

# Shift-invariant ring feature for 3D shape

Shuhui Bu · Pengcheng Han · Zhenbao Liu ·  
Ke Li · Junwei Han

Published online: 10 May 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** In this paper, we present a shift-invariant ring feature for 3D shape, which can encode multiple low-level descriptors and provide high-discriminative representation of local region for 3D shape. First, several iso-geodesic rings are created at equal intervals, and then low-level descriptors on the sampling rings are used to represent the property of a feature point. In order to boost the descriptive capability of raw descriptors, we formulate the unsupervised basis learning into an L1-penalized optimization problem, which uses convolution operation to address the rotation ambiguity of descriptors resulting from different starting points in rings. In the following extraction procedure of high-level feature, we use the learned bases to calculate the sparse coefficients by solving the optimization problem. Furthermore, to make the coefficients irrelevant with the sequential order in ring, we use Fourier transform to achieve circular-shift invariant ring feature. Experiments on 3D shape correspondence and retrieval demonstrate the satisfactory performance of the proposed intrinsic feature.

**Keywords** 3D shape descriptor · Sparse coding · Shape correspondence · Shape retrieval

## 1 Introduction

3D models have been extensively applied in the domains of multimedia, graphics, virtual reality, amusement, design, and manufacturing [39] due to the rich information preserving the surface, color, and texture of real objects. A huge number of publicly available models such as Google 3D Warehouse have been quickly spread online. Moreover, with the development of RGB-D devices, e.g., Microsoft Kinect, users can obtain 3D models in a convenient and efficient way, which further leads to the explosion of 3D data. This rapid growth requires effective retrieval and classification techniques [3] for their management and reusing.

There have been many solutions to 3D shape recognition, correspondence, classification, and retrieval in the past decade. These solutions are directly related with shape descriptor, and the comprehensive and excellent reviews can be found in an early work [39] and the latest works [23, 24]. 3D shape descriptors are used to characterize important global or local geometric characteristic, which are distinctively discriminative with other shapes or local regions. Local descriptor captures important geometric changes on local regions of 3D surface, which depicts the different properties of local regions on various shapes. An earlier and representative work is spin images [14]. Recently, Darom and Keller [8] extend the spin images to make them have the capability of scale-invariant and interest point detection. Sipiran et al. [35] adopt 3D Harris detector to find interesting points for 3D shape retrieval, which can be seen as an extension from 2D Harris detector measuring the variation in the gradient of a given function (e.g., the intensity function of a image). 3D SURF descriptor [18, 25] is recently proposed for classifying and retrieving similar shapes.

The above local descriptors depend on extrinsic properties constrained by location and orientation of 3D mesh, or a local

---

S. Bu · P. Han · Z. Liu (✉) · J. Han  
Northwestern Polytechnical University, Xi'an, China  
e-mail: liuzhenbao@nwpu.edu.cn

S. Bu  
e-mail: bushuhui@nwpu.edu.cn

K. Li  
Information Engineering University, Zhengzhou, China

coordinate system defined on a mesh vertex. To overcome the limitations, several intrinsic descriptors have been proposed in recent decade, which do not need to specify the descriptor position relative to an arbitrarily defined coordinate system. Therefore, they achieve much better discriminative capability for 3D shape analysis. Laplace–Beltrami operator, which is a generalization of the Laplacian from flat space to manifold, is appealing for 3D shape retrieval because of sparse, symmetric, and intrinsic properties of its robustness to rigid transformation and deformation. Retrieval methods [9, 12, 20, 41] extract main eigenvalues and eigenvectors of Laplace matrix generated on local regions to match different regions of 3D shapes. Laplace–Beltrami operator also provides an efficient way of computing a conformal map from a manifold mesh to a homeomorphous surface with constant Gaussian curvature, and the histogram of conformal factors [2] serves as a robust pose-invariant signature of 3D shape, which is regarded as an attribute of a graph node to identify segmented parts in bipartite graph matching for 3D shape retrieval [30]. In a recent work [29], 3D shape is also partitioned into several connected iso-surfaces (annuluses) of conformal factors and expressed with a graph where node substitutes each annulus. Heat kernel signature (HKS) [38], a recently proposed local descriptor, attracts much attention from researchers. It provides rich local geometric information which makes the signature invariant to isometric deformation and multi-scale characteristic, thereby achieving better performance in 3D shape retrieval and matching [5, 6, 37]. In order to overcome the influence of diffusion time change under different shape scales [6], Fourier transform is imposed on HKS at each given vertex to obtain time invariants.

Despite the progressive improvement on 3D shape descriptors, the performance of existing 3D shape descriptors is still far from satisfactory. The main issue results from the insufficiency in describing complex 3D shapes just using local statistics, i.e., most local descriptors only catch a piece of geometric characteristics. However, 3D shape is composed of complex topological structure and visibly variational geometry; consequently, only limited information can be extracted. To further improve the performance of 3D shape descriptor, an alternative approach is to learn hidden states from local descriptors and then use them as feature representation. Castellani et al. [7] propose a scheme of local descriptor extraction based on a generative model. Local patches are modeled as a stochastic process through a set of circular geodesic pathways and learned via hidden Markov model. Another work uses intrinsic shape content [19] to characterize the local shape property. In the method, the shape context is processed in an intrinsic local polar coordinate system; therefore, it is intrinsic and invariant to isometric deformation. Furthermore, Fourier transform is applied to the original shape content data to deal with orientation ambiguity.

