Computers & Graphics 46 (2015) 117–129

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

SMI 2014

Local deep feature learning framework for 3D shape

Shuhui Bu [a], Pengcheng Han [a], Zhenbao Liu [a,*], Junwei Han [a], Hongwei Lin [b]

[a] Northwestern Polytechnical University, China
[b] Zhejiang University, China

ARTICLE INFO

Article history:
Received 8 July 2014
Received in revised form
30 August 2014
Accepted 17 September 2014
Available online 2 October 2014

Keywords:
Bag-of-words
Deep learning
Shape correspondence
Shape symmetry detection
Shape recognition
Local deep feature

ABSTRACT

For 3D shape analysis, an effective and efficient feature is the key to popularize its applications in 3D domain. In this paper, we present a novel framework to learn and extract local deep feature (LDF), which encodes multiple low-level descriptors and provides high-discriminative representation of local region on 3D shape. The framework consists of four main steps. First, several basic descriptors are calculated and encapsulated to generate geometric bag-of-words in order to make full use of the various basic descriptors' properties. Then 3D mesh is down-sampled to hundreds of feature points for accelerating the model learning. Next, in order to preserve the local geometric information and establish the relationships among points in a local area, the geometric bag-of-words are encoded into local geodesic-aware bag-of-features (LGA-BoF). However, the resulting feature is redundant, which leads to low discriminative and efficiency. Therefore, in the final step, we use deep belief networks (DBNs) to learn a model, and use it to generate the LDF, which is high-discriminative and effective for 3D shape applications. 3D shape correspondence and symmetry detection experiments compared with related feature descriptors are carried out on several datasets and shape recognition is also conducted, validating the proposed local deep feature learning framework.

1. Introduction

The advancement of modeling, digitizing, and visualizing techniques for 3D models has led to an increasing amount of 3D models in the fields of multimedia, graphics, virtual reality, amusement, design, and manufacturing [1]. Nowadays, a large number of publicly available models such as Google 3D Warehouse have been quickly spread online. In addition, 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 causes high demand of 3D models techniques including shape retrieval, recognition, classification, and correspondence [2].

Reviewing the implementations of these techniques, we can find that feature-based methods play an important role and some details could be found in an early work [1] and the latest works [3,4]. 3D shape descriptors are used to characterize important global or local geometric characteristics, which are distinctively discriminative with other shapes or local regions. Some descriptors, such as area and volume, shape distributions [5], ratios derived from the object's convex-hull, or the Light Field Descriptor [6] have been proposed and achieved great performance in the task of matching, retrieval, and some other applications [7]. Although these techniques make some breakthrough, there still remain many hard problems badly in need of being solved. For example, the shape descriptors mentioned above are global and just use a single vector to represent an object. However, 3D models have rich information including surface, color, and texture, while a single vector cannot represent an object effectively. In addition, global descriptors are often not invariant to scaling, rotation, or translation, hence they have limited capability to discriminate shape variance. To cope with above-mentioned problems, local descriptors [4,7] which use a single vector to describe the local surface region around a number of sample points on an object, have been studied in recent decades. They have the merits of capturing important geometric changes on local regions of 3D surface, being invariant to scaling, rotation, and isometric transformation.

A good local descriptor is the one that is invariant to “unimportant” geometric changes, especially rotation, translation, scaling, or bending (such as changing the pose of an articulated character) [8]. Because only considering the feature of point itself, the descriptor can be influenced easily by geometric changes. When extracting the local feature of a 3D shape, we should take into account the neighbor area surrounding the feature point. In the last decade, some local descriptors, which will be introduced in the following section in detail, have been proposed and successfully used in many tasks. However, the performance of

[*] Corresponding author.
E-mail address: liuzhenbao@nwpu.edu.cn (Z. Liu).

http://dx.doi.org/10.1016/j.cag.2014.09.007
0097-8493/© 2014 Elsevier Ltd. All rights reserved.

some 3D shape local descriptors is still far from satisfactory. The main issue results from three aspects: First, some local descriptors are insufficient to describe complex 3D shape, i.e., only catching a piece of geometric characteristics. Second, 3D shape is composed of complex topological structure and visibly variational geometry, consequently for one type of descriptor only limited information can be extracted. Third, although some descriptors can collect enough information about one local region of 3D shape, they are redundant, which leads to the inefficient usage. Thus, in order to make the extracted feature boost the performance of shape analysis, it is vital to design an effective and efficient local descriptor which can provide discriminative information from raw data.

