REGULAR PAPER

Intrinsic spatial pyramid matching for deformable 3D shape retrieval

Chunyuan Li · A. Ben Hamza

Abstract   In this paper, we present an intrinsic spatial pyramid matching approach for 3D shape retrieval. Motivated by the fact that the second eigenfunction of Laplace–Beltrami operator not only can capture the global topological structure information, but also is intrinsic, we propose to adopt its level sets as cuts to perform surface partition. The resulting matching scheme is able to consistently estimate the approximate global geometric correspondence among 3D shapes. In particular, we can leverage recent developments in intrinsic shape analysis and perform intrinsic spatial pyramid matching based on dense spectral shape descriptors such as scale-invariant heat kernel signature. Our experiments demonstrate a significant improvement of 3D shape retrieval on two standard benchmarks.

Keywords   Intrinsic partition · Diffusion geometry · Eigenfunction · Shape retrieval

1 Introduction

State-of-the-art image recognition algorithms 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 the scale-invariant feature transform (SIFT) [1] or the histogram of oriented gradients (HOG) [2], and encode them into an overcomplete representation using various algorithms such as the k-means or sparse coding. After coding, global image representations are formed by spatially pooling the coded local descriptors. The methods following such a

C. Li · A. B. Hamza (✉)
Concordia Institute for Information Systems Engineering,
Concordia University, Montreal, QC, Canada
e-mail: hamza@ciise.concordia.ca

pipeline have achieved competitive performance on image classification tasks [3]. During the whole procedure, the spatial pooling step brings a substantial performance improvement. One significant milestone in the construction of this arsenal of tools is the spatial pyramid matching (SPM) introduced in [4], which partitions the image into increasingly fine subregions and then computes histograms of local features found inside each subregion. The empirical success of this technique stems from 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.

The codebook model, as a simplified version of such a pipeline without spatial pooling, has been also considered for 3D shapes. The basic idea of using a codebook to represent a shape as histograms of occurrences of visual words is commonly referred to in the literature as Bag-of-Words (BoW) or Bag-of-Features (BoF) approach. Several authors have introduced such BoF approaches for 3D shape retrieval. Indeed, early research has mainly dealt with the global Euclidean transformations (rigid motion) [5] and multiple views [6]. By defining the visual words on the segmented shape regions, Toldo et al. [7] obtained encouraging shape categorization and retrieval results. Darom et al. [8] achieved state-of-the-art retrieval accuracy by designing the local vertex-wise features, which are robust to scale changes and partial mesh matching. The codebook model has been shown to be a promising method for partial shape retrieval [8,9]. Recent efforts have also focused on finding the deformation invariance for nonrigid shapes by replacing the Euclidean metric with its geodesic counterpart [10]. The geodesic distance, however, suffers from strong sensitivity to topological noise, which limits its usefulness in real applications.

This problem is well handled by the tools from the emerging field of diffusion geometry, which provides a generic

framework for many intrinsic methods in the analysis of geometric shapes. Diffusion geometry formulates the heat diffusion processes on manifolds. Coifman and Lafon [11] introduced invariant metrics known as diffusion distances, which correspond to the L_2-norm difference of energy distribution between two points initialized with unit impulse functions after a given time. The diffusion distance is more robust to topological noise than geodesic one. Reuter et al. [12] adopted the eigenvalues of the Laplace–Beltrami (LB) operator to construct a global shape descriptor, called ShapeDNA. Based on the theoretical works in [13], Lévy [14] showed that the eigenfunctions of LB operator can be well adapted to the geometry and the topology of an object. Later, several spectral descriptors were proposed to characterize the geometric features of a 3D surface [15–17]. By aggregating these spectral descriptors, the Shape Google algorithm [18,19] was proposed as a classical method for deformable shape retrieval. It uses the multiscale diffusion heat kernels as "geometric words", and constructs a compact and informative shape representation by means of the codebook approach.

More recently, there have been several attempts to adapt 2D planar shape contexts [20], popular image feature detectors [21] and descriptors [22], to 3D surfaces. This line of works partially inspires our proposed approach. Another inspiration is due in part to the great success of SPM in the image domain. Spatially enhanced techniques for 3D shape recognition were explored earlier in [23,24], but these works are not intrinsic, i.e., shape deformations affect the descriptors. "Geometric expressions" [18] was an earlier work that explored the exploitation of intrinsic geometry, but the authors only dealt with the local relative spatial position, by considering the diffusion distance between pairwise vertices. Our approach, on the other hand, models the global absolute spatial positions, which allow us to retain and exploit the information contained in the whole 3D shape.

Our contributions are threefold: (1) we propose to adopt the second eigenfunction of the LB operator in a bid to construct a global surface coordinate system, which is insensitive to shape deformation, (2) we develop a proper generalization of the SPM for surfaces and show a numerical way to construct it, and (3) we experimentally demonstrate that introducing the global spatial context significantly improves the discriminative power of the descriptor in 3D matching and retrieval.

The rest of this paper is organized as follows. Section 2 provides a brief background on the LB operator, its discretization and eigenanalysis, followed by the codebook model. In Sect. 3, we propose the intrinsic spatial pyramid matching (ISPM) approach. Experimental results on two 3D datasets are presented in Sect. 4. Finally, we conclude and point out future work directions in Sect. 5.

2 Background

2.1 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) $\boldsymbol{x}$ defined from a connected open set (parametrization domain) $U \subset \mathbb{R}^2$ to $\mathbb{M} \subset \mathbb{R}^3$ such that

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

where $\boldsymbol{u} = (u^1, u^2) \in U$. Note that the components of $\boldsymbol{x}$ and $\boldsymbol{u}$ are denoted by superscripts in place of subscripts. This superscript convention stems from the use of tensor notation which greatly simplifies the formalism of the theory of surfaces [25].

Given a twice-differentiable function $f : \mathbb{M} \to \mathbb{R}$, the LB operator [26] is a second-order partial differential operator defined as

$$\Delta_{\mathbb{M}} f = -\frac{1}{\sqrt{|g|}} \sum_{i,j=1}^{2} \frac{\partial}{\partial u^j} \left(\sqrt{|g|}\, g^{ij} \frac{\partial f}{\partial u^i} \right)$$

$$= -\sum_{i,j=1}^{2} g^{ij} \frac{\partial}{\partial u^j} \frac{\partial f}{\partial u^i} + \text{(lower order terms)} \tag{2}$$

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. Furthermore, 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.