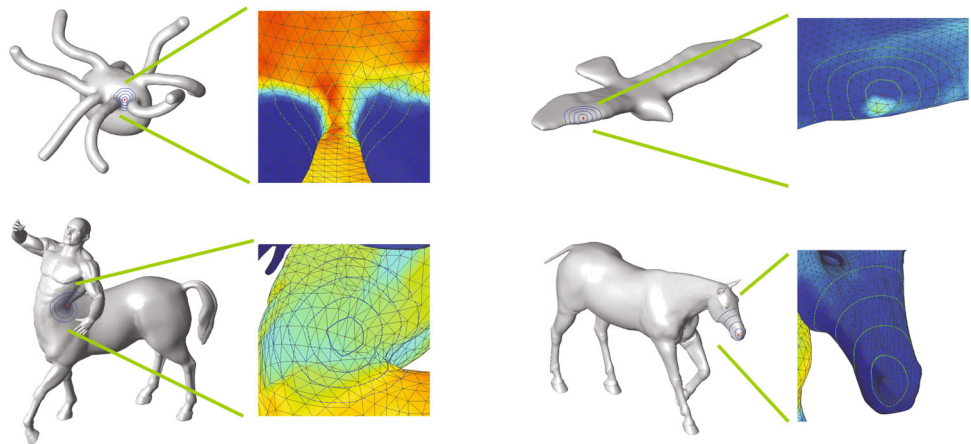
Extracting discriminative information from low-level data is an important work in signal processing, image processing, and other domains, due to that the effective feature will boost the performance of information processing. In this paper, we propose a local ring feature for 3D shape which can encapsulate multi-descriptor into a higher discriminative feature. The idea is to establish iso-geodesic rings of feature point on the mesh, and the iso-geodesic rings are sampled at equal intervals. And then the low-level descriptor values on the sampling rings are used to represent the property of the feature point. However, there exist two problems by applying this feature directly into shape analysis, such as shape correspondence and retrieval. First, this feature is redundant which leads to low discriminative and low efficiency. Second, during the sampling stage, although starting points of the rings for one feature point can be determined and made with same direction, for different feature points the orientations of starting points are difficult to be selected, which consequently makes the comparison or similarity calculation of the feature inefficient and inaccurate.

In signal processing and pattern recognition domains, there has been growing interest in sparse coding [21, 27] with a learned dictionary instead of a predefined one, which is advocated as an effective mathematical description for information processing. Sparse coding is an unsupervised learning algorithm that learns a discriminative high-level description of the unlabeled input data. It represents each input with a sparse linear combination of a set of basis functions and has shown to be useful in many application fields.

Inspired by the success of sparse coding in image and pattern recognition and also to overcome the above mentioned problems, in this work, we introduce sparse coding into our method for extracting high-level feature. The first problem can be easily resolved using sparse coding to learn bases and then extracting sparse coefficients regarded as the high-level feature. In order to settle the second problem, we adopt an improved sparse coding method named shift-invariant sparse coding (SISC) [10, 36], which allows basis to be replicated at each shift position. Therefore, even starting points of rings are different for different feature points, the bases can be learned against the influence. After the bases are obtained, the sparse coefficients can be efficiently got by solving an L1-penalized optimization problem. Finally, to make the coefficients irrelevant with the sequential order of ring, we apply Fourier transformation to the coefficients and then taking the absolute values acting as shift-invariant ring feature (SI-RF).

Several experiments are conducted in both 3D shape correspondence and retrieval tasks. Results and comparisons with state-of-the-art methods indicate that the proposed method has promising performance.

**Fig. 1** Illustration of extracting geodesic rings of four shapes. For each shape, the original shape is drawn in *left*, while extracted rings are plotted in the *right*. The feature points are plotted in *red*, and the *shading color* represents the SDF value



The rest of this paper is organized as follows: Sect. 2 introduces the ring descriptor. Section 3 describes the details of the SI-RF. Experiments of shape correspondence and retrieval on several shape datasets are presented in Sect. 4. We conclude this paper in Sect. 5.

## 2 Ring descriptor for 3D shape

### 2.1 Iso-geodesic rings extraction

For a 3D shape, the description value of the vertex does not provide sufficient discriminative information, especially for the low-level descriptor. Usually, the neighboring vertices and their topological connections provide much more information. Therefore, an effective way to extract highly representative feature of a feature point is to encode the local area's property. The straightforward method to generate local neighborhood region is creating  $n$ -rings of a feature point on the mesh by their topological connections. However, since the edges have variational length, the vertices in the rings have different geodesic distances to the feature point. In order to overcome the shortcoming, we create rings through the geodesic measure.

For a feature point  $v_i$  of the mesh  $M$ , we compute geodesic distances by fast marching method [17] from the feature point  $v_i$  to all other vertices. And then several level sets of the geodesic function are computed using a fixed number of increasing distances  $d_1 < \dots < d_{N_r}$ , where  $N_r$  is the number of rings. Since the number of the sampled points in each ring is different, we use linear interpolation to generate a fixed number of sampling points  $N_s$  in each ring and also make them equally spaced. Points of each ring are collected in a uniform direction (same in clockwise or anti-clockwise) and with the form of  $R^i = [p_1^{d_i}, \dots, p_{N_s}^{d_i}] \in \mathbb{R}^{N_s \times 3}$ . The local iso-geodesic rings are described as  $\mathbf{R} = [R^1; \dots; R^{N_r}] \in \mathbb{R}^{N_s \times N_r \times 3}$ , which are used to represent the local area of the feature point  $v_i$ . The iso-geodesic rings have two advantages:

they are robust to isometric variations and also have the same dimension. Several examples of extracted geodesic rings are shown in Fig. 1. In this study, we set the sampling point of ring  $N_s = 80$  and number of ring  $N_r = 4$ .

### 2.2 Low-level 3D shape descriptors

In order to accurately encode the local property of sampling rings for feature point, we next focus on extracting a set of local shape descriptors. Three representative methods, shape diameter function (SDF) [31], scale-invariant HKS [6], and average geodesic distance (AGD) [11], are adopted as low-level 3D shape descriptors, since these local descriptors are robust against non-rigid and complex shape deformations.

*Shape diameter function* The SDF [31] is a volume-based scalar function measuring the diameters of different parts. The SDF value is computed by sending 30 rays inside a small cone with angle of  $30^\circ$  to intersect with the opposite side of the boundary and averaging these weighted ray lengths. The values remain similar on the neighborhood of the same part and oblivious to articulated deformation.

