.889	0.887	0.876	0.872
		7 (128)	0.889	0.888	0.878	**0.880**	0.889	**0.890**	0.889	0.887
		9 (512)	**0.891**	0.890	0.868	0.873	0.888	0.889	**0.891**	0.890
SIHKS	L_1	1 (2)	0.853	0.849	0.873	0.870	0.861	0.857	0.852	0.850
		4 (16)	0.878	0.876	0.888	0.888	0.887	0.884	0.878	0.876
		7 (128)	0.889	0.888	0.895	0.894	0.890	0.895	0.889	0.887
		9 (512)	**0.891**	0.890	**0.896**	0.896	**0.898**	0.897	**0.891**	0.890
	L_2	1 (2)	0.845	0.843	0.868	0.864	0.871	0.866	0.845	0.843
		4 (16)	0.869	0.867	0.885	0.885	0.886	0.885	0.869	0.867
		7 (128)	0.878	0.877	**0.889**	0.888	0.888	**0.889**	0.878	0.877
		9 (512)	**0.880**	0.879	0.888	0.888	0.889	0.889	**0.880**	0.879
HMS	L_1	1 (2)	0.780	0.784	0.759	0.764	0.801	0.804	0.777	0.781
		4 (16)	0.823	**0.828**	0.833	**0.837**	0.845	**0.850**	0.821	**0.826**
		7 (128)	0.815	0.819	0.819	0.814	0.831	0.835	0.814	0.818
		9 (512)	0.825	0.823	0.804	0.803	0.836	0.836	0.825	0.822
	L_2	1 (2)	0.782	0.786	0.777	0.779	0.794	0.798	0.782	0.786
		4 (16)	0.827	**0.837**	0.831	**0.838**	0.835	**0.842**	0.827	**0.832**
		7 (128)	0.821	0.824	0.806	0.814	0.823	0.828	0.821	0.825
		9 (512)	0.830	0.828	0.796	0.803	0.828	0.828	0.830	0.828
WKS	L_1	1 (2)	0.757	0.746	0.777	0.738	0.765	0.755	0.751	0.741
		4 (16)	0.829	0.824	**0.834**	0.828	**0.841**	0.838	0.828	0.723
		7 (128)	0.839	0.8382	0.824	0.828	0.826	0.829	0.838	0.837
		9 (512)	0.839	**0.839**	0.813	0.818	0.819	0.822	0.839	**0.839**
	L_2	1 (2)	0.718	0.709	0.761	0.754	0.781	0.776	0.678	0.707
		4 (16)	0.797	0.793	0.813	**0.823**	0.822	**0.830**	0.797	0.792
		7 (128)	0.804	**0.804**	0.779	0.789	0.787	0.796	0.803	**0.804**
		9 (512)	0.800	0.803	0.756	0.764	0.776	0.782	0.800	0.802
GPS	L_1	1 (2)	0.691	0.693	0.763	**0.777**	0.764	**0.776**	0.691	0.693
		4 (16)	0.712	0.711	0.692	0.700	0.691	0.670	0.712	0.712
		7 (128)	0.718	0.718	0.684	0.686	0.683	0.686	0.719	0.718
		9 (512)	**0.719**	0.719	0.681	0.683	0.680	0.682	**0.720**	0.719
	L_2	1 (2)	0.705	0.708	0.752	**0.759**	0.751	**0.757**	0.705	0.708
		4 (16)	0.726	0.725	0.705	0.716	0.705	0.715	0.726	0.725
		7 (128)	0.728	0.729	0.678	0.683	0.678	0.683	0.729	0.729
		9 (512)	**0.729**	0.729	0.671	0.675	0.670	0.674	**0.729**	0.729

TABLE 2.14: The DCG measure of HKS based on different pairs of parameters.

HKS		α				
Factor t_0	**Scale**	1.24	1.64	2	2.5	3
4.24	2	0.701	0.741	0.765	0.791	0.803
	4	0.723	0.778	0.802	0.830	**0.831**
	6	0.746	0.801	0.817	0.827	0.822
	8	0.756	0.805	0.823	0.822	0.817
	10	0.765	0.811	0.820	0.811	0.820
	12	0.775	0.805	0.811	0.812	0.815
8.24	2	0.752	0.781	0.812	0.835	0.827
	4	0.781	0.821	**0.840**	0.839	0.829
	6	0.780	0.831	0.837	0.827	0.823
	8	0.810	0.833	0.828	0.825	0.814
	10	0.814	0.832	0.831	0.819	0.817
	12	0.824	0.829	0.817	0.817	0.811
10.24	2	0.763	0.799	0.817	0.835	0.843
	4	0.793	**0.844**	0.833	0.838	0.829
	6	0.806	0.843	0.836	0.834	0.824
	8	0.823	0.840	0.837	0.825	0.815
	10	0.826	0.828	0.826	0.813	0.814
	12	0.836	0.835	0.816	0.816	0.810
12.24	2	0.772	0.807	0.821	0.830	0.829
	4	0.791	0.840	0.831	0.829	0.823
	6	0.817	**0.846**	0.832	0.822	0.815
	8	0.833	0.834	0.827	0.820	0.815
	10	0.842	0.835	0.830	0.814	0.819
	12	0.839	0.828	0.820	0.811	0.808
16.24	2	0.787	0.810	0.809	0.822	0.832
	4	0.805	0.828	0.835	0.819	0.821
	6	0.827	0.836	0.819	0.813	0.814
	8	0.844	0.825	0.819	0.810	0.803
	10	**0.848**	0.824	0.822	0.810	0.813
	12	0.835	0.820	0.813	0.811	0.799

TABLE 2.15: The DCG measure of SIHKS based on different pairs of parameters.

SIHKS		α				
Timescale	**Frequency**	1.24	1.64	2	2.5	3
20	6	0.795	0.842	0.867	0.853	0.846
	50	0.792	0.851	0.861	0.847	0.847
	100	0.794	0.847	0.869	0.852	0.845
	150	0.793	0.852	0.864	0.853	0.854
	193	0.795	0.852	0.866	0.848	0.855
25	6	0.774	0.867	0.872	0.854	0.849
	50	0.770	0.859	0.869	0.864	0.852
	100	0.769	0.861	0.871	0.862	0.851
	150	0.764	0.861	0.862	0.864	0.848
	193	0.757	0.861	**0.877**	0.862	0.845
30	6	0.811	0.860	0.859	0.857	0.850
	50	0.808	0.869	0.868	0.857	0.853
	100	0.819	0.872	0.865	0.863	0.856
	150	0.809	0.866	0.861	0.859	0.854
	193	0.809	0.868	0.870	0.853	0.854

TABLE 2.16: The DCG measure of HMS based on different pairs of parameters.

HMS	t							
τ	40τ	60τ	80τ	100τ	120τ	140τ	160τ	180τ
0.005	0.676	0.692	0.697	0.690	0.734	0.712	0.729	0.740
0.01	0.702	0.722	0.744	0.714	0.708	0.734	0.712	0.714
0.02	0.713	0.726	0.712	0.742	0.719	0.720	0.722	0.723
0.05	0.724	0.734	0.729	0.744	0.736	0.733	0.727	0.737
0.1	0.744	0.739	0.742	0.717	0.712	0.729	0.713	0.705
0.2	0.723	0.716	0.712	0.718	0.749	0.753	0.751	0.753
0.4	0.739	0.744	0.744	0.753	0.749	0.753	0.745	0.752
0.6	0.735	0.747	0.752	0.750	0.747	0.754	0.747	0.742
1	0.751	0.751	0.753	**0.754**	0.744	0.744	0.739	0.738
2	0.739	0.744	0.729	0.733	0.723	0.736	0.731	0.725

TABLE 2.17: The DCG measure of WKS based on different pairs of parameters.

WKS	σ				
M	0.12	0.16	0.20	0.24	0.28
20	0.708	0.720	0.718	0.724	0.697
50	0.722	0.724	**0.727**	0.725	0.703
80	0.713	0.720	0.725	0.724	0.706
100	0.717	0.712	0.726	0.720	0.708
120	0.720	0.720	0.718	0.715	0.713
150	0.716	0.724	0.723	0.713	0.714

TABLE 2.18: Performance comparison of descriptors and their optimal parameters on SHREC 2010 dataset.

	Spectral Descriptors						
	HKS	SIHKS	HMS	WKS	GPS	ShapeDNA	EVD
Parameters	$\tau = 1/4$ $T = 10$ $t_0 = 16.24$ $\alpha = 1.24$	$\tau = 1/16$ $T = 25$ $F = 193$ $\alpha = 2$	$\tau = 2$ $t = 100\tau$	$M = 50$ $\sigma = 0.2$			
DCG	0.848	0.877	0.754	0.727	0.757	0.801	0.636

CHAPTER 3

SPECTRAL GRAPH WAVELET SIGNATURE

In this chapter, we present a spectral graph wavelet framework for the analysis and design of efficient shape signatures for nonrigid 3D shape retrieval. Although this work focuses primarily on shape retrieval, our approach is, however, fairly general and can be used to address other 3D shape analysis problems. In a bid to capture the global and local geometry of 3D shapes, we propose a multiresolution signature via a cubic spline wavelet generating kernel. The parameters of the proposed signature can be easily determined as a tradeoff between effectiveness and compactness. Experimental results on two standard 3D shape benchmarks demonstrate the much better performance of the proposed shape retrieval approach in comparison with three state-of-the-art methods. Additionally, our approach yields a higher retrieval accuracy when used in conjunction with the intrinsic spatial partition matching.

3.1 INTRODUCTION

The content-based differentiation between 3D objects from different classes is being pursued in a number of established and emerging fields, including animation, molecular biology and medicine, computer-aided design, multimedia entertainment, and mobile game development. With the increasing use of 3D scanners and as a result of emerging multimedia computing technologies, large databases of 3D models are distributed freely or commercially on the World Wide Web. The availability and widespread usage of such databases, coupled with the need to explore 3D models in depth as well as in breadth, has sparked the need to organize and search these vast repositories, and efficiently retrieve the most relevant selec-

tions. The shape retrieval problem has been extensively investigated in the literature, from comprehensive surveys [1, 2, 3, 4] to comparable benchmarks [82, 81, 83].