In this paper, we propose a novel framework to learn and extract local descriptor for 3D shape. The fundamental of the framework is to extract an intermediate representation preserving its surrounding information from low-level 3D descriptors. However, this intermediate representation is redundant which results in the low efficiency. Recently, the deep learning [9–11] has been applied successfully in speech recognition, image processing, and so on. First of all, it can provide a powerful solution to get the high-level feature that is discriminative and robust. So it is critical for pattern recognition. In addition, this method can achieve better generalization because of that the high-level feature is learned from the low-level features. Thus, we adopt deep belief networks (DBNs) to extract compact feature from the intermediate representation. Through the unsupervised learning, model parameters of DBNs are optimized, and the output of the DBNs for newly input data is regarded as the high-level feature which is called as local deep feature (LDF).

The advantages of this framework are as follows:

1. The framework is not only limited to SI-HKS or AGD, other local descriptors are also supported.
2. Multiple features can be fused to provide abundant description.
3. The learning procedure is fully unsupervised.
4. Unlike other machine learning methods which need to tune parameter manually for obtaining the best performance, there are no parameters to be tuned in the learning procedure. Some other parameters, which are used for generating intermediate representation, have little influence on the performance and it is easy to select proper parameters.

Several experiments are conducted in 3D shape correspondence, symmetry detection, and shape recognition tasks. Results and comparisons with related descriptors indicate that the proposed framework reaches promising performance.

2. Related work

Extrinsic descriptors: Some local descriptors are extracted based on location and orientation of 3D mesh, or a local coordinate system defined on a vertex. An earlier and representative work is spin images [12]. Recently, Darom et al. [13] extend the spin images to possess the capability of scale-invariant and interest point detection. Sipiran et al. [14] adopt 3D Harris detector to locate 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 [15,16] is recently proposed for classifying and retrieving similar shapes. Although these features have been applied in many 3D shape processing applications, they belong to the extrinsic descriptors and usually cannot preserve the rich information on local region of 3D shape.

Intrinsic descriptors. To overcome the above limitations, several intrinsic descriptors have been proposed in recent decades, 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 [17–20] 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. The histogram of conformal factors [21] 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 [22]. In a recent work [23], 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 [24], a recently proposed local descriptor, absorbs researchers' much attention. It provides rich local geometric information which makes the signature invariant to isometric deformation and has multi-scale characteristics, thereby achieving better performance in 3D shape retrieval and matching [25–28]. In order to overcome the influence of diffusion time change under different shape scales [25], Fourier transform is imposed on heat kernel signature at each given vertex to obtain scale invariant. Another work uses intrinsic shape context (ISC) [29] 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.

Learning features. Feature learning based methods attract attention of many researchers in the last decade because of their capability of improving discriminability of low-level feature.

In the research of Shape Google [27,28], despite the introduction of spatial-sensitive bag-of-features (SS-BoF), the authors also present a similarity-sensitive hashing method to achieve the best discriminability and compact representation. A middle-level feature extraction scheme through learning hidden states from local basic descriptors is proposed by Castellani et al. [30,31]. In the method, local patches are modeled as a stochastic process through a set of circular geodesic pathways and learned via hidden Markov model. Bu et al. [32] propose shift-invariant ring feature (SI-RF) based on iso-geodesic rings and shift-invariant sparse coding for 3D shape analysis. It represents the local region of a feature point efficiently and has great performance on correspondence and retrieval tasks.

The Laplacian-based descriptors achieve state-of-the-art performance, however, they usually focus on different properties of shape and are suitable for specified task. In order to provide a generic feature descriptor for 3D shape, Litman et al. [33] propose a learning scheme for the construction of optimized spectral descriptors. In order to collect rich information from the raw data and select the most significant feature, Barra et al. [34] propose a method utilizing multiple kernel learning to find optimal linear combination of kernels in classification and retrieval.

Above-mentioned methods focus on feature itself, but ignore the structure consistency. Structural learning, which can produce high-level semantic labels from low-level features through a global optimization, has been successfully applied to segmentation or labeling. Kalogerakis et al. [35] introduce a data-driven approach to simultaneous segmentation and labeling of parts in 3D meshes. They adopt conditional random field model with defined terms assessing the consistency of faces with labels and terms between labels of neighboring faces. To realize the automatic recognition of functional parts of man-made 3D shapes, Laga et al. [36] use graph to represent 3D shape, and then model the context of a shape part as walks in the graph. In the method, the similarity computation can be efficiently performed with graph kernels.