*Scale-invariant heat kernel signature* HKS [38] is derived from a heat diffusion equation using Laplace–Beltrami operator on surfaces, which has the advantages of providing rich local geometric information, invariant to isometric deformation, and multi-scale characteristic. However, a limitation of the HKS is that it is sensitive to the scale of shape, for example, when the shape becomes large the region described by the HKS becomes small in the same time range. To cope with the problem, Bronstein and Kokkinos [6] proposed a scale-invariant heat kernel signature (SI-HKS) by Fourier transform of the difference of the HKS. The detail of this method can be found in [6].

*Average geodesic distance* The AGD is introduced by Hilaga et al. [11] for the purpose of shape matching. However, the original AGD is not robust when using extremum value as a normalization factor, e.g. the use of the intra-class geometric variations make the local descriptor change eas-

ily. It is, therefore, difficult to be applied to generate robust descriptor from a set of models. We modify the normalization factor to the mean of geodesic distances between all pairs of vertices to cope with the above limitation. For any model, the modified AGD descriptor has a fixed mean value 1.

### 2.3 Ring descriptor for 3D shape

Finally, we concatenate SDF, first six frequency components of SI-HKS, and AGD descriptors to form a low-level shape descriptor as

$$f(x_i) = (\text{SDF}(x_i), \text{SI-HKS}(x_i)[\omega_1, \dots, \omega_6], \text{AGD}(x_i)), \quad (1)$$

where the dimension of the feature is  $f_m = 8$ . For the SI-HKS, the time-scale is set to be  $[1, 20]$  with an interval of 0.2, the number of eigenfunction is set to 100, and the log time base  $\alpha = 2$ .

In the next step, the low-level descriptors are interpolated at vertices to form descriptors on the iso-geodesic rings  $\mathbf{F}(v_i) \in \mathbb{R}^{N_s \times N_r \times f_m}$  with the following form:

$$\mathbf{F}(v_i) = \begin{bmatrix} f(p_1^1) & f(p_2^1) & \dots & f(p_{N_s}^1) \\ f(p_1^2) & f(p_2^2) & \dots & f(p_{N_s}^2) \\ \dots & \dots & \dots & \dots \\ f(p_1^{N_r}) & f(p_2^{N_r}) & \dots & f(p_{N_s}^{N_r}) \end{bmatrix}, \quad (2)$$

where  $v_i$  is a feature point on the mesh,  $N_s$  is the point number in the ring, and  $N_r$  is the ring number. Because mesh has adequate density of elements, the descriptor values in one facet on the mesh change slowly. According to this fact, linear interpolation is adopted in this study to interpolate from descriptor values at around vertices to iso-geodesic rings. In the following processing, we use  $\mathbf{F}^l \in \mathbb{R}^{N_s \times N_r}$  to denote one dimension of the low-level descriptors, where  $l \in \{1, \dots, f_m\}$  is the descriptor index.

### 3 Shift-invariant ring feature for 3D shape

In order to generate intrinsic feature for 3D shape, the proposed method is carried out as the following steps:

1. Down-sampling: the 3D mesh often has thousands of vertices and directly using all of the vertices to learn the dictionary is time-consuming; therefore, in the first step 3D mesh is down-sampled to hundreds of points for accelerating the dictionary learning.
2. Basis learning: we use a large set of extracted low-level descriptors on the iso-geodesic rings to learn a set of bases via sparse coding.

3. Feature extracting: for a given point in the 3D mesh, we extract the high-level features of the point through solving an L1-penalized optimization problem.

The overall flowchart of the proposed method is depicted in Fig. 2.

#### 3.1 Feature point selection

Usually, 3D shapes need more than thousands of vertices to represent them accurately; however, the feature of a given vertex is similar to its neighbor vertices. In addition, using the full set of vertices is computationally intractable for dense meshes. Therefore, in the first stage of the method, a few points on the mesh are selected as feature points. We adopt the uniform sampling that is the farthest point sampling (FPS) strategy [28], which is adopted to compute subset feature points  $\mathbf{V} = \{v_i \in \mathbf{M}, i = 1, \dots, n_s\}$  on the mesh  $M$ , where  $n_s$  is the desired sampling point number. The initial point  $v_1 \in M$  is sampled at random.

#### 3.2 Dictionary learning

Sparse coding [21,27] has attracted many researchers from the domain of image and vision to complete tasks of image analysis, e.g., image retrieval, classification, recognition, and segmentation. The advantages of sparse coding are twofold. The first one is that it can capture higher-level features via learning basis functions from unlabeled data, and these features contain more semantic information and are adaptable to complex recognition tasks. Another one is that sparse coding can learn over-complete basis sets, which adequately represent objects more than limited orthogonal basis, and thus they can capture a large number of patterns from the input data.

Through the ring descriptor processing, the originally low-level descriptors around a point are encapsulated into a fixed dimension array, as a consequence, the dictionary can be efficiently learned by sparse coding method. However, for different feature points on mesh, it is difficult to determine starting points in the iso-geodesic rings, which may cause low representative of the basis functions. For the sake of addressing the issue, in this study we introduce SISC [10,36] which allows each basis function to be replicated at each position in the iso-geodesic rings, and consequently the learned bases are insensitive to the circular shifts of the input data. The SISC is an extension of tradition sparse coding [21,27], in which the multiplication operation is replaced by convolution operation as shown in Eq. (3). One dimension of the feature (2) can be represented as