Recently, there has been a surge of interest in the spectral analysis of the Laplace-Beltrami (LB) operator, resulting in many applications to manifold learning [84], object recognition and deformable shape analysis [46, 16, 85, 15, 34]. It is worth pointing out that spherical harmonics [86] are nothing but the LB eigenfunctions on the sphere. The truncated sequence of the LB eigenvalues was proposed by Reuter *et al.* as an isometry-invariant global shape descriptor, dubbed shape-DNA [16]. Reuter also introduced a Morse-theoretic method for shape segmentation and registration using the topological features of the eigenfunctions [85]. These eigenfunctions are computed via a cubic finite element method on triangular meshes, and are arranged in increasing order of their associated eigenvalues. Rustamov [15] proposed a feature descriptor referred to as the global point signature (GPS), which is a vector whose components are scaled eigenfunctions of the LB operator evaluated at each surface point. GPS is invariant under isometric deformations of the shape, but it suffers from the problem of eigenfunctions' switching whenever the associated eigenvalues are close to each other. This problem was lately well handled by the heat kernel signature (HKS) [11], which is a temporal descriptor defined as an exponentially-weighted combination of the eigenfunctions. It is a local shape descriptor that has a number of desirable properties, including robustness to small perturbations of the shape, efficiency and invariance to isometric transformations. The idea of HKS was also independently proposed by Gębal *et al.* [87] for 3D shape skeletonization and segmentation under the name of auto diffusion function. Using the Fourier transform's magnitude, Kokkinos *et al.* introduced the scale invariant heat kernel signature (SIHKS) [12], which is constructed based on a logarithmically sampled scale-space. Also, by faithfully reflecting the temperature distribution during the heat process, the heat mean signature (HMS) was recently proposed by Fang *et al.* for shape segmentation [13] and retrieval [17].

From the graph Fourier perspective, it can be seen that HKS is highly dominated by information from low frequencies, which correspond to macroscopic properties of a shape. To give rise to substantially more accurate matching than HKS, the wave kernel signature (WKS) [14] was proposed as an alternative in an effort to allow access to high-frequency information. Despite being physically inspired, both WKS and HKS can be regarded as filters. On the other hand, in order to construct a good task-specific spectral descriptor, one has to be in the position of defining the spectral content of the geometric "signal" and the "noise". Bronstein [88] proposed to learn the signal and noise from examples in a way that resembles the construction of a Wiener filter that passes frequencies containing more signal than noise, while attenuating those where the noise covers the signal.

In this chapter, we introduce a wavelet-based signature for nonrigid 3D shape retrieval. Wavelets are a class of a functions used to localize a given function in both space and scaling [89]. The wavelet analysis has some major advantages over Fourier transform, which makes it an interesting alternative for many applications. In particular, unlike the Fourier transform, wavelet analysis is able to perform local analysis and also makes it possible to perform a multiresolution analysis. Classical wavelets are constructed by translating and scaling a mother wavelet, which is used to generate a set of functions through the scaling and translation operations. The wavelet transform coefficients are then obtained by taking the inner product of the input function with the translated and scaled waveforms. The application of wavelets to graphs (or triangle meshes) is, however, problematic and not straightforward due in part to the fact that it is unclear how to apply the scaling operation on a signal (or function) defined on the mesh vertices. To tackle this problem, Coifman *et al.* [90] introduced the diffusion wavelets, which generalize the classical wavelets by allowing for multiscale analysis on graphs. The construction of diffusion wavelets interacts with the underlying graph through repeated applications of a diffusion operator, which induces a scaling process. More recently, Hammond *et al.* [91] showed that the wavelet transform can be performed in the graph Fourier domain, and proposed a spectral graph wavelet transform that is defined in terms of the eigensystem of the graph Laplacian matrix. In the same vein, Kim *et al.* [92] introduced a wavelet-based multiscale descriptor for the analysis of cortical surface signals (such as cortical thickness) using the spectral graph wavelet transform. While building on earlier efforts, we take a rather different approach in this chapter by proposing a novel multiresolution shape signature that is not only isometric invariant, but also compact, easy to compute and combines the advantages of both band-pass and low-pass filters.

3.1.1 Contributions

Our main contributions in this chapter may be summarized as follows:

(i) We present a general and flexible framework for the analysis and design of shape signatures from the spectral graph wavelet perspective.

(ii) We propose a multiresolution shape signature for deformable 3D shape retrieval using a cubic spline generating kernel. The main attractive properties of the proposed signature may be summarized as follows: It can capture both global and local geometry of shapes in a multiresolution fashion; its parameters can be automatically determined as a tradeoff between effectiveness and compactness; it nicely fits the intrinsic spatial partition matching; and it yields the best retrieval

accuracy on two standard 3D shape benchmarks compared to spectral signatures within the diffusion geometric framework.

The rest of the chapter is organized as follows. In Section 3.2, we provide some background on the eigenanalysis of LB operator. Then, we briefly review two closely related spectral signatures to our proposed shape descriptor. In Section 3.3, we propose a multiresolution shape signature in the spectral graph wavelet framework. Using a cubic spline wavelet generating kernel, we introduce an efficient approach for nonrigid 3D shape retrieval in Section 3.4. Experimental results on two standard 3D shape datasets are provided in Section 3.5.

3.2 Spectral Geometric Signatures

3.2.1 Eigenanalysis

The eigenvalues λ_ℓ and the associated eigenfunctions χ_ℓ of the LB operator can be computed by solving the following generalized eigenvalue problem:

$$C\chi_\ell = \lambda_\ell R\chi_\ell, \qquad \ell = 1,2,\ldots,n \tag{3.1}$$

Assuming that we have a connected triangle mesh, we may sort the eigenvalues in ascending order as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n = \lambda_{\max}$ with associated orthonormal eigenfunctions as $\chi_1, \chi_2, \ldots, \chi_n$. The eigensystem $\{\lambda_\ell, \chi_\ell\}_{\ell=1,\ldots,n}$ of the LB operator enjoys nice properties, including intrinsicness to the mesh surface and isometry invariance.

3.2.2 Spectral Shape Signatures

Several shape descriptors based on the eigensystem of the LB operator have been proposed in the literature [16, 15, 11, 14, 13]. The spectral signature HKS and WKS are particularly the most closely related to our work. Both HKS and WKS have an elegant physical interpretation: HKS describes the amount of heat remaining at a mesh vertex $j \in \mathcal{V}$ after a certain time, whereas WKS is the probability of measuring a quantum particle with the initial energy distribution at j. The HKS descriptor at a vertex j is defined as:

$$\mathfrak{s}_{t_k}(j) = \sum_{\ell=1}^{n} \exp(-\lambda_\ell t_k)\chi_\ell^2(j) \tag{3.2}$$

and contains information mainly from low frequencies, which correspond to macroscopic features of the shape; and thus exhibits a major discrimination ability in shape retrieval tasks. With multiple scaling

factors t_k, a collection of low-pass filters are established. The larger is t_k, the more high frequencies are suppressed. However, different frequencies are always mixed in HKS, and high-precision localization task may fail due in part to the suppression of the high frequency information, which corresponds to microscopic features. To circumvent these disadvantages, the WKS descriptor was introduced [14] and it is defined as:

$$\mathfrak{s}_{t_k}(j) = \sum_{\ell=1}^{n} C_{t_k} \exp\left(\frac{-(\log t_k - \log \lambda_\ell)^2}{\sigma^2}\right) \chi_\ell^2(j) \tag{3.3}$$

where C_{t_k} is a normalization constant [14]. The WKS explicitly separates the influences of different frequencies, treating all frequencies equally. Thus, different spatial scales are naturally separated, making the high-precision feature localization possible. Given a range of discrete scales t_k, a bank of filters is constructed for each signature, and thus a vertex j on the mesh surface can be described by the following signature vector:

$$s(j) = \{\mathfrak{s}_{t_k}(j) \mid k = 1, \ldots, K\}. \tag{3.4}$$

3.2.3 Graph Fourier Transform

The Fourier transform is a powerful mathematical tool for the analysis of non-periodic functions, and defines a relationship between a function in the time domain and its representation in the frequency domain. It basically maps a function defined on physical space to a function defined on the space of frequencies, whose values quantify the "amount" of each periodic frequency contained in the original function. The inverse Fourier transform then reconstructs the original function from its transformed frequency components. The integrals defining the Fourier transform and its inverse are, remarkably, almost identical. More precisely, the continuous Fourier transform is the expansion of a function f (defined on the real line) in terms of the eigenfunctions $e^{\imath\omega x}$ of the 1D Laplacian operator $\frac{d^2}{dx^2}$, and it is given by

$$\hat{f}(\omega) = \langle f, e^{-\imath\omega x} \rangle = \int_{-\infty}^{\infty} f(x) e^{-\imath\omega x} dx, \quad \forall \omega \in \mathbb{R} \tag{3.5}$$

where $\imath$ denotes the unit complex number ($\imath^2 = -1$). When the variable x represents time, the transform variable ω represents angular frequency. The inverse Fourier transform is given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{\imath\omega x} d\omega, \quad \forall x \in \mathbb{R}. \tag{3.6}$$

The eigensystem $\{\lambda_\ell, \chi_\ell\}$ of the LB operator can be interpreted in the same vein as the Fourier basis: the eigenvalues λ_ℓ act as the frequencies ω, while the eigenfunctions χ_ℓ play the role of the complex exponentials (Fourier basis functions) $e^{\imath\omega x}$. Analogously to (3.5) and (3.6), for any function $f \in \mathbb{R}^n$ defined on

the mesh vertex set $\mathcal{V}$, the forward and inverse graph Fourier transforms are defined, respectively, by

$$\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{n} \chi_\ell^*(i) f(i), \quad \ell = 1, \ldots, n \tag{3.7}$$

and

$$f(j) = \sum_{\ell=1}^{n} \hat{f}(\ell) \chi_\ell(j), \quad j \in \mathcal{V} \tag{3.8}$$

where $*$ denotes operation of complex conjugate. Notice that the graph Fourier transform is the expansion of f in terms of the eigenfunctions, while the inverse Fourier transform provides a representation of f as a superposition of the eigenfunctions.