Best view selection is an important procedure in view-based shape retrieval, in order to achieve better performance, Laga [37] proposes a framework to automatically select the best views of 3D models by learning sets of 2D views that not only maximize the similarity between shapes of the same class, but also make the views discriminate shapes in different classes. To deal with the problem of low compactness and discrimination power of view-based descriptors, Tabia et al. [38] adopt vectors of locally aggregated tensors to generate descriptor, and then use principal component analysis to reduce the dimension of the descriptor. Secord et al. [39] propose a perceptual model for best view selection, in which the goodness measure relies on weights determined via a large user study. Gao et al. [40] propose a 3D object retrieval method with Hausdorff distance learning. In their method, relevance feedback information is employed to select positive and negative view pairs with a probabilistic strategy and a view-level Mahalanobis distance metric is learned to estimate the Hausdorff distances between objects.

Interactive feature learning is another important learning method because it usually provides semantic information according to the interaction with human. Moreover, it has the merit of robustness, avoiding to generate abnormal results. Leng et al. [41] present an interactive learning mechanism, which creates a mapping from feature points in low-level feature space to point in a high-level semantic space. The mechanism receives long-term relevance feedback from users via recorded retrieval history, which is adopted to capture users' semantic information to refine retrieval results.

3. Framework of local deep feature

The proposed novel feature learning framework for 3D shape is carried out in the following four stages, while the flowchart is depicted in Fig. 1.

3.1. Basic 3D shape descriptors

In this research, we adopt scale-invariant heat kernel signature and improved average geodesic distance as the low-level 3D shape descriptors which are used for generating intermediate representation.

Scale-invariant heat kernel signature: HKS [24] 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. To cope with the problem, Bronstein and Kokkinos [25] proposed a scale-invariant heat kernel signature (SI-HKS) by Fourier transform of the difference of the HKS.

Average geodesic distance: The average geodesic distance (AGD) [42] is initially introduced for the purpose of shape matching. However, the AGD is not robust when using extremum as a normalization factor, e.g. the use of the intra-class geometric variations make the local descriptor change easily. It is therefore difficult to be applied to generate bag-of-words 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.

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

$$F(x_i) = (SIHKS(x_i)[\omega_1, \dots, \omega_6], AGD(x_i)), \tag{1}$$

where the dimension of the feature is $M=7$. 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 $\alpha=2$. Due to each dimension has varying value ranges and scales, they are linearly normalized to $[-1, 1]$ according to each dimension's maximum and minimum values. Feature weighting is described in Section 3.3.

3.2. 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 neighbors. In addition, using the full set of vertices is computationally intractable for dense meshes. Therefore, in this work, a few points on the mesh are selected as feature points for training the model. We adopt farthest point sampling (FPS) strategy [43] as the uniform sampling, which is to compute subset feature points $V = \{v_i \in X, i = 1, \dots, N_s\}$ on the mesh X, where N_s is the desired sampling point number. The initial point $v_1 \in X$ is sampled at random.

3.3. Local information encoding

For a 3D shape, the description value of the vertex does not provide sufficient discriminative information especially for the low-level descriptor. Usually, the neighbor vertices and their topological connections provide much more information. Therefore, an effective way to extract high-representative feature for a feature point is to encode the local area's property. Nevertheless, it is difficult to collect local information because of the variational length of edges and the complex structure of mesh. To overcome these challenges, we propose a method to encode the local area's property into local geodesic-aware bag-of-features regarded as intermediate representation, which expresses the occurrence probability of geometric words and extracts the rich information on 3D shape efficiently. Also, this way makes the intermediate representation have the same dimension.

In order to generate geometric vocabulary, the most widely used un-supervised method such as k-means is frequently used. However, each dimension has varying contribution for discrimination. As a consequence, the traditional k-means lacks of capability to automatically distribute the weights for each dimension of feature. In this work, we adopt an enhanced k-means method which uses Minkowski metric and automatic feature weighting [44] to generate geometric words more precisely.