$$\mathbf{F}^l = \sum_{j=1}^{N_b} \mathbf{s}_j * \mathbf{a}_j, \quad (3)$$

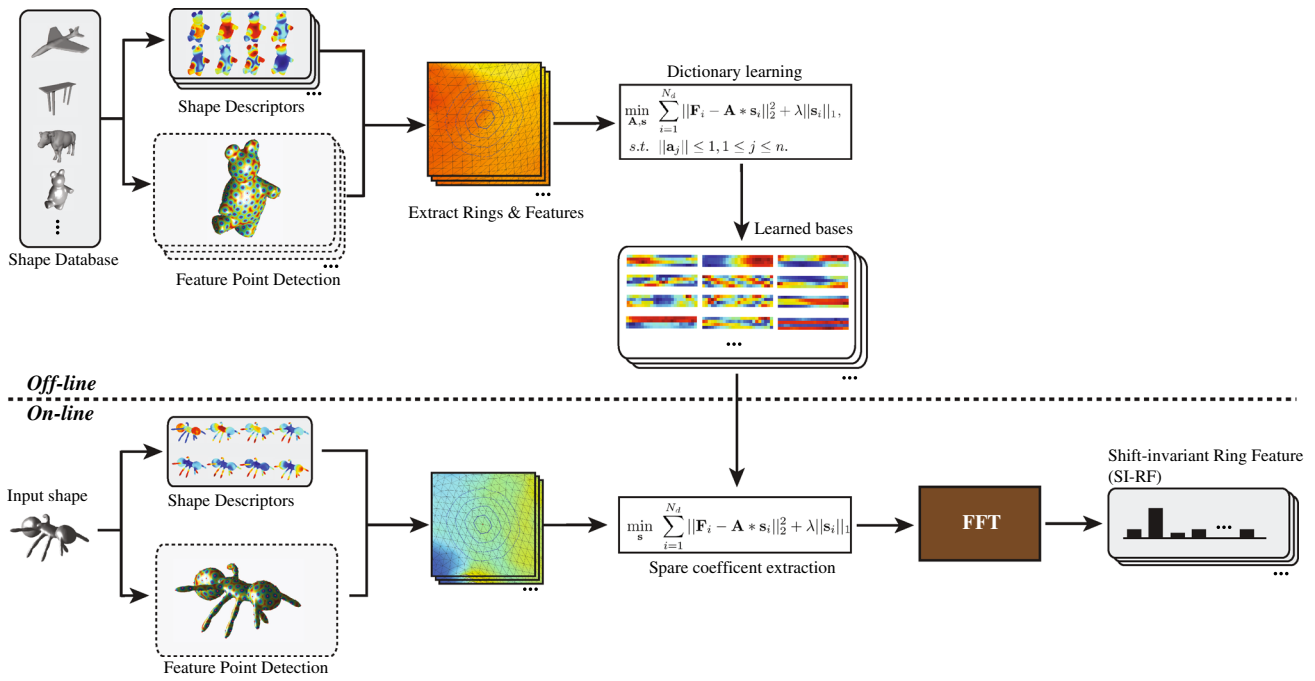


Fig. 2 The flowchart of the proposed method

where  $\mathbf{a}_j \in \mathbb{R}^{q \times N_r}$  is basis which has a lower dimension than the input data ( $q \leq N_s$ ) because the ring number is smaller than that of sampling points of a ring. Based on the input  $\mathbf{F}^l$  and basis  $\mathbf{a}_j$ , the corresponding vector  $\mathbf{s}_j \in \mathbb{R}^{N_s - q + 1}$  represents the coefficient for each possible circular offset of the basis  $\mathbf{a}_j$ . The corresponding optimization problem for learning bases  $\mathbf{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_{N_b}\}$  and coefficients  $\mathbf{s}$  is then described as

$$\min_{\mathbf{A}, \mathbf{s}} \sum_{i=1}^{N_d} \|\mathbf{F}_i - \mathbf{A} * \mathbf{s}_i\|_2^2 + \lambda \|\mathbf{s}_i\|_1, \tag{4}$$

$$\text{s.t. } \|\mathbf{a}_j\| \leq 1, 1 \leq j \leq N_b. \tag{5}$$

Objective function (4–5) is a joint optimization problem, which is commonly non-convex. Nevertheless, it can be seen as a convex problem with respect to one variable while keeping the other one fixed. Specifically,  $\mathbf{s}$  can be solved in the convex optimization problem with  $\mathbf{A}$  being fixed, and vice versa. These sub-problems are solved alternatively and iteratively to obtain the optimal  $\mathbf{A}$ . After the convergence of the objective function, the sparsest coefficients can be computed with the learned basis functions. We adopt a fast computation approach described in [10] to solve the optimization problem for each dimension of the low-level descriptor. After all types of descriptors have been computed, a set of bases

$$\{\mathbf{A}^1, \dots, \mathbf{A}^{f_m}\}$$

are used to compute sparse coefficients treated as high-level feature for feature point.

### 3.3 Shift-invariant feature extraction

Since bases are learned from a set of the feature points on different shapes, for one point in a mesh the high-level feature can be represented using the learned bases from solving the optimization problem of (5) with a learned  $\mathbf{A}$ . For one dimension of the low-level descriptor, the sparse coefficient is represented by  $\mathbf{s}^l$ . During the generation step of iso-geodesic rings, starting points of rings are selected with same orientation. But for different feature points, they are not coherent, which consequently results in a rotation ambiguity. So as to provide an *intrinsic* feature for 3D shape, we introduce Fourier transform to achieve rotation invariance. By applying Fourier transform to the sparse coefficients, we can get

$$\mathcal{F}\{s_k\}(\omega) = \sum_k s_k \exp(-ik\omega), \tag{6}$$

$$\mathcal{F}\{s_{k+c}\}(\omega) = \mathcal{F}\{s_k\}(\omega) \exp(-ic\omega), \tag{7}$$

where  $k$  represents the index of one basis’s corresponding sparse coefficient, and  $c$  is the circular shift of the sparse coefficient. Taking the absolute value we have  $|\mathcal{F}\{s_k\}(\omega)| = |\mathcal{F}\{s_{k+c}\}(\omega)|$ , which can eliminate the influence from the incoherence sequential orders of iso-geodesic rings.

Finally, one dimension of low-level descriptor corresponding to intrinsic feature  $\theta^l$  is calculated as

$$\theta^l = \left| \mathcal{F}\{s^l\} \right|. \quad (8)$$

All the dimensions of intrinsic feature for low-level descriptors are concatenated to form the SI-RF  $\theta = [\theta^1, \dots, \theta^{f_m}]$ .