Fourier analysis allows us to study the cyclical nature of a signal in the frequency domain. In spite of its utility, however, under the Fourier transform, the time information of a signal is lost. Because of this loss of information, it is hard to distinguish transient relations or to identify when structural changes do occur. To circumvent these limitations, wavelet analysis performs the estimation of the spectral components of a signal as a function of time, revealing how the different periodic components of the signal change over time. A major advantage of using wavelet analysis over Fourier analysis is the possibility of tracing transitional changes across time.

3.3 Proposed Approach

Similar to the Fourier transform which decomposes a signal into its constituent frequencies, the wavelet transform is a powerful multiresolution analysis tool that enable decomposition of a signal into a wavelet basis which allows simultaneous localization in space and frequency. Wavelet analysis provides a time-scale representation and extends frequency analysis to scale, while Fourier analysis only gives the frequency information [89]. The idea of wavelets is based on the use of two main operations on the signal, namely shifting and scaling. Using these two operations, a signal f can be represented as the sum of shifted and scaled versions of the so-called mother wavelet function, ψ, and shifted versions of the so-called scaling function, ϕ. The mother wavelet and scaling functions act as band-pass and low-pass functions, respectively.

3.3.1 Classical Continuous Wavelet Transform

The continuous wavelet transform maps the original signal, which is a function of just one variable (time) into a function of two variables (time and frequency), providing highly redundant information. More specifically, for a given mother wavelet ψ, a family $\psi_{t,a}$ of daughter wavelets can be obtained by simply

scaling and translating ψ as follows:

$$\psi_{t,a}(x) = \frac{1}{t}\psi\left(\frac{x-a}{t}\right), \tag{3.9}$$

where t is a positive scaling parameter that controls the width of the wavelet, and a is a translation parameter that controls the location of the wavelet. Scaling a wavelet simply means stretching it (if $t > 1$) or compressing it (if $t < 1$), while translating a wavelet simply means shifting its position in time. Note that the translation parameter does not have a counterpart in the Fourier basis functions, where the position information is totally missing. It should also be noted that the scaling parameter in the wavelet analysis is similar to the scale used in maps. As in the case of maps, high scales correspond to a non-detailed global view of the signal, while low scales correspond to a detailed view.

Given a square-integrable signal f, the continuous wavelet transform (CWT) with respect to the mother wavelet ψ is expressed by the following integral

$$W_f(t,a) = \langle \psi_{t,a}, f \rangle = \frac{1}{t}\int_{-\infty}^{\infty} f(x)\psi^*\left(\frac{x-a}{t}\right)dx, \tag{3.10}$$

which is also referred to as the wavelet coefficient at scale t and location a. The position of the wavelet in the time domain is given by the translational value a, while its position in the frequency domain is given by the scale t. Thus, the CWT gives us information simultaneously on time and frequency. Unlike Fourier transform, the CWT possesses the ability to construct a time-frequency representation of a signal that offers very good time and frequency localization. The CWT may be invertible when the mother wavelet ψ satisfies the admissibility condition, $C_\psi = \int_0^\infty \frac{|\hat{\psi}(\omega)|^2}{\omega} \mathrm{d}\omega < \infty$, where $\hat{\psi}$ is the Fourier transform of ψ. The inverse CWT is given by

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^{\infty} W_f(t,a)\psi_{t,a}(x)\frac{da\,dt}{t}. \tag{3.11}$$

For a fixed scale t, the CWT may be interpreted as an operator taking a function f and returning the function $(T^t f)(a) = W_f(t,a)$. In other words, the translation parameter can be considered as the independent variable of the function returned by the operator T^t. The CWT may also be expressed in the Fourier domain as [91]:

$$(T^t f)(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}^*(t\omega)\hat{f}(\omega)e^{\imath\omega x}d\omega, \tag{3.12}$$

where $\hat{\psi}^*(t\omega)$ is the complex conjugate of the Fourier transform of the wavelet ψ at scale t, and $\hat{f}(\omega)$ is the Fourier transform of the signal f. The scaling parameter t appears only in the argument of $\hat{\psi}^*(t\omega)$, showing that the scaling operation can be completely transferred to the Fourier domain. It is clear that the wavelet transform at each scale can be viewed as a Fourier multiplier operator, determined by filters that are derived from scaling a single filter $\hat{\psi}^*(\omega)$. This idea was adopted by Hammond *et al.* [91] to provide the analogue of the wavelet transform on weighted graphs via spectral graph theory.

3.3.2 Spectral Graph Wavelet Transform

Recall that the wavelet transform is based on the scaling and translation operations. It is, however, not straightforward how to apply the scaling operation on a signal (or function) defined on the mesh vertices. In other words, for a function $f(i)$ defined on a mesh vertex $i \in \mathcal{V}$, it is meaningless to interpret $f(ti)$ on its own domain for a scaling constant t. To tackle this problem, Hammond *et al.* [91] introduced the spectral graph wavelet transform (SGWT). The idea is to cast the problem in the Fourier domain and then define the required scaling in that domain. The SGWT is determined by the choice of a spectral graph wavelet generating kernel $g : \mathbb{R}^+ \to \mathbb{R}^+$, which is analogous to the Fourier domain wavelet $\hat{\psi}^*$ in Eq. (3.12). To act as a band-pass filter, the kernel g should satisfy $g(0) = 0$ and $\lim_{x\to\infty} g(x) = 0$.

Wavelet Function: Let g be a given kernel function and denote by T_g^t the wavelet operator at scale t. This operator acts on a given function f by modulating its Fourier domain representation as follows:

$$\widehat{T_g^t f}(\ell) = g(t\lambda_\ell)\hat{f}(\ell) \tag{3.13}$$

Thus, the inverse graph Fourier transform is given by

$$(T_g^t f)(i) = \sum_{\ell=1}^{n} \widehat{T_g^t f}(\ell)\chi_\ell(i) = \sum_{\ell=1}^{n} g(t\lambda_\ell)\hat{f}(\ell)\chi_\ell(i). \tag{3.14}$$

Applying the wavelet operator T_g^t to an impulse function, that is $f(i) = \delta_j(i)$ on each mesh vertex $i \in \mathcal{V}$ (i.e. $\delta_j(i) = \delta(i-j)$), and using the fact the graph Fourier transform of δ_j is

$$\hat{\delta}_j(\ell) = \langle \chi_\ell, \delta_j \rangle = \sum_{i=1}^{n} \chi_\ell^*(i)\delta_j(i) = \chi_\ell^*(j), \tag{3.15}$$

it follows that the spectral graph wavelet $\psi_{t,j}$ localized at vertex j and scale t is given by

$$\psi_{t,j}(i) = \widehat{T_g^t \delta_j}(i) = \sum_{\ell=1}^{n} g(t\lambda_\ell)\chi_\ell^*(j)\chi_\ell(i). \tag{3.16}$$

It should be noted that $g(t\lambda_\ell)$ is able to modulate the spectral wavelets $\psi_{t,j}$ only for λ_ℓ within the domain of the spectrum of LB operator. Thus, an upper bound on the largest eigenvalue $\lambda_{\max}$ is required to provide knowledge on the spectrum in practical applications.

Hence, the spectral graph wavelet coefficients of a given function f can be generated from its inner product with the spectral graph wavelets:

$$W_f(t,j) = \langle \psi_{t,j}, f \rangle = \sum_{\ell=1}^{n} g(t\lambda_\ell)\hat{f}(\ell)\chi_\ell(j), \tag{3.17}$$

which closely resembles Eq. (3.12). In other words, the LB eigenfunctions are the natural generalization of Fourier modes on Euclidean spaces, which are the (generalized) eigenfunctions of the 1D Euclidean Laplacian.

Scaling Function: Similar to the low-pass scaling functions in the classical wavelet analysis, a second class of waveforms $h : \mathbb{R}^+ \to \mathbb{R}$ will be used as low-pass filters to better encode the low-frequency content of a function f defined on the mesh vertices. To act as a low-pass filter, the function h should satisfy $h(0) > 0$ and $h(x) \to 0$ as $x \to \infty$. Similar to the wavelet kernels, the scaling functions are given by $\phi_j = T_h \delta_j$, and their spectral coefficients by

$$S_f(j) = \langle \phi_j, f \rangle = \sum_{\ell=1}^{n} h(\lambda_\ell) \hat{f}(\ell) \chi_\ell(j). \tag{3.18}$$

A major advantage of using the scaling function is to ensure that the original signal f can be stably recovered when sampling scale parameter t with a discrete number of values t_k. As demonstrated in [91], given a set of scales $\{t_k\}_{k=1}^K$, the set $F = \{\phi_j\}_{j=1}^n \cup \{\psi_{t_k,j}\}_{k=1}^K{}_{j=1}^n$ forms a spectral graph wavelet frame with bounds

$$A = \min_{\lambda \in [0, \lambda_{\max}]} G(\lambda) \quad \text{and} \quad B = \max_{\lambda \in [0, \lambda_{\max}]} G(\lambda), \tag{3.19}$$

where

$$G(\lambda) = h(\lambda)^2 + \sum_k g(t_k \lambda)^2. \tag{3.20}$$

The stable recovery of f is ensured when A and B are away from zero. Additionally, the crux of the scaling function is to smoothly represent the low-frequency content of the signal on the mesh. Thus, the design of the scaling function h is uncoupled from the choice of wavelet generating kernel g.