After the geometric words $\mathcal{C} = \{c_1, c_2, \dots, c_K\}$ of size K are obtained, the next step is to quantize the low-level descriptor space in order to obtain a compact representation. For each point $x \in X$ with the descriptor $F(x)$, we define the feature distribution $\theta(x) = (\theta_1(x), \dots, \theta_K(x))^T$, a $K \times 1$ vector whose elements are

$$\theta_i(x) = N_1(x) \exp\left(-\frac{\|F(x) - c_i\|_2^2}{k_{BoF}\ \sigma_{min}^2} \right), \tag{2}$$

the constant $N_1(x)$ is selected with the constraint $\|\theta(x)\|_1 = 1$. The above equation is a "soft" version of vector quantization, not only the nearest word is selected, but also some similar words also have feature value. $\theta_i(x)$ can be interpreted as the probability of the point x to be associated with the geometric word c_i. The benefit of soft quantization is that it can generate more representative probability feature values. In order to control the range of similar words selection, two parameters are used in this study, σ_{min}, the

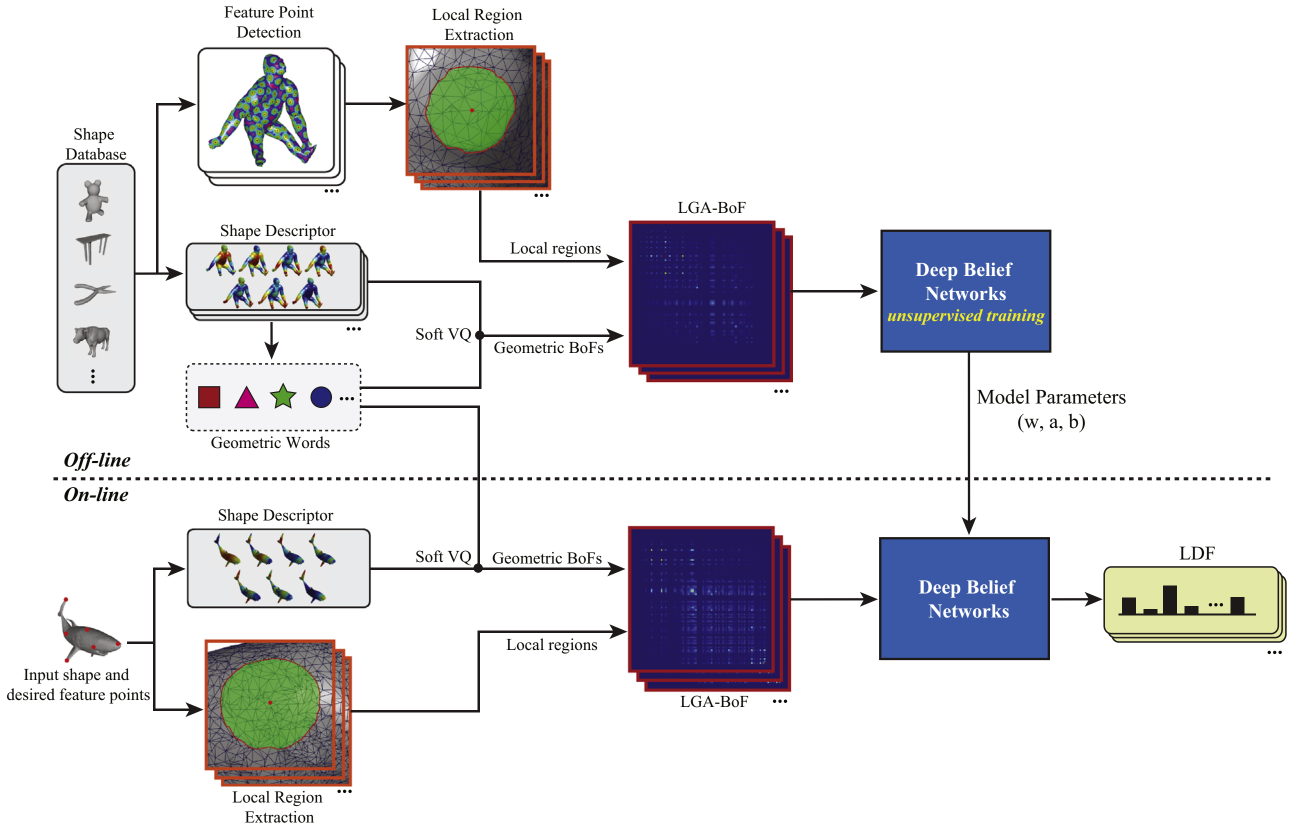

Fig. 1. The flowchart of the proposed method.

minimum distance between any two geometric words, and k_{BoF}, a parameter controlling the decay coefficient for soft quantization.