2.2 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 $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ is the set of vertices, $\mathcal{E} = \{e_{ij}\}$ is the set of edges, and $\mathcal{T} = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n\}$ is the set of triangles. Each edge e_{ij} (denoted by $[\boldsymbol{v}_i, \boldsymbol{v}_j]$ or simply $[i, j]$) connects a pair of vertices $\{\boldsymbol{v}_i, \boldsymbol{v}_j\}$. Two distinct vertices $\boldsymbol{v}_i, \boldsymbol{v}_j \in \mathcal{V}$ are adjacent (denoted by $\boldsymbol{v}_i \sim \boldsymbol{v}_j$ 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 $\boldsymbol{v}_i$ is the set $\boldsymbol{v}_i^\star = \{\boldsymbol{v}_j \in \mathcal{V} : i \sim j\}$.

Several discretizations of the LB operator are available in the literature. In this paper, we use the approach devel-

oped in [27], which employs a mixed finite element/finite volume method on triangle meshes. Hence, the value of $\Delta_{\mathbb{M}} f$ at a vertex $\boldsymbol{v}_i$ (or simply i) can be approximated using the cotangent weight scheme:

$$\Delta_{\mathbb{M}} f(i) \approx \frac{1}{a_i} \sum_{j \sim i} \frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} \left(f(j) - f(i) \right) \qquad (3)$$

where α_{ij} and β_{ij} are the angles $\angle(\boldsymbol{v}_i \boldsymbol{v}_{k_1} \boldsymbol{v}_j)$ and $\angle(\boldsymbol{v}_i \boldsymbol{v}_{k_2} \boldsymbol{v}_j)$ of two faces $\boldsymbol{t}^\alpha = \{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_{k_1}\}$ and $\boldsymbol{t}^\beta = \{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_{k_2}\}$ that are adjacent to the edge $[i, j]$, and a_i is the area of the Voronoi cell. 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 [28].

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

$$\omega_{ij} = \begin{cases} \frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2a_i} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \qquad (4)$$

Then, for a function $f : \mathcal{V} \to \mathbb{R}$ that assigns to each vertex $i \in \mathcal{V}$ a real value $f(i)$ (we can view f as a column vector of length m), we may write the LB operator given by Eq. (3) as

$$Lf(i) = \sum_{j \sim i} \omega_{ij} \left(f(i) - f(j) \right), \qquad (5)$$

where the matrix $L = (\ell_{ij})$ is given by

$$\ell_{ij} = \begin{cases} d_j & \text{if } i = j \\ -\omega_{ij} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \qquad (6)$$

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

2.3 Eigenanalysis and spectral signatures

Note that $\omega_{ij} \neq \omega_{ji}$ implies L is not a symmetric matrix. Thus, the spectrum (set of eigenvalues) of the eigenvalue problem $L\boldsymbol{\varphi}_i = \lambda_i \boldsymbol{\varphi}_i$ may not be real [29]. Noting that $\omega_{ij} = \kappa_{ij}/a_i$, where

$$\kappa_{ij} = \begin{cases} \frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \qquad (7)$$

we may factorize the matrix L as $L = A^{-1}C$, where $A = \text{diag}(a_i)$ is a positive-definite diagonal matrix called *stiffness* matrix and $C = (c_{ij})$ is a sparse symmetric matrix referred to as *lumped mass* matrix, given by

$$c_{ij} = \begin{cases} \sum_{i=1}^m \kappa_{ij} & \text{if } i = j \\ -\kappa_{ij} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \qquad (8)$$

Therefore, we may write the eigenvalue problem $L\boldsymbol{\varphi}_i = \lambda_i \boldsymbol{\varphi}_i$ as a generalized problem

$$C\boldsymbol{\varphi}_i = \lambda_i A \boldsymbol{\varphi}_i, \quad i = 1, 2, \ldots, m \qquad (9)$$

The eigenvalues λ_i and associated eigenfunctions $\boldsymbol{\varphi}_i$ of the LB operator can be computed by solving the above generalized problem. That is, $\boldsymbol{\varphi}_i$ 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 $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_m$, where each eigenfunction $\boldsymbol{\varphi}_i = (\varphi_i(\boldsymbol{v}_1), \ldots, \varphi_i(\boldsymbol{v}_m))'$ is an m-dimensional vector. Note that the eigensystem $\{\lambda_i, \boldsymbol{\varphi}_i\}_i$ is intrinsic to the manifold and enjoys a nice property of being isometric invariant.

Based on the obtained eigenfunctions and eigenvalues, several spectral signatures have been proposed in the literature to describe a single vertex on a surface. Sun et al. [15] introduced the heat kernel signature (HKS) based on the fundamental solution of the heat equation (heat kernel). Its scale-invariant version (SIHKS) was developed in [17]. Another physically inspired descriptor is the wave kernel signature (WKS), which was proposed in [16]. Unlike the HKS, the WKS separates influences of different frequencies, treating all frequencies equally. These descriptors have been shown to achieve an excellent performance in 3D shape analysis and recognition.

2.4 Bag-of-feature model

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 geometric word vocabulary in the codebook model may be constructed in various ways, e.g., by approximate k-means [30] or hierarchical k-means [31]. We use the simple k-means method, which is also used in the Shape Google algorithm [19]. Thus, the "geometric words" of a vocabulary $P = \{\boldsymbol{p}_k, k = 1, 2, \ldots, K\}$ are obtained as the K centroid of k-means clustering in the signature space. From any shape, a specific type of local spectral descriptor $S = \{\boldsymbol{s}_t, t = 1, 2, \ldots, T\}$ is used for comparison. Obviously, each local descriptor $\boldsymbol{s}_t$ (represented as a vector) is associated with its nearest geometric word $NN(\boldsymbol{s}_t)$ in the codebook. By a certain vector coding technique, such as hard counting or ambiguity modeling, 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 essential for the codeword-cumulative histogram representation.