3.3.3 Proposed Multiresolution Shape Signature

Wavelets are useful in describing functions at different levels of resolution. For example, an image can be described at different levels of resolution. Understanding and characterizing the differences between functions at different levels of resolution is what wavelets are all about. To characterize the localized context around a mesh vertex $j \in \mathcal{V}$, we assume that the signal on the mesh is a unit impulse function, that is $f(i) = \delta_j(i)$ at each mesh vertex $i \in \mathcal{V}$. Thus, it follows from Eq. (3.14) that the spectral graph wavelet coefficients are

$$W_{\delta_j}(t, j) = \langle \psi_{t,j}, \delta_j \rangle = \sum_{\ell=1}^{n} g(t\lambda_\ell) \chi_\ell^2(j) \tag{3.21}$$

and that the coefficients of the scaling function are

$$S_{\delta_j}(j) = \sum_{\ell=1}^{n} h(\lambda_\ell) \chi_\ell^2(j). \tag{3.22}$$

Following the multiresolution analysis, the spectral graph wavelet and scaling function coefficients are collected to form the following multiresolution shape signature:

$$\mathcal{S}_R(j) = \{s_L(j) \mid L = 1, \ldots, R\}, \tag{3.23}$$

which we refer to as the *spectral graph wavelet signature* (SGWS), where R is the resolution parameter, and $s_L(j)$ is the shape signature at resolution level L given by

$$s_L(j) = \{W_{\delta_j}(t_k, j) \mid k = 1, \ldots, L\} \cup \{S_{\delta_j}(j)\}. \tag{3.24}$$

The wavelet scales t_k ($t_k > t_{k+1}$) are selected to be logarithmically equispaced between maximum and minimum scales t_1 and t_L, respectively. Thus, the resolution level L determines the resolution of scales to modulate the spectrum. The proposed shape signature can be represented as a pyramid, as depicted in Figure 3.1, which shows that at resolution level $L = 1$, the signature $s_L(j)$ consists of two elements: one element, $W_{\delta_j}(t_1, j)$, of spectral graph wavelet function coefficients and another element, $S_{\delta_j}(j)$, of scaling function coefficients. And at level $L = 4$, the signature $s_L(j)$ consists of five elements (four elements of spectral graph wavelet function coefficients and one element of scaling function coefficients). Hence, if the resolution is set to $R = 4$, then the multiresolution signature $\mathscr{S}_R(j)$ is composed of a total of 14 elements, as illustrated in Figure 3.1.

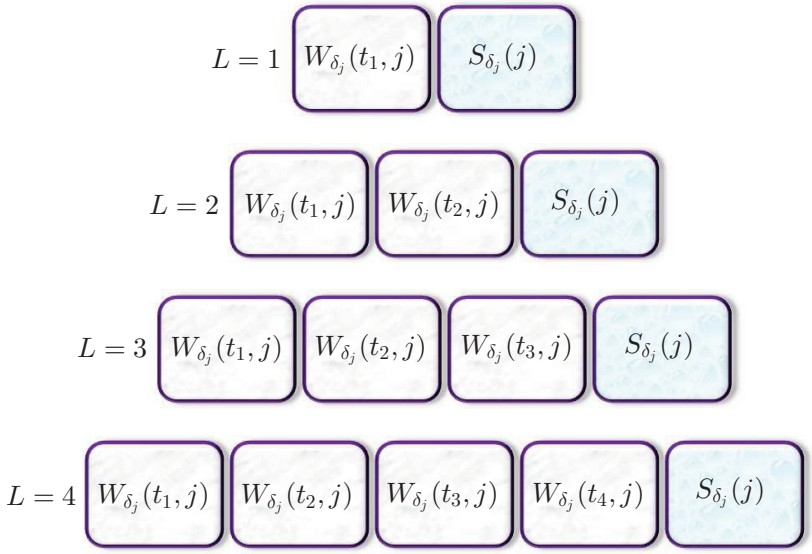


FIGURE 3.1: Pyramid representation of the proposed spectral graph wavelet signature at various resolutions levels $L = 1, \ldots, R$, where $R = 4$.

The spectral graph wavelet coefficients, given by Eq. (3.21), yield a general form for spectral signatures, which includes both HKS and WKS as particular cases, indicating the close relationship between these two signatures and our proposed SGWG. Moreover, our approach provides a general and flexible framework for the analysis and design of shape signatures from the wavelet viewpoint. Unlike HKS and WKS, our proposed signature allows a multiresolution representation of shapes. The multiresolution analysis using spectral graph wavelets model the shape content in different levels. These levels are used to cap-

ture different details inherently found in different structures, as shown in Figure 3.2. Given any vertex j on the mesh surface, we calculate the dissimilarity between j and a reference vertex j_{ref} (the yellow point on the horse's back left leg in Figure 3.2) at resolution level L, using the χ^2-distance given by

$$d(j, j_{\text{ref}}) = \sum_{k=1}^{L+1} \frac{\left(s_k(j) - s_k(j_{\text{ref}})\right)^2}{s_k(j) + s_k(j_{\text{ref}})}, \tag{3.25}$$

where s_k denote the elements of the signature vector s_L of length $L+1$. For visualization purposes, we normalize all the χ^2-distances to $[0,1]$ (dividing by $\max d(j, j_{\text{ref}})$ on the surface). As can be seen in Figure 3.2, the discriminative power from the reference point to other points on the horse surface is dominated by the resolution level $L = 2$. By removing some local geometry, $L = 1$ exhibits the second discriminative ability. For $L = 3$, a detailed geometry is included, which is suitable for distinguishing between the reference point and the vertices around the horse feet. Moreover, compared with their dissimilarity value, other points have a much smaller dissimilarity. From $L = 3$ to $L = 6$, more detailed geometry starts to dominate the representation and it always maintains a large dissimilarity between the reference and feet, resulting in relatively small dissimilarities for other points.

3.4 Cubic Spline Wavelet for Deformable Shape Retrieval

In this section, we discuss what a reasonable good descriptor is for shape retrieval [88], and how to design a signature that satisfies the following three properties:

- *Invariance*: The descriptor should be invariant or at least insensitive to a certain class of transformations that the shape may undergo. In this chapter, we consider shapes with only isometric or near-isometric transformations. Since the eigensystem of LB operator is intrinsic, the spectral signature is naturally isometry invariant.

- *Efficiency*: The descriptor should capture as much information as possible within as little number of dimensions as possible. Additionally, a wavelet provides very compact support because of its band-pass nature, and thus less redundant information is contained among different supports. To guarantee the localization ability of the spectral graph wavelet in the limit of fine scale, the kernel function g should behave as a monic power of a mesh vertex near the origin [91].

- *Discriminative Power*: The descriptor should be able to distinguish between shapes belonging to different classes. First, shapes from different classes usually have different micro structures. We use the multiresolution strategy to capture the micro structures at different resolution levels in a

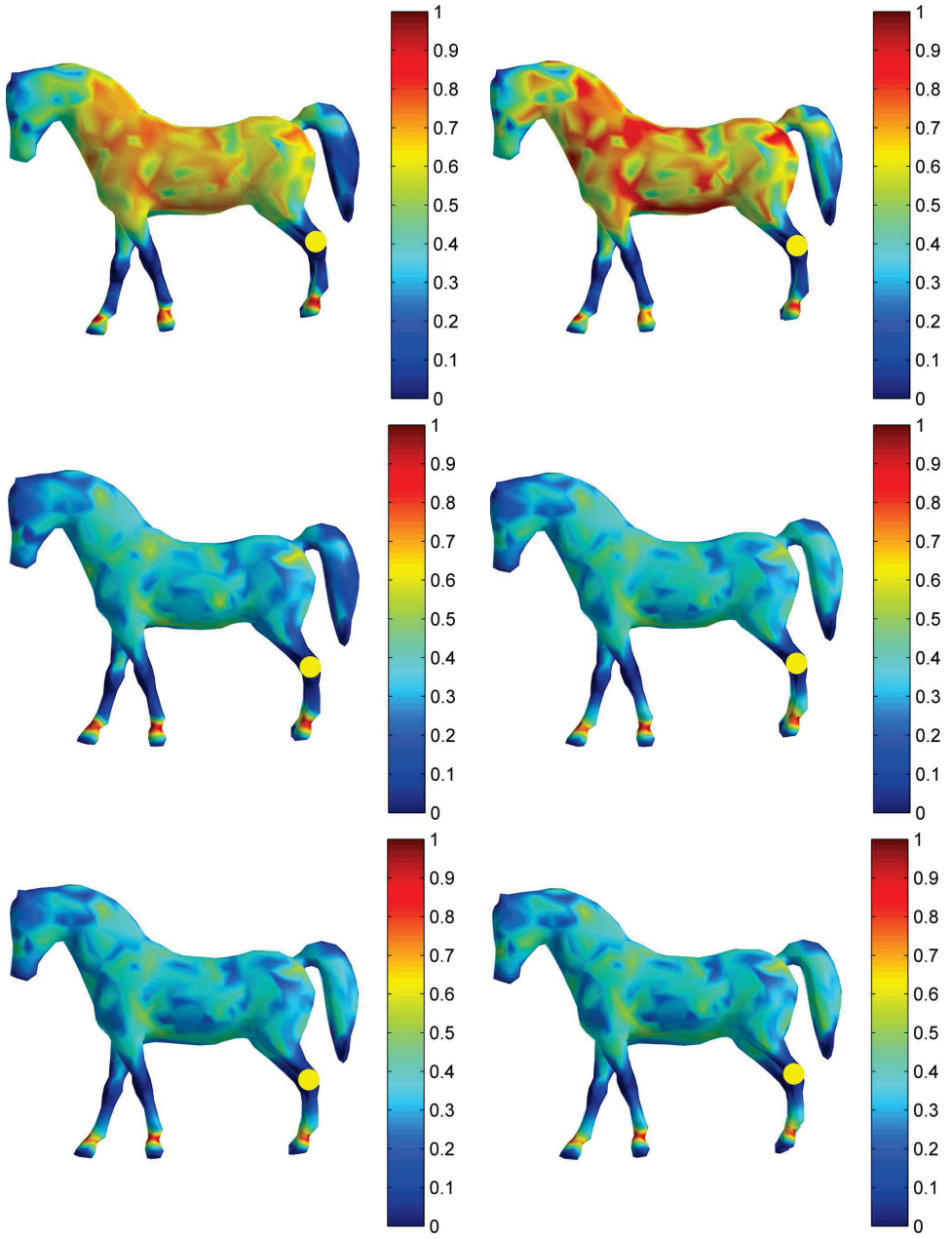


FIGURE 3.2: Normalized χ^2-distance between a reference point (yellow colored point on the horse's back left leg) and other surface points using SGWS at different resolution levels. From top to bottom and left to right, the resolution levels are $L = 1, 2, 3, 4, 5$ and 6.

principled fashion. Second, low-frequency information represents the macro structures, which are critical to shape comparison. Consequently, the scaling function h is an integral part of our proposed shape signature. The guidance to design a proper h is not only to make up for the information loss by the wavelet kernel g in low frequency, but also not to overlap with g.