## 4 Experiments

For evaluating the proposed SI-RF, experiments on shape correspondence and retrieval are performed. In the shape correspondence experiments we adopt the surface correspondence benchmark [16] including three data sets that have a variety of objects with ground-truth correspondences. In the shape retrieval experiment, we use the McGill shape benchmark [34] which contains 457 models including shapes with articulating parts and without articulation. The set of articulated shapes consists of 255 models in 10 categories, and there are 20–30 models per category.

### 4.1 Experiments on correspondence

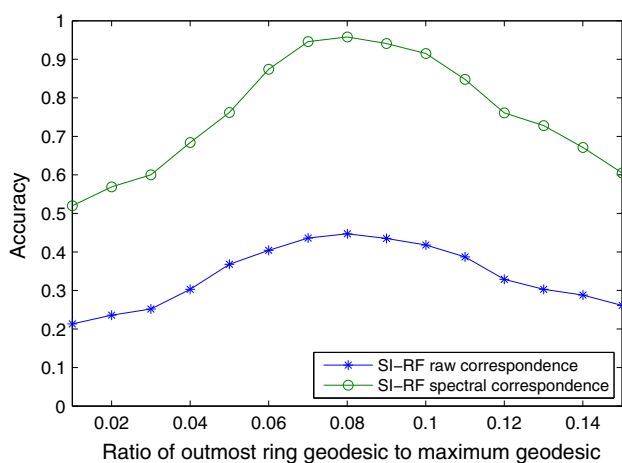
The feature discriminative capability of the proposed method is first evaluated via shape correspondence experiments, while the recent works on this filed can be found in [16, 26, 40]. In the first two experiments, we use the watertight models in surface correspondence benchmark [16] as the ground-truth data. There are 20 shape categories; among them we just use 11 classes shapes because they have well-defined correspondences as follows: human, glasses, air-plane, ant, teddy, hand, plier, fish, bird, armadillo, and four-legged animal.

First, we investigate how the radii of iso-geodesic rings will affect the correspondence performance. Here we set out-

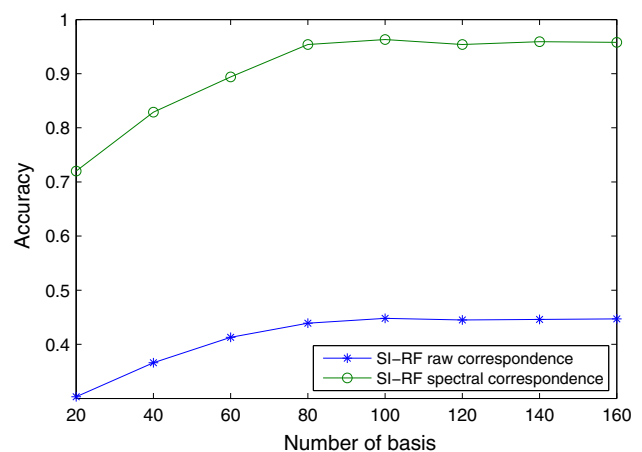
most ring's geodesic  $d_{N_r}$  to 1 ~ 15 % of shape's maximum geodesic among any pair of vertices. Two correspondence estimation methods are adopted here: the first one is raw correspondence which selects pair with minimal feature distance as correspondence, and the other one is spectral correspondence [22]. The averaged correspondence accuracy is used as the evaluation measure, where it is calculated by averaging the percentage of correct correspondence for all pairs of shapes. The comparison results are shown in Fig. 3, where the horizontal axis represents geodesic ratio, and the vertical axis represents the averaged correspondence accuracy. From the figure, we can get that when the sampling range is small the performance is not good. Below 0.08 of the maximum geodesic, the performance is increasing; however, after that geodesic range the performance decreases. According to the results, we select 0.08 of the maximum geodesic as the sampling range for the following experiments.

In the second experiment, the performances of correspondence with different number of bases  $N_b$  are evaluated. We set the basis number to 20, 40, 60, 80, 100, 120, 140, and 160, respectively. The averaged correspondence accuracies are calculated through above mentioned means, and the results are shown in Fig. 4. From the figure, we can conclude that the accuracy increases fast until the number of basis approaches 80, while the change becomes slow after the number of basis exceeds 100. Although the larger number of bases results in better performance, the computation time of dictionary learning increases rapidly, which causes low computation efficiency. Therefore, a trade-off basis size 80 is determined for the following experiments. Several correspondence results are plotted in Fig. 5; above mentioned means are used to compute correspondence for comparing our feature with the recent SI-HKS which is highly discriminative.

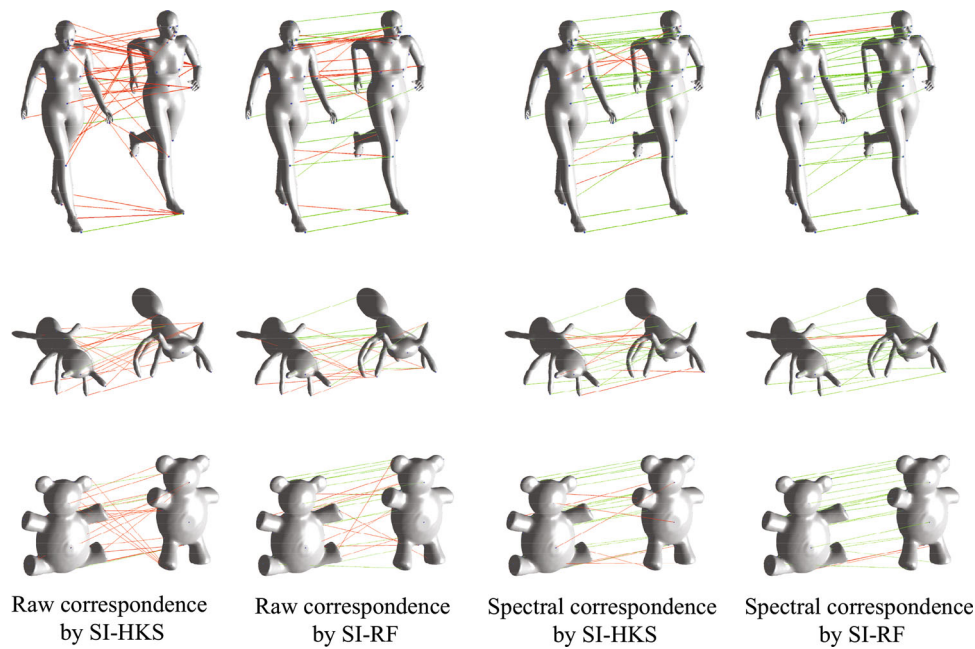
Since the optimal parameters have been obtained, we adopt them in the third experiment which uses TOSCA