The *traditional codebook* is the histogram of the number of local descriptors assigned to each geometric word. For each codeword $\boldsymbol{p}_k$, the differences of vector $\boldsymbol{s}_t$ assigned to

p_k are accumulated by the L_0 norm as follows:

$$q_i = \sum_{s_t:NN(s_t)=i} \|s_t - p_i\|_0 \tag{10}$$

However, this may be disadvantageous because one local descriptor can be modeled better by multiple geometric words. The hard counting strategy in traditional codebook will result in quantization loss. Another similar method is the *codeword uncertainty* [32], in which one image region distributes its probability mass to more than one codeword. In the Shape Google algorithm, the authors also adopted this technique, which 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 Gaussian kernel values for all codewords p_i in the vocabulary

$$q_i = \sum_{t=1}^{T} \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\|s_t - p_i\|_2^2}{\sigma^2}}}{\sum_{k=1}^{K} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\|s_t - p_k\|_2^2}{\sigma^2}}} \tag{11}$$

Thus, in the traditional codebook, the local descriptor selects the best candidate geometric word, whereas the codeword uncertainty does not solely assign the descriptor to the best fitting word, but is also divided over multiple codewords. Both of them can be compared via the Chi-squared kernel. In the following section, we introduce an ISPM kernel.

3 Intrinsic spatial pyramid matching

3.1 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:

$$f = \sum_{i=1}^{m} \alpha_i \varphi_i, \quad \text{where} \quad \alpha_i = \langle f, \varphi_i \rangle. \tag{12}$$

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 Fig. 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

$$\lambda_2 = \inf_{f \perp \varphi_1} \frac{f'Cf}{f'Af} = \inf_{f \perp \varphi_1} \frac{\sum_{i \sim j} c_{ij}(f(v_i) - f(v_j))^2}{\sum_i f(v_i)^2 a_i} \tag{13}$$

Fig. 1 **a–c** 3D horse model colored by $\varphi_2, \varphi_3, \varphi_4$. **d–f** Level sets of $\varphi_2, \varphi_3, \varphi_4$
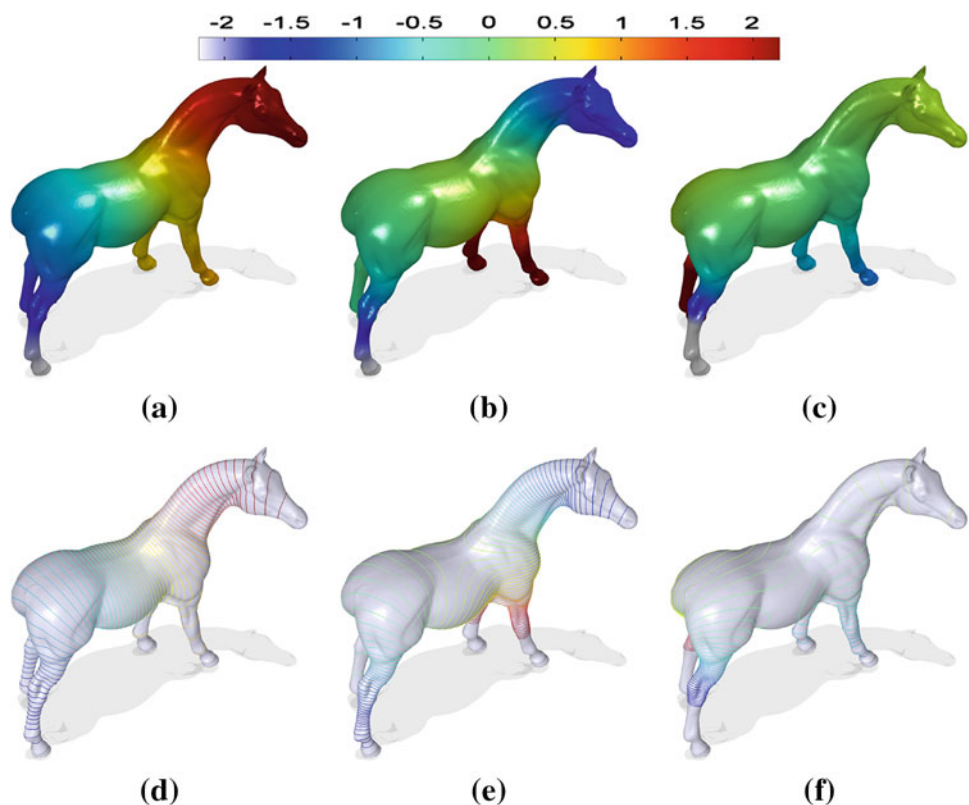
and $\boldsymbol{\varphi_2} = (\varphi_2(\boldsymbol{v}_1), \ldots, \varphi_2(\boldsymbol{v}_m))'$ is its corresponding eigenvector.

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 [33]. On the other hand, Uhlenbeck [34] 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 construction of the associated intrinsic isocurves.

3.2 Intrinsic spatial partition

Motivated by the invariance properties of the second eigenfunction of the LB operator and also by its generic property as a Morse function as well as by the fact that intuitively the second eigenvalue corresponds to the sound we hear the best [33], we propose to use the level sets (isocontours) of the second eigenfunction as cuts to partition a surface. In Fig. 2a–c, we show some examples of the level curves of φ_2. In Fig. 2a, 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 is displayed in Fig. 2b, which shows shapes that include various topological structures. Finally, the consistency of isocontours on the shapes from different classes is displayed in Fig. 2c. Although the shapes are explicitly different, their isocontours can capture their intrinsic correspondence well.

The level sets of the second eigenfunction have been previously used to extract curve skeletons of nonrigid shapes [35], which is a vivid clue that these isocontours capture the global topological structure of shapes.

3.3 Matching by intrinsic spatial partition

Instead of representing the whole shape by the codeword model without considering spatial layout of local descriptors, we propose to 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:

$$H = [h^1, h^2, \ldots, h^i, \ldots, h^R] \qquad (14)$$

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 Fig. 2a, 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:

$$T = [h^R, h^{(R-1)}, \cdots, h^i, \cdots, h^1]. \qquad (15)$$

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

$$\mathcal{B}^R(P, Q) = \min(\mathcal{A}^R(H_P, H_Q), \mathcal{A}^R(H_P, T_Q)) \qquad (16)$$