As suggested in [91], we choose the cubic spline wavelet and scaling function kernels given by

$$g(x) = \begin{cases} x^2 & \text{if } x < 1 \\ -5 + 11x - 6x^2 + x^3 & \text{if } 1 \le x \le 2 \\ 4x^{-2} & \text{if } x > 2 \end{cases} \tag{3.26}$$

and

$$h(x) = \gamma \exp\left(-\left(\frac{x}{0.6\lambda_{\min}}\right)^4\right), \tag{3.27}$$

respectively, where $\lambda_{\min} = \lambda_{\max}/20$, γ is set such that $h(0)$ has the same value as the maximum value of g. The maximum and minimum scales are set to $t_1 = 2/\lambda_{\min}$ and $t_L = 2/\lambda_{\max}$.

The geometry captured at each resolution level L of the SGWS can be viewed as the area under the curve G shown in Figure 3.3. For a given resolution level L, we can understand the information from a specific range of the spectrum as its associated areas under G. As the resolution level L increases, the partition of spectrum becomes tighter, and thus a larger portion of the spectrum is highly weighted.

Relation to HKS: The kernel function $g_{\text{HKS}} = \exp(-t\lambda)$ yields HKS. Since $g_{\text{HKS}} \neq 0$, it follows that g_{HKS} does not satisfy the admissibility condition. As can be seen in Figure 3.3(a), there is a rich redundance along the spectrum between different scales. However, multiresolution analysis is not possible for HKS. Since g_{HKS} acts as a low-pass filter, HKS will fail to capture micro structures. Consequently, HKS is not able to conduct high-precision feature localization. As can be observed in the top row of Figure 3.4, the χ^2-distance changes slowly along the surface. The same distance value remains with the isometric transformation. When a human body (shape from a different class) compares with the reference, large distances occupy most of the area.

Relation to WKS: The WKS is obtained using the kernel function $g_{\text{WKS}} = C_t \exp\left(\frac{-(\log t - \log\lambda)^2}{\sigma^2}\right)$, which is a log-normal distribution function that forms a wavelet generating kernel, where C_{t_k} is a normalization constant and σ^2 is the variance of the distribution [14]. Unlike our proposed shape signature, there is, however, no scaling function for WKS. Therefore, less low-frequency information is used in WKS, as illustrated in Figure 3.3(b), resulting in a substantial loss of global geometry of shapes. In the middle row of Figure 3.4, small χ^2-distances appear on a large percentage of the surface even when comparing the reference to a body shape. The advantage of WKS is also vividly depicted in the sense that micro structures that are different with the local geometry of the reference are detected.

Proposed Signature Revisited: Observed from the reconstruction perspective, the cubic spline kernel is devised obeying the rules of the wavelet generating kernel function. Therefore, the signal defined on the surface can be stably recovered, whereas neither HKS nor WKS allows stable recovery. Figure 3.3(c)-(h) display the cubic wavelet spline kernels and their squared sum function G for different values of the

resolution parameter R, ranging from 1 to 6. Each value of G indicates the energy contribution from each frequency, and it also acts as a good tool to visualize the resolution oscillation for each resolution R. The last row of Figure 3.4 shows our proposed signature with $R = 2$. As can be seen, the SGWS integrates the advantages of both HKS (global geometry for discriminative power) and WKS (local geometry for localization). The former is demonstrated by the body surface colored in red. The χ^2-distances are larger than the horse model. The latter is evidenced by sharply detected features close to the end of the four legs of the horse. As a whole, that is why we observe that the proposed signature brings steeper color change within a neighborhood and deeper red on a different shape.

3.5 Experimental Results

To assess the performance of our proposed approach on 3D shape retrieval, we conducted experiments on two standard benchmarks: SHREC 2010 [80] and SHREC 2011 [81]. Instead of adopting the Gaussian kernel for the codeword ambiguity modeling [34], we use the traditional L_0-norm codebook because it can evaluate signatures without the effect of the kernel size. Shape comparison is then performed via the proposed intrinsic spatial pyramid matching (ISPM) in Chapter 2. The key idea of ISPM is to adopt the level sets of the second eigenfunction of LB operator as cuts to perform surface partition. ISPM is largely motivated by spatial pyramid matching introduced in [35], which partitions an image into increasingly fine sub-regions and then computes histograms of local features found inside each sub-region. For the convenience of comparison, we evaluated the shape retrieval performance using the Discounted Cumulative Gain (DCG) [79].

3.5.1 Settings

Comparing Signatures: We compared the proposed method with classical spectral signatures, including SIHKS, HKS and WKS. The SIHKS is chosen because of its excellent performance in the Shape Google algorithm [34]. The first 150 eigenvalues and eigenvectors of the LB operator on each shape are used. We experimentally select the best parameters for each signature as well as on each SHREC dataset.

Complexity: As the same settings as 2, We report the results on the simplified mesh dataset. For a set of approximately 2×10^5 descriptors, the runtime (in seconds) for a vocabulary with size 32 ranges from 77s for $R = 1$ to 501s for $R = 6$. On the other hand, the runtimes for SIHKS, HKS and WKS with their best parameters are 5702s, 340s and 725s, respectively.

3.5.2 RESULTS ON SHREC 2010

First, we examine the performance of the SGWS with varying resolutions. In Table 3.1, the best DCG value of the SGWS matched by bag-of-feature (BoF) is displayed in the row named "Level 0", and it was obtained using a resolution $R = 2$. Interestingly, for $R = 2$ the proposed shape signature has only 5 elements, but it achieves a comparably good result. The element represented by the scaling function captures the global features, whereas the element represented by the spectral graph wavelet function captures more detailed local features and gives evidence that macro structures dominate the deformable shape retrieval. This argument is demonstrated again using $R = 2$, where the best result 0.8635 rises when only the sub low-frequency band is added. In addition, it can be seen in Table 3.1 that the accuracy is consistently better than HKS and WKS from $R = 2$ all the way up to $R = 6$. We stop at $R = 6$ for two main reasons: (i) The higher is the resolution, the more expensive is the computational cost. However, the performance remains relatively stable; (ii) In Figure 3.3, each value of G indicates the energy contribution of each frequency. The first local minimum close to 0 becomes the global minimum since $R = 6$, indicating the end of low-frequency domination.

To summarize, the proposed signature is highly efficient since it yields high accuracy with a very compact representation. Additionally, it is practical due largely to its stable results and simple parameters' selection.

Next, we examine the behavior of the SGWS with intrinsic spatial pyramid matching using different pyramid levels ł and different surface partitions. The rows named "Level 1" to "Level 4", in Table 3.1, list the DCG performance achieved at different pyramid levels. The improvement of ISPM over BoF is displayed in the last row of the table, where it can be seen that WKS has improved most significantly by gaining 0.0354, while SIHKS and HKS are improved slightly by 0.0122 and 0.0083, respectively. This can be understood as the potential of signatures, independently occupying the frequency band, is well improved by ISPM. However, WKS still gives the worst result because of its low baseline. The problem is well-balanced by the SGWS since the critical low-frequency is also incorporated independently. As expected, the improvement as well as the accuracy of the SGSW are always higher than HKS. On the other hand, all the signatures are improved considerably as we go from ł = 1 to a multi-level setup. The proposed SGWS achieves the best result at ł = 2, while the other signatures at ł = 3 or higher.

TABLE 3.1: Performance (DCG) of different spectral descriptors on SHREC 2010 with intrinsic spatial pyramid matching.