The disadvantage of bag-of-features is the fact that they consider only the distribution of the words and lose the relations between them. In case of shapes, the phenomenon may be more pronounced due to the poorer of shape features, and consequently shapes tend to have many similar geometric words. In order to overcome this problem, text search engines commonly use vocabularies consisting not only of single words, but also of combination of words or expressions. The analogical expression in shapes would be sets of spatially related geometric words. Being different from previous works [27,28], we use geodesic to measure relationship between geometric words which avoids the possible influence from time scale and shape size under the condition of using heat kernel. Therefore, we define the local geodesic-aware bag-of-features (LGA-BoF):

$$V(x) = N_2(x) \sum_{x_i \in \odot_x} \sum_{x_j \in \odot_x} \theta(x_i)\theta(x_j)^T \cdot \exp\left(-k_{gd}\frac{g(x_i,x_j)}{\sigma_{gd}}\right), \tag{3}$$

where σ_{gd} is the maximal geodesic distance of any vertices in the mesh, k_{gd} is distance decay rate which will be discussed in the experiment section, and $N_2(x)$ is a normalization factor which makes features have a fixed maximum value of 1. The resulting representation V is a $K \times K$ matrix, representing the frequency of appearance of nearby geometric words of vertices i and j. $\odot_x$ means a local region of vertex x, we establish the local area through the geodesic measure and the region size is determined with a geodesic threshold d_l that is an important parameter discussed in experiment section. This expression not only provides a position-independent representation of a feature point but also expresses the relationship between the vertices in the local region.

Some examples of local regions are plotted in Fig. 2. Under the most common conditions, the local region is a circle area which just has a boundary. While in some special cases, the boundary will degrade to several isolated lines. Therefore, traditional methods [29,30,32], which use rings to represent the local region, might be failed under this condition. It is worthwhile to note that the proposed method merely uses descriptors on the vertices in the local region, which avoids above-mentioned trouble and leads to a robustness description.

In this procedure, the intermediate representation of the local region can be got. However, if the number of the words is very big, the extracted feature will be redundant and inefficient. In this study we use the deep learning method [9–11] to improve the performance of the feature.

3.4. Learning local deep feature

Recently, deep learning [9–11] based feature learning has become a promising research topic, which can extract structural information from low-level features. Because it does not require high-level structure to be constructed by human, and the high-level features are extracted in an un-supervised manner. As a consequence, it has been successfully applied to image retrieval, image segmentation, image recognition, speech recognition and so on, and it is found to achieve highly competitive performance. However, due to the intrinsic difference between structural of 3D mesh data (graph data) and image and speech data (constant and simple structure relationship), it is difficult to be applied to 3D shape recognition and retrieval directly. In this work, in order to conquer this limitation, we adopt intermediate representation as the input of the deep learning, and then use the learned feature in 3D shape correspondence and symmetry detection.

Recent works on deep belief networks (DBNs) [10,11] have shown that it is feasible to learn multiple layers of non-linear features that are useful for object classification without requiring