**Fig. 3** Averaged correspondence accuracy with different geodesic range  $d_{N_r}$



**Fig. 4** Averaged correspondence accuracy with different number of basis  $N_b$



**Fig. 5** Illustration of correspondences of three shapes using SI-HKS and proposed SI-RF features. Two correspondence methods are used: raw correspondence which selects minimal feature distance between source and target shapes, and spectral correspondence method. The *first column* shows the correspondence results using raw correspondence

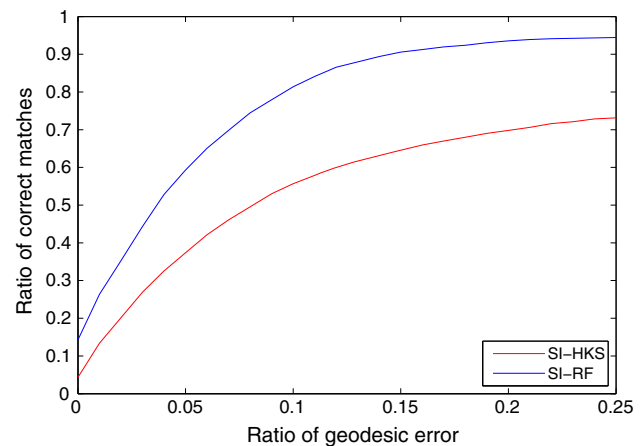
with SI-HKS, and the *second column* shows the results of raw correspondence with proposed SI-RF. The *third column* shows the results of spectral correspondence with SI-HKS, and the *last column* shows the results using SI-RF. In these figures, the *red line* indicates wrong correspondence, while the *green line* indicates correct correspondence

shapes [4] and ground-truth maps from [16]. For each shape, 1,000 points are sampled via FPS [28]. Points from the source shape are matched to the target shape, and here we define the geodesic distance between ground-truth corresponding vertex and matched vertex on the target shape as geodesic error. The analysis results are plotted in Fig. 6, where the horizontal axis denotes the geodesic error, and the vertical axis represents the ratio of correct correspondence. Note in the figure, at the geodesic error 0.15 (ratio of geodesic error to shape's max geodesic distance) the accuracy can reach 90 % with the proposed method, while only 64.5 % via using SI-HKS.

#### 4.2 Experiments on retrieval

Besides the correspondence experiments, shape retrieval experiment is also tested for evaluating whether the feature is suitable for similar shape matching.

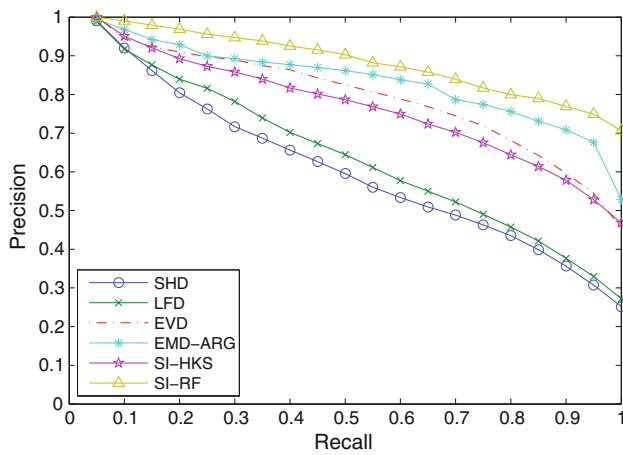
In this experiment, we implement a retrieval method based on the idea of Shape Google [5], and McGill shape benchmark [34] is used as retrieval data. The first step is to sample 1,500 points on the mesh using the FPS [28], and then a dictionary is learned through low-level descriptors on iso-geodesic rings of the sampling points. The shape's global feature is calculated using the spatially sensitive bag-of-features (SS-BoF). The process of extracting sparse coefficients has the similar function as bag of features; hence we directly



**Fig. 6** Ratio of correct matches vs. geodesic error using SI-HKS and proposed SI-RF

use them to generate SS-BoF. The similarity of two shapes is measured by the L2 distance. For the SI-HKS, the original SS-BoF calculation method is used to extract features of shapes.

We use six standard evaluation metrics to assess the performance of the proposed method. They are precision-recall curve, nearest neighbor (NN), first tier (FT), second tier (ST), E-measure (E), and discounted cumulative gain (DCG), where the detailed definitions can be found in [33].



**Fig. 7** Recall-precision curve of some previous methods and the proposed method on McGill dataset

**Table 1** Retrieval performance of the proposed methods using five standard measures on the McGill shape benchmark

Methods	NN	FT	ST	$E$	DCG
EMD-ARG [1]	93.3	69.2	88.9	–	90.8
SI-HKS	89.1	57.4	71.2	50.7	85.1
SI-RF	<b>93.5</b>	<b>72.7</b>	<b>90.3</b>	<b>64.3</b>	<b>92.3</b>

Bold indicates better values than the other methods

The recall-precision curves of the proposed method and state-of-the-art methods are shown in Fig. 7, which includes shape harmonic descriptor [15], light-field descriptor [32], eigenvalue descriptor of affinity matrix [13], earth movers distance and attributed relation graph (EMD-ARG) [1]. The figure shows that the retrieval performance of the proposed high-level feature is superior to other methods. The numerical results are listed in Table 1. From the table, we can draw the conclusion that all of the measures are improved from SI-HKS to proposed high-level features. The gain of DCG is 8.5 %, which demonstrates that the proposed method has the capability to improve retrieval performance using learned bases.