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 $\mathcal{A}^R(\cdot, \cdot)$ is defined as

$$\mathcal{A}^R(H_P, H_Q) = \sum_{i=1}^{R} \sum_{k=1}^{K} \Psi(h^i_P(k), h^i_Q(k)) \qquad (17)$$

where $\Psi(\cdot, \cdot)$ can be any histogram comparison metric. In this paper, we use the Chi-squared kernel so that $h^i_P(k)$ and $h^i_Q(k)$ 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 extend the spatial pyramid, 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 $\{R = 2^\ell, \ell = 0, \ldots, L\}$ 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

$$\mathcal{D}^L(P, Q) = \mathcal{B}^L(P, Q) + \sum_{\ell=0}^{L-1} \frac{1}{2^{L-\ell}}(\mathcal{B}^\ell(P, Q) - \mathcal{B}^{\ell+1}(P, Q))$$

$$= \frac{1}{2^L} \mathcal{B}^0(P, Q) + \sum_{\ell=1}^{L} \frac{1}{2^{L-\ell+1}} \mathcal{B}^\ell(P, Q) \qquad (18)$$

The weight associated with each level is set to $1/2^{L-\ell}$, which is inversely proportional to the cell width at that level. Intuitively, we want to penalize matches found in larger cells because they involve increasingly dissimilar features. Concerning the implementation, one issue that arises is that of normalization. To easily compare the methods of single-level partition and ISPM, we normalize the histogram of each resolution using the L_1-norm.

Fig. 2 a Isocontours are invariant under both global and local deformations. **b** Proportionality correspondence of pairwise nonrigid shapes with varied topological structure. **c** Isocontours are consistent among different classes of shapes.

(a)

(b)

(c)
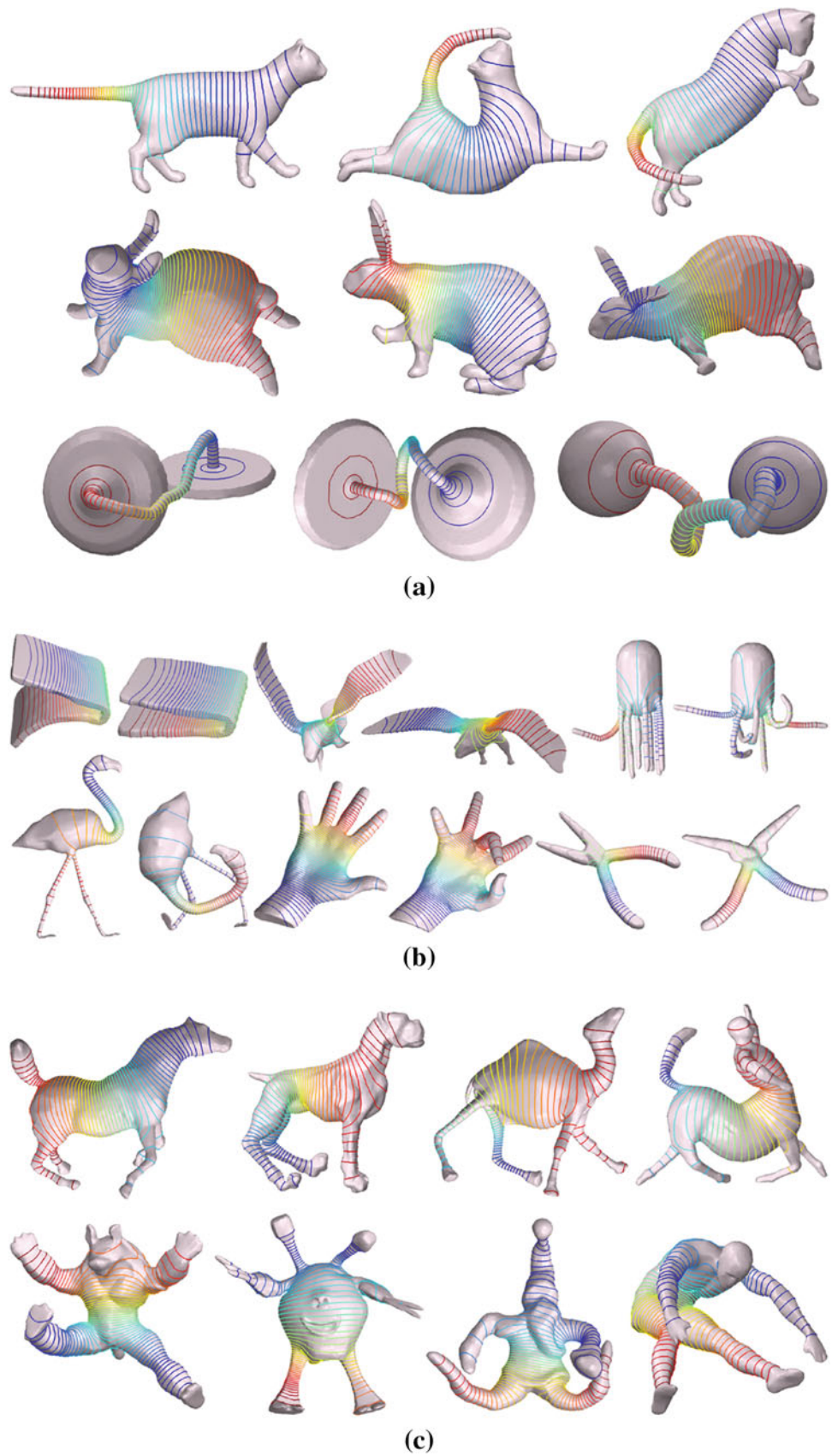
4 Experimental results

The performance of our proposed intrinsic spatial pyramid was evaluated on two datasets, namely SHREC 2011 Benchmark [36] and the TOSCA-based robust shape retrieval database used in [19]. The first dataset is used to validate the discriminative power of ISPM between different shape categories, and the second one is used to test the robustness of ISPM.

4.1 SHREC 2011 database

SHREC 2011 database 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 Fig. 3, we show two shapes of each class in the dataset. The retrieval performance is evaluated using the discounted cumulative gain (DCG) [37]. All normalized DCG calculations are relative values in the interval [0, 1]. Higher numbers are better, and the results are cross-query comparable.

We performed ISPM based on HKS and SIHKS dense descriptors. These descriptors showed excellent performance with the codebook model in the Shape Google algorithm. The

first 150 eigenvalues and eigenvectors of the LB operator on each shape are used. We experimentally select the best parameters for HKS and SIHKS on SHREC 2011 dataset. For HKS, we formulate the diffusion time as $t = t_0 \alpha^\tau$, where τ is sampled from 0 to a given *scale T* with a resolution $1/4$. $T = 5$, $t_0 = 0.01$ and $\alpha = 4$ are set in our case. In order to construct the SIHKS, we use $t = \alpha^\tau$, where τ ranges from 1 to a given scale with finer increments of $1/16$. $T = 25$ and $\alpha = 2$ are chosen. After applying the logarithm, derivative and Fourier transform, all the frequencies are used to obtain the best result.