Level l (Partitions)	SIHKS	HKS	WKS	Spectral Graph Wavelet Signature (R)					
				1	2	3	4	5	6
0 (1)	0.8719	0.8448	0.7280	0.8124	**0.8635**	0.8541	0.8454	0.8547	0.8497
1 (2)	0.8734	0.8420	0.7290	0.8172	0.8593	0.8603	0.8498	0.8554	0.8527
2 (4)	0.8793	0.8509	0.7509	0.8789	0.8792	0.8732	0.8681	0.8665	0.8634
3 (8)	0.8817	0.8531	0.7612	0.8781	0.8737	0.8682	0.8651	0.8604	0.8615
4 (16)	0.8841	0.8520	0.7634	0.8687	0.8721	0.8636	0.8640	0.8563	0.8590
Improvement	0.0122	0.0083	0.0354	0.0665	0.0157	0.0191	0.0227	0.0118	0.0137

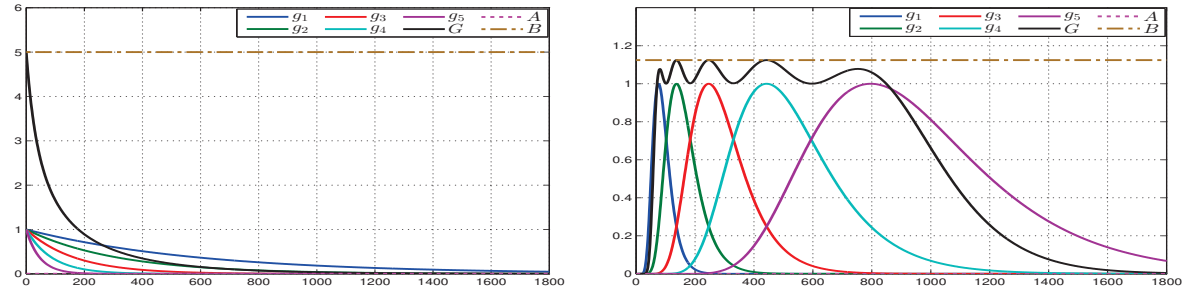
3.5.3 RESULTS ON SHREC 2011

Table 2.13 gives a breakdown of shape retrieval accuracy for different resolutions R and at different pyramid levels ł. For ł = 0, our best result is 0.8948, which is achieved with resolution $R = 2$. This significantly exceeds 0.8262 obtained by SIHKS. The proposed method with other resolutions are very convincing as well, as shown in Table 3.2. The behavior with ISPM on this database is kind of surprising. The best accuracy of the SGWS is obtained at ł = 4, and the performance rises to 0.9526, which is much higher than all the signatures in the literature within the framework of diffusion geometry. It can also be observed that the improvement gains of proposed shape signature via ISPM are also higher than SIHKS and HKS. This is consistent with the results on SHREC 2010.

TABLE 3.2: Performance (DCG) of different spectral descriptors on SHREC 2011 with intrinsic spatial pyramid matching.

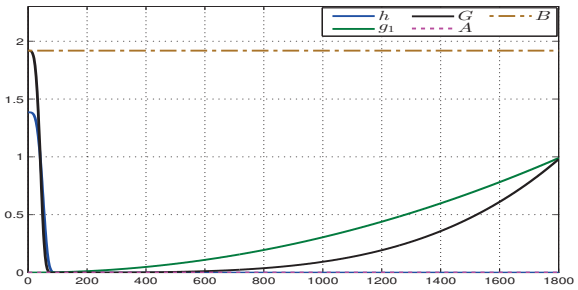
Level l (Partitions)	SIHKS	HKS	WKS	Spectral Graph Wavelet Signature (R)					
				1	2	3	4	5	6
0 (1)	0.8262	0.8114	0.6801	0.8043	**0.8948**	0.8536	0.8731	0.8613	0.8633
1 (2)	0.8436	0.8277	0.7097	0.8419	0.9208	0.8858	0.9002	0.8873	0.8911
4 (16)	0.8671	0.8721	0.7933	0.9203	0.9508	0.9443	**0.9526**	0.9471	0.9517
7 (128)	0.8771	0.8878	0.8042	0.9132	0.9427	0.9366	0.9383	0.9344	0.9359
9 (512)	0.8793	0.8902	0.8029	0.8982	0.9344	0.9241	0.9319	0.9277	0.9272
Improvement	0.0531	0.0788	0.1241	0.1160	0.0560	0.0907	0.0795	0.0858	0.0884

Figure 3.5 shows the object classes improved or confused by our proposed method. We compare our shape signature with the state-of-the-art approaches under both BoF and ISPM. Our method succeeds on most of the classes, such as Birds, Camel, Cat and Dinosaur, etc. The least successful or failed classes include Ant, Hand, Horse and Flamingo, etc.

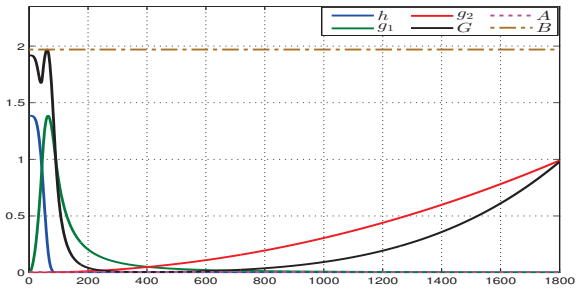


(a) Heat kernel

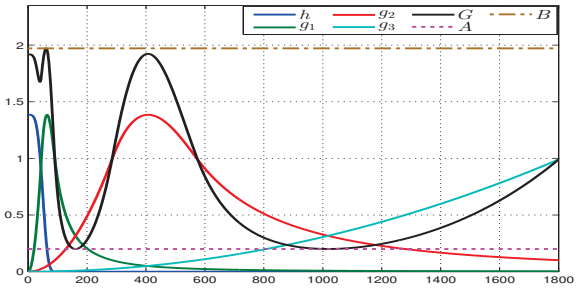
(b) Wave kernel



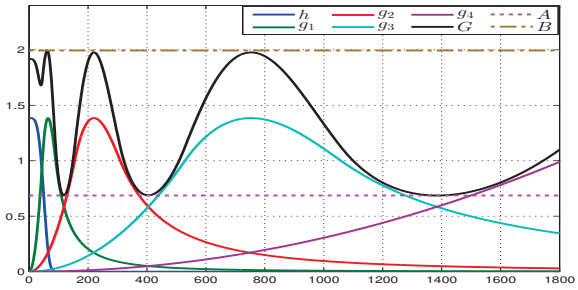
(c) Cubic spline kernel for $R = 1$



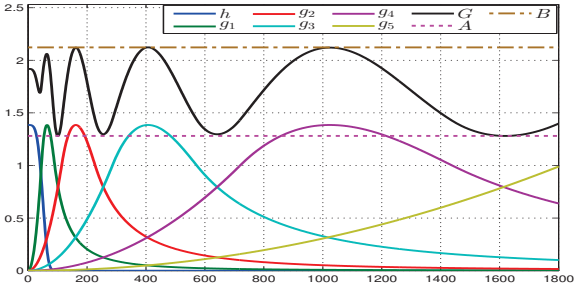
(d) Cubic spline kernel for $R = 2$



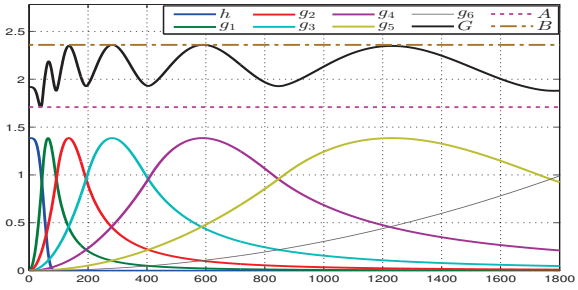
(e) Cubic spline kernel for $R = 3$



(f) Cubic spline kernel for $R = 4$



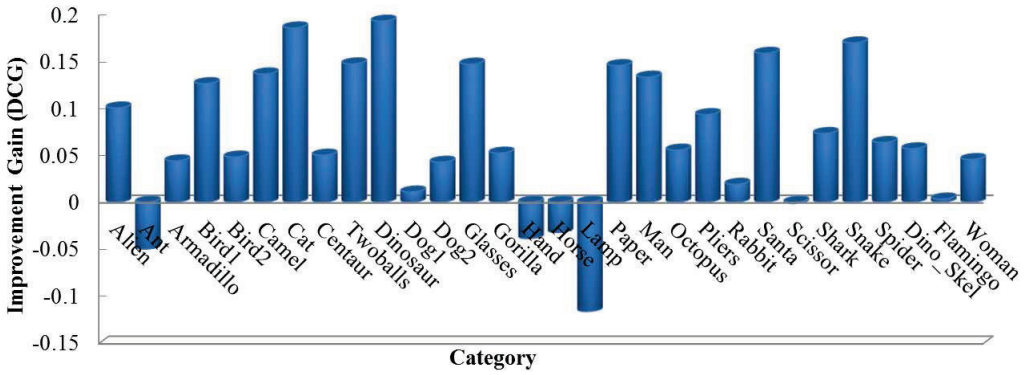
(g) Cubic spline kernel for $R = 5$



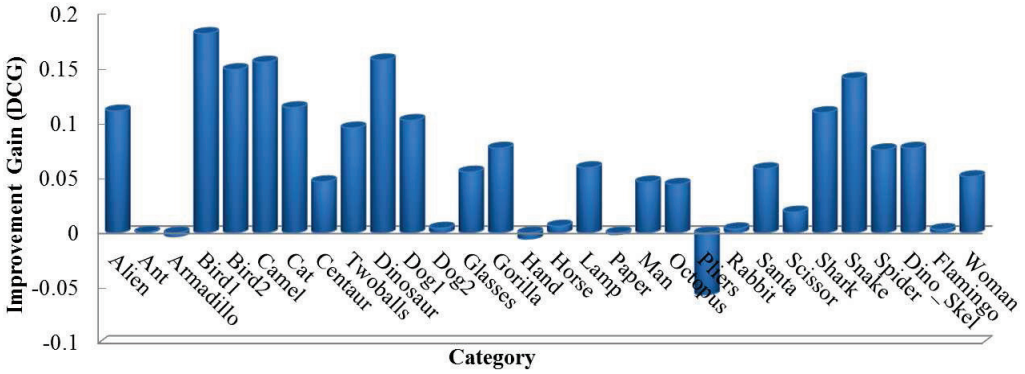
(h) Cubic spline kernel for $R = 6$

FIGURE 3.3: Spectrum modulation using different kernel functions at various resolutions. The dark line is the squared sum function G, while the dash-dotted and the dotted lines are upper and lower bounds (B and A) of G, respectively.

FIGURE 3.4: Normalized χ^2-distance between a reference point (yellow colored on the horse's back left leg) and other surface points using different signatures. Top row: heat kernel. Middle row: wave kernel. Bottom row: cubic spline wavelet kernel with $R = 2$.



(a) Comparison via BoF. Proposed SGWS (with $R = 2$) vs. SIHKS.