## 5 Conclusion

Extracting high-level feature of 3D shape is still a challenging topic up to date, owing to the complex structure compared with image data. In this paper, we present a novel high-level 3D shape feature extraction framework for correspondence and retrieval of 3D shapes. The unsupervised dictionary learning is adopted to learn high representative bases for generating local feature of shapes. In the dictionary learning, for boosting the computation speed, some uniformly sampled points on the mesh are used as feature points. And then, iso-geodesic rings and corresponding low-level descriptors

on them are collected for representing the local region of the feature point. SISC is used to learn bases for feature extraction due to that it can avoid the orientation ambiguity. The sparse coefficients which are more discriminative are generated through L1-penalized optimization, and then Fourier transformation is applied to the sparse coefficients so as to settle the shift-sensitive problem. Finally, the spectrum of sparse coefficients act as SI-RF. The experiment results demonstrate that the learned high-level features have better performance both on correspondence and retrieval tasks.

Although the proposed method achieves better performance, several improvements should be made in the future. First, we use iso-geodesic rings to represent local region of feature point; however, this method has low efficiency of computation and possible singularities on them which may degrade performances. In the following research, better local region extracting methods will be investigated. In addition, currently only one-layer sparse coding is adopted to extract feature, and we consider that convolutional neural network which has better performance in many domains will be studied.

**Acknowledgments** This work was supported partly by grants from National Natural Science Foundation of China (61202185, 61003137, 41201390), Northwestern Polytechnical University Basic Research Fund (JC201202, JC201220), the Fundamental Research Funds for the Central Universities, Shaanxi Natural Science Fund (2012JQ8037), and Open Project Program of the State Key Lab of CAD&CG (A1306), Zhejiang University.

## References

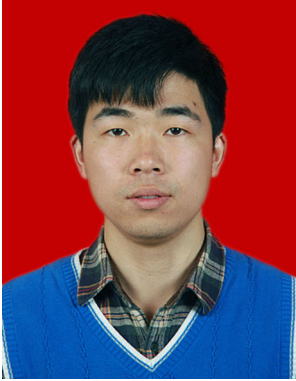
- Agathos, A., Pratikakis, I., Papadakis, P., Perantonis, S., Azariadis, P., Sapidis, N.S.: 3d articulated object retrieval using a graph-based representation. *Vis. Comput.* **26**(10), 1301–1319 (2010)
- Ben-Chen, M., Gotsman, C.: Characterizing shape using conformal factors. In: *Eurographics Workshop on 3D Object Retrieval*, pp. 1–8 (2008)
- Bimbo, A.D., Pala, P.: Content-based retrieval of 3d models. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMCCAP)* **2**(1), 20–43 (2006)
- Bronstein, A.M., Bronstein, M., Bronstein, M.M., Kimmel, R.: *Numerical Geometry of Non-Rigid Shapes*. Springer, New York (2008)
- Bronstein, A.M., Bronstein, M.M., Guibas, L.J., Ovsjanikov, M.: Shape google: geometric words and expressions for invariant shape retrieval. *ACM Trans. Graph. (TOG)* **30**(1), 1 (2011)
- Bronstein, M.M., Kokkinos, I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1704–1711. IEEE (2010)
- Castellani, U., Cristani, M., Fantoni, S., Murino, V.: Sparse points matching by combining 3d mesh saliency with statistical descriptors. In: *Computer Graphics Forum*, vol. 27, pp. 643–652. Wiley, New York (2008)
- Darom, T., Keller, Y.: Scale-invariant features for 3-d mesh models. *IEEE Trans. Image Process.* **21**(5), 2758–2769 (2012)