The computation of the vocabulary is performed offline in advance. To confirm getting optimal results, the clustering is repeated three 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. The running time 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 2,000 faces for each shape, we have a set of approximately 6×10^5 descriptors. The vocabulary size is fixed as 32, the dimension of HKS is 21 and the dimension of SIHKS is 385. It is important to point out that we performed the BoF experiments on SHREC2011 with various codeword sizes of 8, 12, 16, 24,
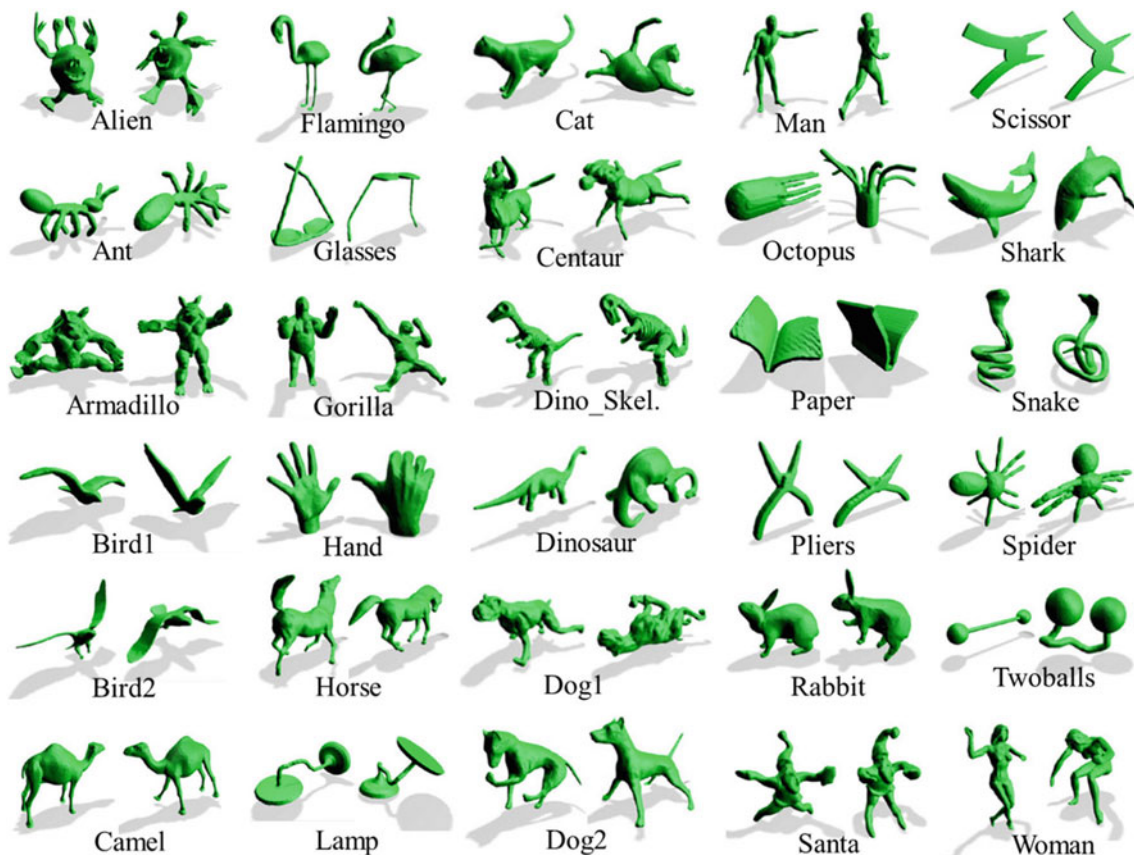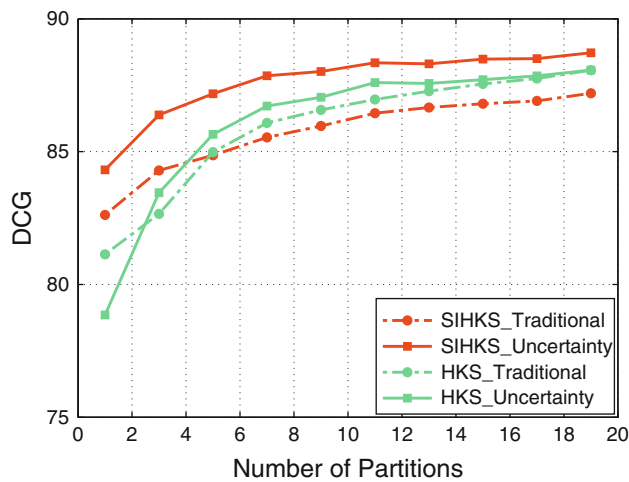

Fig. 3 Sample shapes in SHREC 2011 dataset

Fig. 4 Performance improvement by increasing the number of intrinsic partitions on SHREC 2011 dataset

Table 1 Performance (DCG) comparison of ISPM and single-level partition on SHREC 2011

Spectral descriptor	Level L (partitions)	Codebook Models			
		Traditional		Uncertainty	
		Single	Pyramid	Single	Pyramid
HKS	1 (2)	0.8345	0.8277	0.8497	0.8431
	4 (16)	0.8766	0.8721	0.8892	0.8870
	7 (128)	0.8893	0.8878	0.8892	**0.8903**
	9 (512)	0.8911	0.8902	0.8885	0.8891
SIHKS	1 (2)	0.8457	0.8436	0.8714	0.8661
	4 (16)	0.8688	0.8671	0.8862	0.8853
	7 (128)	0.8778	0.8771	0.8887	**0.8890**
	9 (512)	0.8798	0.8793	0.8888	0.8890