(b) Comparison via ISPM. Proposed SGSW (with $R = 4$ and ł $= 4$) vs. HKS (with ł $= 9$)

FIGURE 3.5: Improvement gain of the proposed shape signature over HKS and SIHKS.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

This thesis has presented spectral geometric approaches for nonrigid 3D shape retrieval using the Laplace-Beltrami operator and the graph wavelet transform. We have demonstrated through extensive experiments the much better performance of the proposed methods in comparison with existing techniques in the literature.

In Section 4.1, the contributions made in each of the previous chapters and the concluding results drawn from the associated research work are presented. Suggestions for future research directions related to this thesis are also provided in Section 4.2.

4.1 CONTRIBUTIONS OF THE THESIS

4.1.1 INTRINSIC SPATIAL PARTITION MATCHING

We reviewed and compared five recent spectral descriptors and varied codebook ambiguity models for nonrigid 3D shape retrieval. It turns out that SIHKS is the most discriminative spectral descriptor, and that codeword uncertainty yields the best ambiguity modeling ability between codewords, without taking codeword plausibility into account. By integrating the spatial cues with our proposed intrinsic partition approach, the retrieval performance was significantly improved. The intrinsic spatial pyramid matching, which works by repeatedly partitioning the surface and computing histograms of spectral descriptors over the resulting subpatches, showed promising results on two standard benchmarks of 3D models.

4.1.2 Spectral Graph Wavelet Signature

We introduced a graph wavelet framework for generalized spectral shape signatures. Our proposed approach provides a general and flexible framework to design shape descriptors for specific applications. By concentrating on finding informative spectrum for 3D shape retrieval, we devised a surface representation that is multiresolution, compact, highly discriminative, and parameter-insensitive. We also demonstrated through extensive experiments the effectiveness of the SGWS by achieving state-of-the-art results on two standard repositories of 3D shapes.

4.2 Future Research Directions

Several interesting research directions, motivated by this thesis, are discussed below:

4.2.1 Unifying Topology and Geometry

Viewed from the Morse-theoretic perspective, the eigenfunctions of LB operator capture the topological features of shapes. One potential research direction is to extract the skeleton Reeb graph of a 3D shape guided by these eigenfunctions. Additionally, integrating the surface geometry into the shape skeleton would provide a unified intrinsic framework of both topology and geometry for deformable 3D shape retrieval.

4.2.2 Applications of the Global Intrinsic Coordinate System

Using the second eigenfunction of the LB operator as a global intrinsic coordinate system, intrinsic versions of many other aggregation-based compact representations popular in image analysis, such as Fisher vector, can be designed. We intend to explore these constructions in our future work. Moreover, other applications like texture transfer among isometric shapes can also be improved by the intrinsic coordinate.

4.2.3 Design of Wavelet Generating Kernels

In its current form, the proposed SGWS is generated using a cubic spline kernel, and it has been shown to yield superior performance only with isometric or near-isometric transformations. In the future, we will look more carefully into the optimal choice of other wavelet generating kernel functions, thus extending the scope of the SGWS to more general classes of deformations. Additionally, designing appropriate

signatures for other shape analysis applications such as surface denoising is a promising future work direction that we plan to explore.

4.2.4 From Image Processing to Geometry Processing

Generally speaking, this thesis provides two bridges to borrow ideas from image processing for geometry processing, namely the intrinsic global coordinate system and the wavelet framework for shape descriptors' design. Abstractly, it generalizes methods in the Euclidean space to the weighted graph space, resulting in a fruitful way to understand 3D shapes by extending sophisticated methods in image domain via these tools. Our future plan is to explore other tools to link these two fields, such as finding a proper generalization of sparse coding and low rank matrix recovery based methods in the image domain for 3D surfaces.

REFERENCES

[1] Y. Yang, H. Lin, and Y. Zhang, "Content-based 3-D model retrieval: A survey," *IEEE Trans. Systems, Man, and Cybernetics, Part C*, pp. 1081-1098, 2007.

[2] A.D. Bimbo and P. Pala, "Content-based retrieval of 3D models". *TOMCCAP*, pp. 20-43, 2006.

[3] J.W.H. Tangelder and R.C. Veltkamp, "A survey of content based 3D shape retrieval methods," *Multimedia Tools Appl.*, pp. 441-471, 2008.

[4] B. Bustos, D.A. Keim, D. Saupe, T. Schreck, and D.V. Vranic, "Feature-based similarity search in 3D object databases," *ACM Comput. Surv*, pp. 345-387, 2005.

[5] V. Jain and H. Zhang, "A spectral approach to shape-based retrieval of articulated 3D models". *Computer-Aided Design*, 2007, pp. 398-407.

[6] D. Macrini, K. Siddiqi, and S.J. Dickinson, "From skeletons to bone graphs: Medial abstraction for object recognition", *Proc. CVPR*, 2008.

[7] K. Siddiqi, J. Zhang, D. Macrini, A. Shokoufandeh, S. Bouix, and S.J. Dickinson, "Retrieving articulated 3-D models using medial surfaces". *Mach. Vis. Appl.*, 2008, pp. 261-275.

[8] K. Siddiqi and S. Pizer, Eds., *Medial representations: mathematics, algorithms and applications*. Springer, 2008.

[9] C. Li and A. Ben Hamza, "Skeleton path based approach for nonrigid 3D shape analysis and retrieval". *Proc. IWCIA'2011*, LNCS, pp. 84-95.

[10] W. Mohamed and A. Ben Hamza, "Reeb graph path dissimilarity for 3D object matching and retrieval", *The Visual Computer*, 2012, pp. 305-318.

[11] J. Sun, M. Ovsjanikov, and L.J. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion", *Comput. Graph. Forum*, 2009, pp. 1383-1392.

[12] I. Kokkinos, M. M. Bronstein, and A. Yuille, "Dense scale-invariant descriptors for images and surfaces". *Research Report*, INRIA RR-7914, 2012, Mar.

[13] Y. Fang, M. Sun, M. Kim, and K. Ramani, "Heat-mapping: A robust approach toward perceptually consistent mesh segmentation". *Proc. CVPR*, 2011, pp. 2145-2152.

[14] M. Aubry, U. Schlickewei, and D. Cremers, "The wave kernel signature: A quantum mechanical approach to shape analysis". *Proc. Computational Methods for the Innovative Design of Electrical Devices*, 2011, pp. 1626-1633.

[15] R.M. Rustamov, "Laplace-Beltrami eigenfunctions for deformation invariant shape representation". *Proc. Symposium on Geometry Processing*, 2007, pp. 225-233.

[16] M. Reuter, F. Wolter, and N. Peinecke, "Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids". *Computer-Aided Design*, 2006, pp. 342-366.

[17] Y. Fang, M. Sun, and K. Ramani, "Temperature distribution descriptor for robust 3D shape retrieval". *Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment, CVPR*, 2011.

[18] R. EL Khoury, J.-P. Vandeborre, and M. Daoudi: Indexed heat curves for 3D-model retrieval. In: Proc. ICPR (2012).

[19] S. Rosenberg, *The Laplacian on a Riemannian Manifold*, Cambridge University Press, 1997.

[20] A.M. Bronstein, M.M. Bronstein, R. Kimmel, *Numerical geometry of non-rigid shapes*, Springer, 2008.

[21] D.G. Lowe, "Distinctive image features from scale-invariant keypoints". *International Journal of Computer Vision*, 2004, pp. 91-110.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection". *Proc. CVPR*, 2005, pp. 886-893.

[23] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge". *International Journal of Computer Vision*, 2010, pp. 303-338.

[24] J. Fehr and A. Streicher and H. Burkhardt, "A bag of features approach for 3D shape retrieval". *Advances in Visual Computing*, 2009, pp. 34-43.

[25] Z. Lian, A. Godil, and X. Sun, “Visual similarity based 3D shape retrieval using bag-of-features”. *Proc. SMI*, 2010, pp. 25-36.

[26] H. Tabia, O. Colot, M. Daoudi, and J.P. Vandeborre, “Non-rigid 3D shape classification using bag-of-feature techniques”. *Proc. ICME*, 2011, pp. 1-6.

[27] P. Bérard, G. Besson and S. Gallot, “Embedding Riemannian manifolds by their heat kernel”. *Geometric and Functional Analysis*, 1994.

[28] I. Kokkinos, M.M. Bronstein, R. Litman, and A.M. Bronstein, “Intrinsic shape context descriptors for deformable shapes”. *Proc. CVPR*, 2012, 159-166.

[29] I. Pratikakis, M. Spagnuolo, T. Theoharis, and R. Veltkamp, “A robust 3D interest points detector based on Harris operator”. *Proc. Eurographics Workshop on 3DOR*, 2010.

[30] A. Zaharescu, E. Boyer, K. Varanasi and R. Horaud, “Surface feature detection and description with applications to mesh matching”. *Proc. CVPR*, 2009, pp. 373-380.

[31] X. Li and A. Godil, “Exploring the Bag-of-Words method for 3D shape retrieval”. *Proc. ICIP*, 2009, pp. 437-440.

[32] C. Redondo-Cabrera, and R.J. López-Sastre, and J. Acevedo-Rodriguez and S. Maldonado-Bascón, “SURFing the point clouds: Selective 3D spatial pyramids for category-level object recognition”. *Proc. CVPR*, 2012, pp. 3458-3465.

[33] A.E. Abdel-Hakim and A. Farag, “CSIFT: A SIFT descriptor with color invariant characteristics”. *Proc. CVPR*, 2006, pp. 1978-1983.

[34] A.M. Bronstein, M.M. Bronstein, L.J. Guibas, and M. Ovsjanikov, “Shape google: Geometric words and expressions for invariant shape retrieval”. *ACM Trans. Graph.*, 2011, pp. 1-1.

[35] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories”. *Proc. CVPR*, 2006, pp. 2169-2178.

[36] Y. Shi, R. Lai, S. Krishna, I. Dinov, and A.W. Toga, “Anisotropic Laplace-Beltrami eigenmaps: Bridging reeb graphs and skeletons”. *Proc. IEEE CVPR, Workshops*, Anchorage, AK, 2008, pp. 1-7.