9. Dubrovina, A., Kimmel, R.: Matching shapes by eigendecomposition of the Laplace–Beltrami operator. In: Proceedings of Symposium on 3D Data Processing, vol. 2 (2010)
10. Grosse, R., Raina, R., Kwong, H., Ng, A.Y.: Shift-invariance sparse coding for audio classification. [arXiv:1206.5241](https://arxiv.org/abs/1206.5241) (2012)
11. Hilaga, M., Shinagawa, Y., Kohmura, T., Kunii, T.L.: Topology matching for fully automatic similarity estimation of 3d shapes. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 203–212. ACM (2001)
12. Hu, J., Hua, J.: Salient spectral geometric features for shape matching and retrieval. *Vis. comput.* **25**(5–7), 667–675 (2009)
13. Jain, V., Zhang, H.: A spectral approach to shape-based retrieval of articulated 3d models. *Comput.-Aided Des.* **39**(5), 398–407 (2007)
14. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(5), 433–449 (1999)
15. Kazhdan, M., Funkhouser, T., Rusinkiewicz, S.: Rotation invariant spherical harmonic representation of 3d shape descriptors. In: Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, pp. 156–164. Eurographics Association (2003)
16. Kim, V.G., Lipman, Y., Funkhouser, T.: Blended intrinsic maps. In: ACM Transactions on Graphics (TOG), vol. 30, p. 79. ACM (2011)
17. Kimmel, R., Sethian, J.A.: Computing geodesic paths on manifolds. *Proc. Natl. Acad. Sci.* **95**(15), 8431–8435 (1998)
18. Knopp, J., Prasad, M., Willems, G., Timofte, R., Van Gool, L.: Hough transform and 3d SURF for robust three dimensional classification. In: Computer Vision-ECCV 2010, pp. 589–602. Springer, New York (2010)
19. Kokkinos, I., Bronstein, M.M., Litman, R., Bronstein, A.M.: Intrinsic shape context descriptors for deformable shapes. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 159–166. IEEE (2012)
20. Laga, H., Schreck, T., Ferreira, A., Godil, A., Pratikakis, I.: Bag of words and local spectral descriptor for 3D partial shape retrieval. In: Eurographics Workshop on 3D Object Retrieval (2011)
21. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. *Adv. Neural Inf. Process. Syst.* **19**, 801 (2007)
22. Léordeanu, M., Hebert, M.: A spectral technique for correspondence problems using pairwise constraints. In: Tenth IEEE International Conference on Computer Vision, vol. 2, pp. 1482–1489. IEEE (2005)
23. Lian, Z., Godil, A., Bustos, B., Daoudi, M., Hermans, J., Kawamura, S., Kurita, Y., Lavoué, G., Van Nguyen, H., Ohbuchi, R., et al.: A comparison of methods for non-rigid 3d shape retrieval. *Pattern Recognit.* **46**(1), 449–461 (2012)
24. Liu, Z., Bu, S., Zhou, K., Gao, S., Han, J., Wu, J.: A survey on partial retrieval of 3d shapes. *J. Comput. Sci. Technol.* **28**(5), 836–851 (2013)
25. López-Sastre, R.J., García-Fuertes, A., Redondo-Cabrera, C., Acevedo-Rodríguez, F., Maldonado-Bascón, S.: Evaluating 3d spatial pyramids for classifying 3d shapes. *Comput. Graph.* **37**(5), 473–483 (2013)
26. Mykhalchuk, V., Cordier, F., Seo, H.: Landmark transfer with minimal graph. *Comput. Graph.* **37**(5), 539–552 (2013)
27. Olshausen, B.A., Field, D.J.: Sparse coding with an overcomplete basis set: a strategy employed by v1? *Vis. Res.* **37**(23), 3311–3325 (1997)
28. Peyré, G., Cohen, L.D.: Geodesic remeshing using front propagation. *Int. J. Comput. Vis.* **69**(1), 145–156 (2006)
29. Sfikas, K., Theoharis, T., Pratikakis, I.: Non-rigid 3d object retrieval using topological information guided by conformal factors. *Vis. Comput.* **28**(9), 943–955 (2012)
30. Shapira, L., Shalom, S., Shamir, A., Cohen-Or, D., Zhang, H.: Contextual part analogies in 3d objects. *Int. J. Comput. Vis.* **89**(2–3), 309–326 (2010)
31. Shapira, L., Shamir, A., Cohen-Or, D.: Consistent mesh partitioning and skeletonisation using the shape diameter function. *Vis. Comput.* **24**(4), 249–259 (2008)
32. Shen, Y.T., Chen, D.Y., Tian, X.P., Ouhyoung, M.: 3D model search engine based on lightfield descriptors. *EUROGRAPHICS Interactive Demos, Granada*, pp. 1–6 (2003)
33. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The princeton shape benchmark. In: Proceedings of the Shape Modeling Applications, pp. 167–178. IEEE (2004)
34. Siddiqi, K., Zhang, J., Macrini, D., Shokoufandeh, A., Bouix, S., Dickinson, S.: Retrieving articulated 3-d models using medial surfaces. *Mach. Vis. Appl.* **19**(4), 261–275 (2008)
35. Sipiran, I., Bustos, B., Schreck, T.: Data-aware 3d partitioning for generic shape retrieval. *Comput. Graph.* **37**(5), 460–472 (2013)
36. Smith, E., Lewicki, M.S.: Efficient coding of time-relative structure using spikes. *Neural Comput.* **17**(1), 19–45 (2005)
37. Kovnatsky, A., Bronstein, M., Bronstein, A., Raviv, D., Kimmel, R.: Affine-invariant photometric heat kernel signatures. In: Eurographics Workshop on 3D Object Retrieval, pp. 1–8 (2012)
38. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. In: Computer Graphics Forum, vol. 28, pp. 1383–1392. Wiley, New York (2009)
39. Tangelder, J.W., Velkamp, R.C.: A survey of content based 3d shape retrieval methods. *Multimed. Tools Appl.* **39**(3), 441–471 (2008)
40. Tevs, A., Berner, A., Wand, M., Ihrke, I., Seidel, H.P.: Intrinsic shape matching by planned landmark sampling. In: Computer Graphics Forum, vol. 30, pp. 543–552. Wiley, New York (2011)
41. Wu, H.Y., Zha, H., Luo, T., Wang, X.L., Ma, S.: Global and local isometry-invariant descriptor for 3d shape comparison and partial matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 438–445. IEEE (2010)



**Shuhui Bu** received the Ph.D. degree in College of Systems and Information Engineering from University of Tsukuba, Japan in 2009. Currently, he is an associate professor at Northwestern Polytechnical University, China. Prior to joining Northwestern Polytechnical University, he was an assistant professor at Kyoto University, Japan. His research includes 3D shape analysis, image processing, and computer vision.



**Pengcheng Han** received the BS degree from Northwestern Polytechnical University of China in 2013. Currently, he is working toward the MS degree at Northwestern Polytechnical University, China. His research interests include artificial intelligence, 3D shape analysis, image processing, and computer vision.



**Ke Li** received the Ph.D. degree in Cartography and Geographical Information Engineering from Information Engineering University, China in 2008. He is an associate professor at Information Engineering University, China. His research includes GIS, image processing, and computer vision. He has published over 15 journal and conference papers in the related areas.



**Zhenbao Liu** received the Ph.D. degree in computer science from the College of Systems and Information Engineering, University of Tsukuba, Japan, in 2009. He is an associate professor at Northwestern Polytechnical University, Xi'an. He was a visiting scholar in the GrUVi Lab of Simon Fraser University in 2012. His research interests include 3D shape analysis, matching, retrieval and segmentation.



**Junwei Han** is a currently a professor with Northwestern Polytechnical University, Xi'an. He received his Ph.D. degree in pattern recognition and intelligent systems from the School of Automation, Northwestern Polytechnical University in 2003. His research interests include computer vision and multi-media processing.