32, 48, 64, 80 and 200. It turns out that different descriptors attain the best results with size 32, and change slightly afterwards. As a result, we fixed the codeword size as 32 for SHREC2011. The running times to obtain the vocabulary for HKS and SIHKS are 1,043 and 9,033 s, respectively.
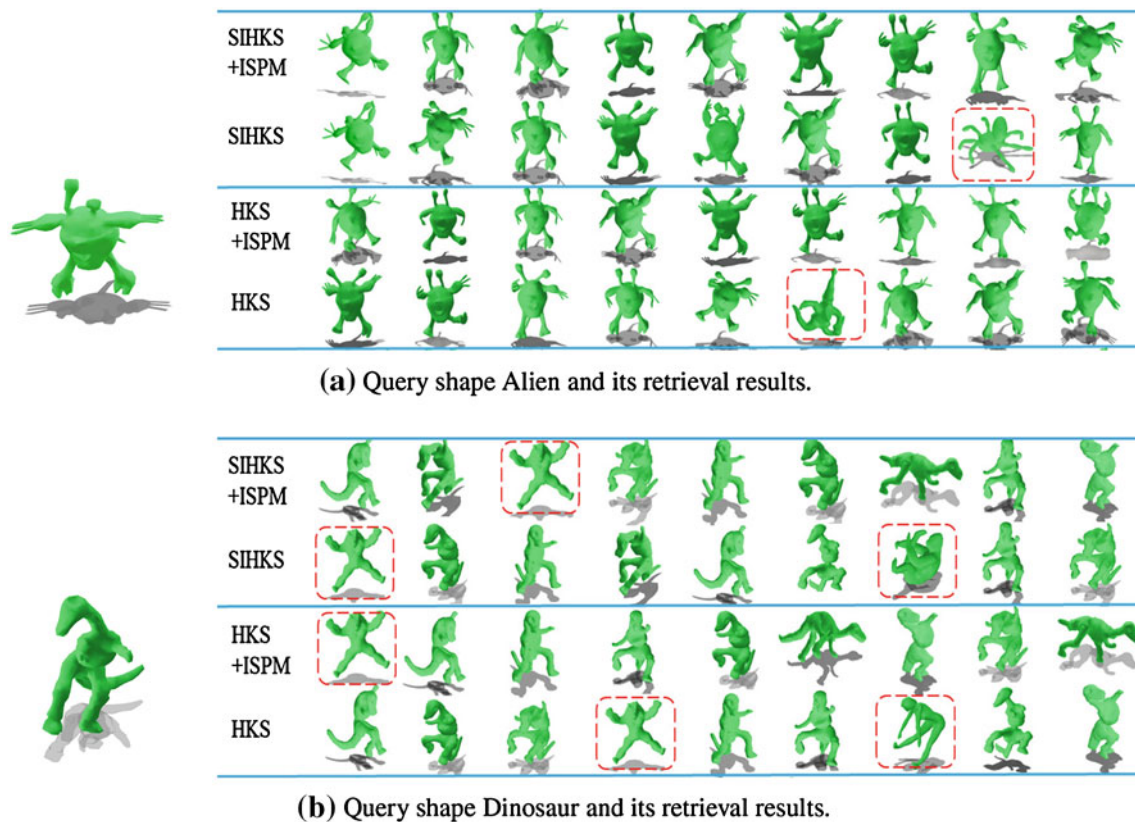
First, we examine the effect of integrating spatial cues on surfaces via the intrinsic partition. Figure 4 shows the performance improvement of the retrieval experiments by matching shapes directly using intrinsic partitions on SHREC 2011 dataset. With the increase of the number of intrinsic partitions, both HKS and SIHKS are improved substantially. The performance of the Shape Google is plotted as points whose partition number is one. Obviously, intrinsic spatial cues on the shape surface proposed in our framework significantly outperform the Shape Google. We only show the result with maximum 19 partitions for visualization purposes. As a matter of fact, the retrieval results will go slightly higher even when a 1,024-partition is adopted in our experiments.

Next, let us examine the behavior of the ISPM. For completeness, Table 1 lists the performance achieved using just the highest level of the pyramid (the "single" columns) as well as the performance of the complete matching scheme using multiple levels (the "pyramid" columns). For both HKS and SIHKS, 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 improvement, using all the levels together helps provide stable results. For HKS with codeword uncertainty, single-level performance actually drops as we go from $L = 7$ to $L = 9$. This means that the highest level of the $L = 7$ 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 $L = 9$ pyramid remains essentially identical to that of the $L = 7$ 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.

In Fig. 5, we show two examples of top nine retrieval results for different methods. There are plenty of examples to demonstrate that our proposed ISPM method improves the Shape Google, we just take two to illustrate the idea. Between each two blue lines in that figure, the upper row is our approach, while the bottom row is Shape Google. For the first query alien, SIHKS confuses it with spider, while HKS confuses it with Santa. 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 separates pipe-like parts and globular parts into different sub-histograms according to the global spatial position, thus resulting in a more descriptive representation. In particular, the Santa and alien models share similar body shape, but the shapes of Santa's hat and the alien's horns are spatially inverted, even though these two parts are similar in terms of the proportion of primitive geometric elements. For the second query dinosaur, ISPM successfully removes the incorrect results of the gorilla and woman models. But the error with armadillo still remains, which turns out the case as ISPM fails. It is understandable since even humans may incorrectly recognize it at first glance. In terms of global shape structures, the armadillo and dinosaur models are almost isometric. So ISPM considers the semantically correspondent parts as good matchings and compares them by their correspondent regions. Because of their similar geometric details of the four legs and the tail, ISPM is still not able to distinguish between these two shapes.

In addition to comparing shapes at their coarsest level (shape-to-shape), ISPM is also able to quantitatively tell the difference at certain detailed levels (patch-to-patch).

(a) Query shape Alien and its retrieval results.

(b) Query shape Dinosaur and its retrieval results.

Fig. 5 Retrieval results using SIHKS and HKS and their ISPM version. Error result is marked in the *red dashed box*. **a** On the *left* is the query shape alien, and the *ten rows on the right* show its top 9 retrieval results. **b** On the *left* is the query shape dinosaur and the *ten rows in the right* show its top 9 retrieval results

4.2 TOSCA database

We also tested our algorithm on the TOSCA database, which has been used to validate the Shape Google [19]. The total positive set size was 531, equally categorized into 13 classes. In each class, the shapes underwent different types of transformations, including null, isometry topology, isometry + topology, triangulation and partiality. We used 456 shapes as negatives. All the shapes were normalized to have approximately the same scale. The retrieval quality was quantitatively measured using the receiver operating characteristic (ROC) curve, which essentially represents a trade-off between the percentage of similar shapes correctly identified as similar (true positive rate or TPR) and the percentage of dissimilar shapes wrongfully identified as similar (false positive rate or FPR). We used the same experimental settings as in [19]. Since the performance of both the traditional codebook and the uncertainty codebook is consistent, we only display the results of the traditional codebook for the sake of convenience. The highest level of ISPM is chosen as $L = 8$, and the dictionary size is set to 48. Figure 6 shows the ROC curves of both BoF and ISPM using HKS and SIHKS for each class of transformations. As can be seen in Fig. 6, the

ISPM method outperforms the BoF of the Shape Google in all cases.