[37] T.A. Funkhouser, P. Min, M.M. Kazhdan, J. Chen, J.A. Halderman, D.P. Dobkin, and D.P. Jacobs, “A search engine for 3D models”. *ACM Trans. Graph.*, 2003, pp. 83-105.

[38] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts". *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, pp. 509-522.

[39] M. Kortgen, G.-J. Park, M. Novotni, and R. Klein, "3D shape matching with 3D shape contexts". *7th Central European Seminar on Computer Graphics*, 2003.

[40] R. Osada, T.A. Funkhouser, B. Chazelle, and D.P. Dobkin, "Shape distributions". *ACM Trans. Graph.*, 2002, pp. 807-832.

[41] M.M. Kazhdan, T.A. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors". *In Proc. Symposium on Geometry Processing*, 2003, pp. 156-165.

[42] A. Elad and R. Kimmel, "On bending invariant signatures for surfaces". *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1285-1295, 2003.

[43] A. Ben Hamza and H. Krim, "Geodesic matching of triangulated surfaces". *IEEE Trans. Image Processing*, 2006, pp. 2249-2258.

[44] V. Jain, H. Zhang, and O.V. Kaick, "Non-rigid spectral correspondence of triangle meshes", *International Journal of Shape Modeling*, 2007, pp. 101-124.

[45] R. Coifman and S. Lafon. "Diffusion maps". *Applied and Computational Harmonic Analysis*, 2006

[46] B. Levy. "Laplace-Beltrami eigenfunctions: Towards an algorithm that understands geometry". *Proc. IEEE Int. Conf. on Shape Modeling and Applications*, 2006, pp. 13-20.

[47] M.M. Bronstein and A.M. Bronstein, "Shape recognition with spectral distances". *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, pp. 1065-1071.

[48] F. Fouss, A. Pirotte, J. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation". *IEEE Trans. Knowl. Data Eng.*, 2007, pp. 355-369.

[49] Y. Lipman, R.M. Rustamov, and T.A. Funkhouser, "Biharmonic distance". *ACM Trans. Graph.*, 2010.

[50] H. Zhang, O.V. Kaick, and R. Dyer, "Spectral mesh processing". *Comput. Graph. Forum*, 2010, pp.1865-1894.

[51] J. Sivic and A. Zisserman, “Video Google: A text retrieval approach to object matching in videos”. *Proc. ICCV*, 2003, pp. 1470-1477.

[52] J.V. Gemert, C.G.M. Snoek, C.J. Veenman, A.W.M. Smeulders, and J. Geusebroek, “Comparing compact codebooks for visual categorization”. *Computer Vision and Image Understanding*, 2010, pp. 450-462.

[53] O. Boiman, E. Shechtman, and M. Irani, “In defense of nearest-neighbor based image classification”. *Proc. CVPR*, 2008.

[54] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Lost in quantization: Improving particular object retrieval in large scale image databases”. *Proc. CVPR*, 2008.

[55] J.V. Gemert, C.J. Veenman, A.W.M. Smeulders, and J. Geusebroek, “Visual word ambiguity”. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, pp. 1271-1283.

[56] J. Yang, K. Yu, Y. Gong, and T.S. Huang, “Linear spatial pyramid matching using sparse coding for image classification”. *Proc. CVPR*, 2009, pp. 1794-1801.

[57] J. Wang, J. Yang, K. Yu, F. Lv, T.S. Huang, and Y. Gong, “Locality-constrained linear coding for image classification”, *Proc. CVPR*, 2010, pp. 3360-3367.

[58] F. Perronnin and C.R. Dance, “Fisher kernels on visual vocabularies for image categorization”. *Proc. CVPR*, 2007.

[59] H. Jégou, M. Douze, C. Schmid, and P. Perez, “Aggregating local descriptors into a compact image representation”. *Proc. CVPR*, 2010, pp. 3304-3311.

[60] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, “Aggregating local images descriptors into compact codes”. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9. pp. 1704-1716, 2012.

[61] D. Picard and P. Gosselin, “Improving image similarity with vectors of locally aggregated tensors”. *Proc. ICIP*, 2011, pp. 669-672.

[62] A.M. Bronstein and M.M. Bronstein, “Spatially-sensitive affine-invariant image descriptors”, *Proc. ECCV*, 2010, pp. 197-208.

[63] S. Savarese, J.M. Winn, and A. Criminisi, “Discriminative object class models of appearance and shape by correlatons”, *Proc. CVPR*, 2006, pp. 2033-2040.

[64] H. Ling and S. Soatto, ”Proximity distribution kernels for geometric context in category recognition", *Proc. ICCV*, 2007, pp. 1-8.

[65] R. Behmo, N. Paragios, and V. Prinet, “Graph commute times for image representation”. *Proc. CVPR*, 2008.

[66] D. Liu, G. Hua, P.A. Viola, and T. Chen, ”Integrated feature selection and higher-order spatial feature extraction for object categorization”, *Proc. CVPR*, 2008.

[67] K. Grauman and T. Darrell, “The pyramid match kernel: Discriminative classification with sets of image features”. *Proc. ICCV*, 2005, pp. 1458-1465.

[68] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang, “Spatial-bag-of-features”. *Proc. CVPR*, 2010, pp. 3352-3359.

[69] Y. Yang and S. Newsam, “Spatial pyramid co-occurrence for image classification”. *Proc. ICCV*, 2011, pp. 1465-1472.

[70] Y. Zhang, Z. Jia, and T. Chen, "Image retrieval with geometry-preserving visual phrases”. *Proc. CVPR, 2011*, pp. 809-816.

[71] Y. Jia, C. Huang, and T. Darrell, “Beyond spatial pyramids: Receptive field learning for pooled image features”. *Proc. CVPR*, 2012.

[72] J. Krapac, J.J. Verbeek, and F. Jurie, “Modeling spatial layout with fisher vectors for image categorization”. *Proc. ICCV*, 2011, pp. 1487-1494.

[73] M. Meyer, M. Desbrun, P. Schröder, A. Barr, “Discrete differential geometry operators for triangulated 2-manifolds”. *Proc. Visual Mathematics*, 2002.

[74] M. Wardetzky, S. Mathur, F. Kälberer, and E. Grinspun, “Discrete Laplace operators: no free lunch”. *Proc. Conf. Computer Graphics and Interactive Techniques*, 2008.

[75] M. Belkin, J. Sun, and Y. Wang: Discrete laplace operator on meshed surfaces. In: Proc. SCG, 278-287 (2008).

[76] K. Hildebrandt and K. Polthier: On approximation of the Laplace-Beltrami operator and the Willmore energy of surfaces. Computer Graphics Forum 30(5), 1513-1520 (2011).

bibitemA:multi-resolution A. Vaxman, M. Ben-Chen, and C. Gotsman, "A multi-resolution approach to heat kernels on discrete surfaces", *ACM Trans. Graph.*, 2010.

[77] E.B. Davies and Y. Safarov (Eds.), *Spectral Theory and Geometry*, Cambridge Univer. Press, 1999.

[78] K. Uhlenbeck, "Generic properties of eigenfunctions". *American Journal of Mathematics*, vol. 98, no. 4, pp. 1059-1078, 1976.

[79] K. Jarvelin and J. Kekalainen, "IR evaluation methods for retrieving highly relevant documents". *In 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2000.

[80] Z. Lian, A. Godil, T. Fabry, T. Furuya, J. Hermans, R. Ohbuchi, C. Shu, D. Smeets, P. Suetens, D. Vandermeulen, and S. Wuhrer, "SHREC'10 Track: Non-rigid 3D Shape Retrieval". *Proc. 3DOR*, 2010, pp. 101-108.

[81] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoue, H.V. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F.P.M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, and D. Vandermeulen, "SHREC '11 Track: Shape Retrieval on Non-rigid 3D Watertight Meshes". *Proc. 3DOR*, 2011, pp. 79-88.

[82] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," *Proc. Shape Modeling Applications*, pp. 167-178, 2004.

[83] B. Li, A. Godil, M. Aono, X. Bai, T. Furuya, L. Li, R. López-Sastre, H. Johan, R. Ohbuchi, C. Redondo-Cabrera, A. Tatsuma, T. Yanagimachi, and S. Zhang, "SHREC'12 Track: Generic 3D shape retrieval," *Proc. Eurographics Workshop on 3D Object Retrieval*, pp. 119-126, 2012.

[84] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: a geometric framework for learning from labeled and unlabeled examples," *Journal of Machine Learning Research*, vol. 7, pp. 2399Ű2434, 2006.

[85] M. Reuter, "Hierarchical shape segmentation and registration via topological features of Laplace-Beltrami eigenfunctions," *Int. Jour. Computer Vision*, vol. 89, no. 2, pp. 287-308, 2010.

[86] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," *Proc. ACM Sympo. Geometry Processing*, pp. 156-164, 2003.

[87] K. Gębal, J. A. Bærentzen, H. Aanæs, and R. Larsen, "Shape analysis using the auto diffusion function," *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1405-1513, 2009.

[88] A.M. Bronstein, "Spectral descriptors for deformable shapes," *CoRR abs/1110.5015*, 2011.

[89] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, 2008.

[90] R. Coifman and M. Maggioni, "Diffusion wavelets," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 53-94, 2006.

[91] D.K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129-150, 2011.

[92] W. H. Kim, D. Pachauri, C. Hatt, M. K. Chung, S. C. Johnson and V. Singh, "Wavelet based multi-scale shape features on arbitrary surfaces for cortical thickness discrimination," *Proc. NIPS*, pp. 1135-1143, 2012.

[93] M. Meyer, M. Desbrun, P. Schröder, A. Barr, "Discrete differential-geometry operators for triangulated 2-manifolds," *Visual Mathematics III*, pp. 35-57, 2003.