4.3 Strengths and weaknesses of the proposed approach

Similar to BoF, the proposed method performs worse than ShapeDNA on SHREC 2011 benchmark. This is due in part to the fact that ShapeDNA is particularly good at retrieving the near-redundant isometrics. On TOSCA-based database, however, which has different classes of mesh transformations, ISPM performs the best on the grounds that the partitions provided by the second eigenfunction are stable. We may summarize the strengths and weaknesses of our approach as follows:

- **Strengths:** (1) The main advantage of ISPM over BoF is its integration of spatial information in a principled way. (2) ISPM provides a coarse correspondence of shapes.
- **Weaknesses:** (1) A major drawback of ISPM is how to determine an appropriate partition number parameter. Such a limitation also exists in the original paper of Lazebnik's et al. [4] on SPM. Thus, using too many par-
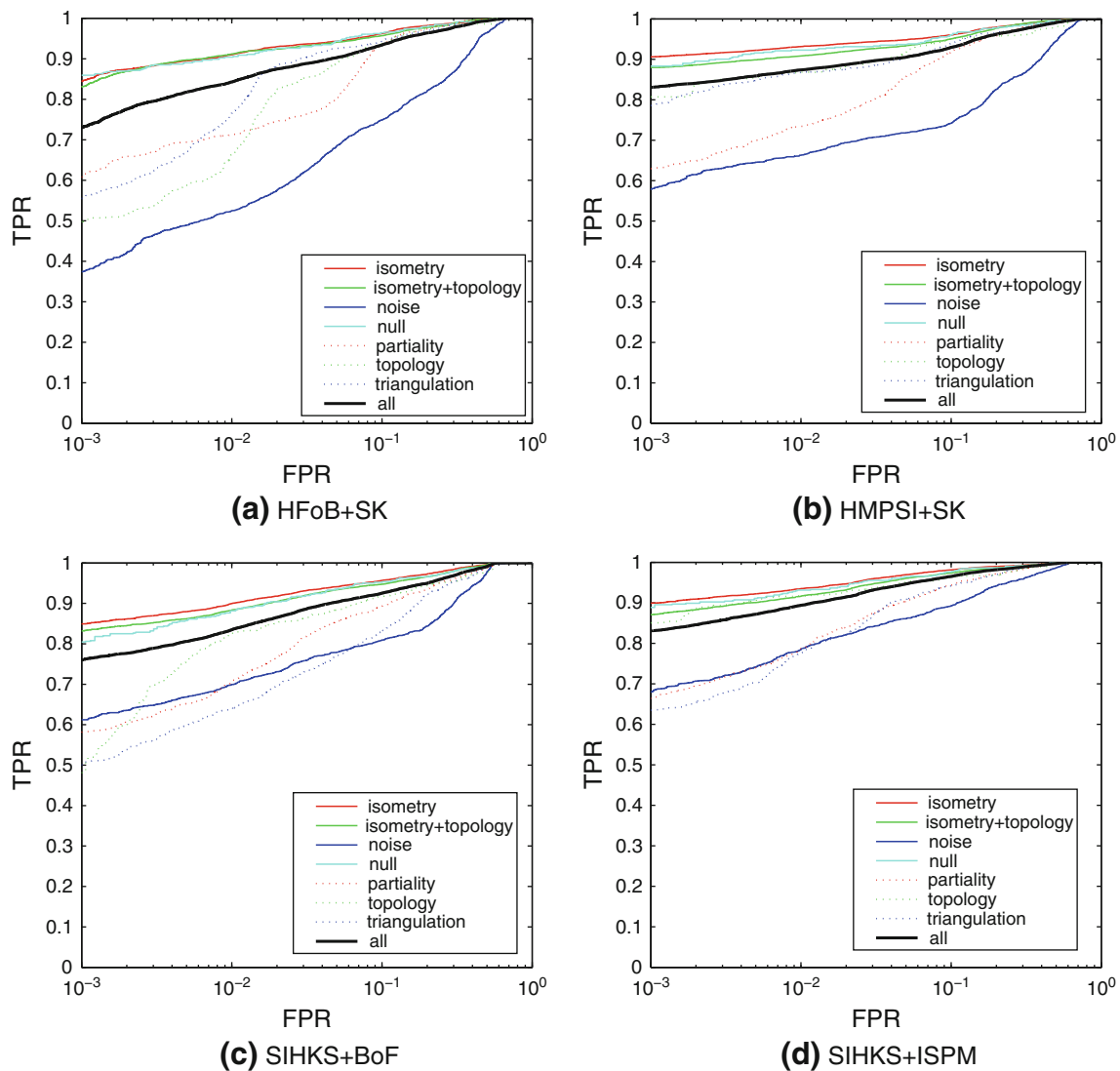
Fig. 6 ROC curves (true positive vs. false positive rate) for different classes of shape transformations using SIHKS and HKS with both BoF and ISPM

titions on the surface tend to degrade the performance of the proposed algorithm, largely because of the mismatching. (2) Unlike the graph-based method, ISPM may still lose the topological information, which is critical in distinguishing between shapes from different classes.

5 Conclusion and future work

We developed an intrinsic version of the SPM, making it suitable for the analysis of deformable 3D shapes. Our construction is based on the isocontours of the second eigenfunction of LB operator on Riemannian manifolds. The proposed partitioning can capture the global shape topological information and provide a deformation invariant representation. Furthermore, the ISPM is able to establish a global

correspondence among shapes. It can be used in combination with any dense shape descriptor, e.g., heat kernel signature or scale-invariant heat kernel signature, and consistently achieves a notable improvement over the BoF model, which only encodes orderless local information.

We plan to extend this work in two directions. First, the spatial pyramid framework offers insights into the success of the different dense shape descriptor in our experiments. Therefore, performing a spatial partition-based investigation on all the recent spectral descriptors, such as wave kernel signature, may provide very practical instructions for further applications. Second, using the proposed global system of coordinates, 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.

References

1. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
2. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR'05), vol 1, pp 886–893
3. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A (2010) The pascal visual object classes (voc) challenge. Int J Comput Vis 88(2):303–338
4. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR'06), vol 2, pp 2169–2178
5. Fehr J, Streicher A, Burkhardt H (2009) A bag of features approach for 3D shape retrieval. In: Proceedings of the international symposium on advances in visual computing (ISVC'09), pp 34–43
6. Lian Z, Godil A, Sun X (2010) Visual similarity based 3D shape retrieval using bag-of-features. In: Proceedings of the IEEE international conference on shape modeling and applications (SMI'10), pp 25–36
7. Toldo R, Castellani U, Fusiello A (2010) The bag of words approach for retrieval and categorization of 3D objects. Vis Comput 26(10):1257–1268
8. Darom T, Keller Y (2012) Scale invariant features for 3-D mesh models. IEEE Trans Image Process 5(21):2758–2769
9. Liu Y, Zha H, Qin H (2006) Shape topics: a compact representation and new algorithms for 3D partial shape retrieval. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR'06), vol 2, pp 2025–2032
10. Tabia H, Colot O, Daoudi M, Vandeborre JP (2011) Non-rigid 3D shape classification using bag-of-feature techniques. In: Proceedings of the IEEE international conference on multimedia and expo (ICME'11), pp 1–6
11. Coifman RR, Lafon S (2006) Diffusion maps. Appl Comput Harmon Anal 21(1):5–30
12. Reuter M, Wolter F, Peinecke N (2006) Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. Comput Aid Des 38(4):342–366
13. Bérard P, Besson G, Gallot S (1994) Embedding Riemannian manifolds by their heat kernel. Geom Funct Anal 4(4):373–398
14. Lévy B (2006) Laplace-Beltrami eigenfunctions: Towards an algorithm that "understands" geometry. In: Proceedings of the IEEE international conference shape modeling and applications, p 13
15. Sun J, Ovsjanikov M, Guibas L (2009) A concise and provably informative multi-scale signature based on heat diffusion. Comput Graph Forum 28(5):1383–1392
16. Aubry M, Schlickewei U, Cremers D (2011) The wave kernel signature: a quantum mechanical approach to shape analysis. In: Proceedings of the computational methods for the innovative design of electrical devices, pp 1626–1633
17. Kokkinos I, Bronstein MM, Yuille A (2012) Dense scale invariant descriptors for images and surfaces. Technical report, INRIA
18. Bronstein AM, Bronstein MM, Guibas LJ, Ovsjanikov M (2011) Shape Google: geometric words and expressions for invariant shape retrieval. ACM Trans Graph 30(1)
19. Ovsjanikov M, Bronstein AM, Bronstein MM, Guibas LJ (2009) Shape Google: a computer vision approach to isometry invariant shape retrieval. In: Proceedings of the international conference on computer vision workshops (ICCVW'09), pp 320–327
20. Kokkinos I, Bronstein MM, Litman R, Bronstein AM (2012) Intrinsic shape context descriptors for deformable shapes. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR'12). IEEE, pp 159–166
21. Sipiran I, Bustos B (2010) A robust 3D interest points detector dased on harris operator. In: Proceedings of the eurographics workshop on 3D object retrieval (3DOR'10), pp 7–14
22. Zaharescu A, Boyer E, Varanasi K, Horaud R (2009) Surface feature detection and description with applications to mesh matching. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR'09). IEEE, pp 373–380
23. Li X, Godil A (2009) Exploring the Bag-of-Words method for 3D shape retrieval. In: Proceedings of the IEEE international conference image processing (ICIP'09), pp 437–440
24. Redondo-Cabrera C, López-Sastre RJ, Acevedo-Rodriguez J, Maldonado-Bascón S (2012) SURFing the point clouds: selective 3D spatial pyramids for category-level object recognition. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR'12), pp 3458–3465
25. Bronstein AM, Bronstein MM, Kimmel R (2008) Numerical geometry of non-rigid shapes. Springer, Berlin
26. Rosenberg S (1997) The Laplacian on a Riemannian manifold. Cambridge University Press, Cambridge
27. Meyer M, Desbrun M, Schröder P, Barr A (2003) Discrete differential-geometry operators for triangulated 2-manifolds. Vis Math III 3(7):35–57
28. Wardetzky M, Mathur S, Kälberer F, Grinspun E (2007) Discrete Laplace operators: no free lunch. In: Proceedings of the eurographics symposium on geometry processing (SGP'07), pp 33–37
29. Rustamov RM (2007) Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In: Proceedings of the eurographics symposium on geometry processing (SGP'07), pp 225–233
30. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2007) Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR'07), pp 1–8
31. Nister D, Stewenius H (2006) Scalable recognition with a vocabulary tree. In: Proceedings of the IEEE computer vision and pattern recognition (CVPR'06), vol 2, pp 2161–2168
32. van Gemert JC, Veenman CJ, Smeulders AWM, Geusebroek JM (2010) Visual word ambiguity. IEEE Trans Pattern Anal Mach Intel 32(7):1271–1283
33. Kac M (1966) Can one hear the shape of a drum? Am Math Mon 73(4):1–23
34. Uhlenbeck K (1976) Generic properties of eigenfunctions. Am J Math 98(4):1059–1078
35. Shi Y, Lai R, Krishna S, Sicotte N, Dinov I, Toga AW (2008) Anisotropic Laplace-Beltrami eigenmaps: Bridging Reeb graphs and skeletons. In: Proceedings of the IEEE computer vision and pattern recognition workshops (CVPRR'08), pp 1–7
36. Lian Z, Godil A, Bustos B, Daoudi M, Hermans J, Kawamura S, Kurita Y, Lavoué G, Van Nguyen H, Ohbuchi R, Ohkita Y, Ohishi Y, Porikli F, Reuter M, Sipiran I, Smeets D, Suetens P, Tabia H, Vandermeulen D (2013) A comparison of methods for non-rigid 3D shape retrieval. Pattern Recogn 46(1):449–461
37. Järvelin K Kekäläinen J (2000) IR evaluation methods for retrieving highly relevant documents. In: Proceedings of the international conference research and development in information retrieval (SIGIR'00), pp 41–48