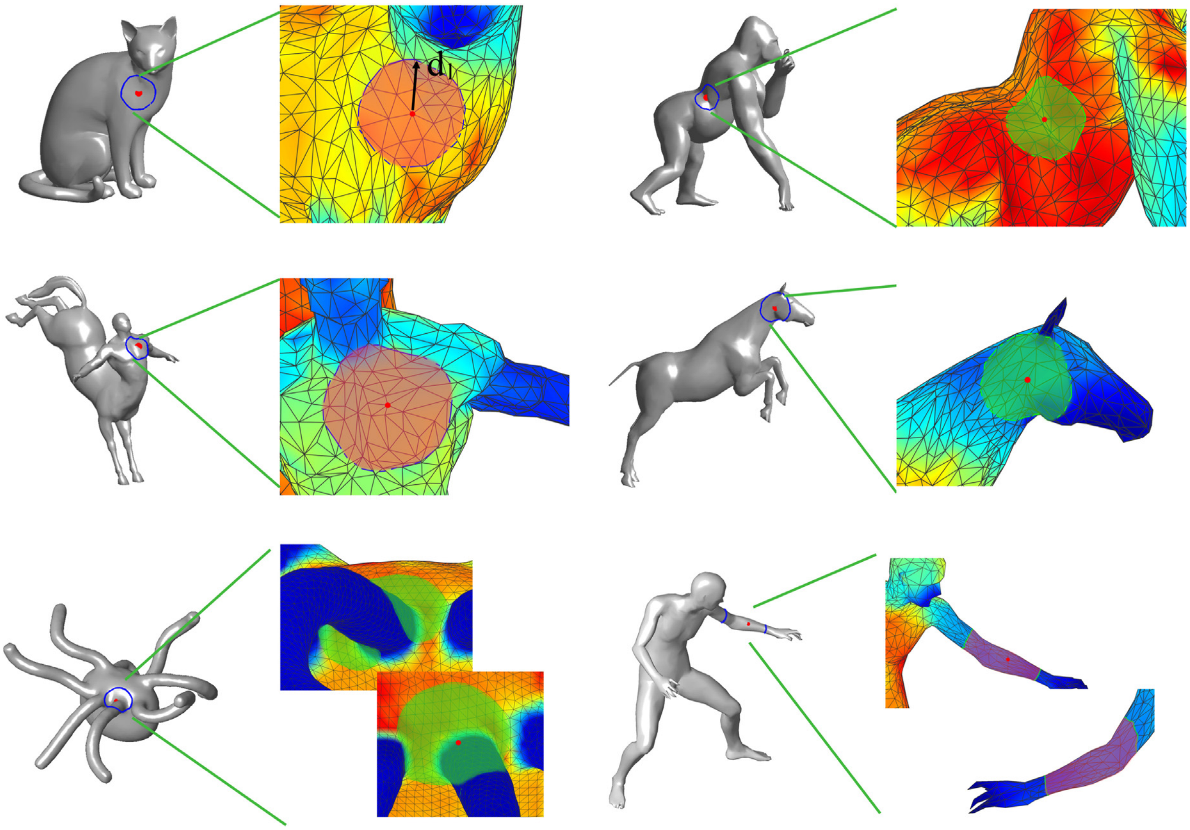

Fig. 2. Illustration of extracting local regions of six shapes. For each shape, the original shape is drawn in left, while extracted local region is plotted in the right. The feature points are plotted in red, d_l is geodesic distance threshold determining the region size. The last row shows some complex regions which can also be described by the proposed method. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

labeled data. The features are trained layer by layer in a restricted Boltzmann machine (RBM) [45–47] by means of contrastive divergence (CD) [47]. The feature activations learned by one layer RBM become the input data for training the next layer RBM. After training, the optimized parameters which are good for modeling the statistical structure in a set of unlabeled data, and the last layer's output is a kind of highly representative feature which encodes the input data.

Table 1
The statistic timing for each step of the proposed LDF framework. The time unit is second.

Procedure	Sample number	All timing	Averaged timing
Geodesic	400 shapes	1049.2	2.623
SI-HKS	400 shapes	471.6	1.179
AGD	400 shapes	2.3	0.005
FPS	400 shapes (400 sample points)	1.4	0.003
k-means	0.5 M	247.9	–
LGA-BoF	80,000	179.0	0.002
DBN training	80,000	327.7	–
DBN testing	80,000	3.8	4.75e–5
Training overall	200 shapes	1516.2	–
Testing	200 shapes	945.0	4.7

3.4.1. *Restricted Boltzmann machines*

In order to make the paper more self-contained, we succinctly discuss the concept of restricted Boltzmann machines. The RBM is a two layer, bipartite, undirected graphical model with a set of binary hidden unit $\mathbf{h}$, a set of (binary or real-valued) visible units $\mathbf{v}$, and symmetric connections between these two layers represented by a weighted matrix W. The joint distribution $p(\mathbf{v},\mathbf{h};\theta)$ over the visible units $\mathbf{v}$ and hidden units $\mathbf{h}$, given the model parameters $\theta=\{\mathbf{w},\mathbf{a},\mathbf{b}\}$, is defined in terms of an energy function $E(\mathbf{v},\mathbf{h};\theta)$ of

$$p(\mathbf{v},\mathbf{h};\theta)=\frac{\exp(-E(\mathbf{v},\mathbf{h};\theta))}{Z}, \tag{4}$$

where $Z=\sum_v\sum_h \exp(-E(\mathbf{v},\mathbf{h};\theta))$ is a normalization factor or partition function and the marginal probability that the model assigns to a visible vector $\mathbf{v}$ is

$$p(\mathbf{v};\theta)=\frac{\sum_h \exp(-E(\mathbf{v},\mathbf{h};\theta))}{Z}. \tag{5}$$

For a Bernoulli (visible)–Bernoulli (hidden) RBM, the energy is

$$E(\mathbf{v},\mathbf{h};\theta)=-\sum_{i=1}^{V}\sum_{j=1}^{H} w_{ij}v_ih_j-\sum_{i=1}^{V} b_iv_i-\sum_{j=1}^{H} a_jh_j, \tag{6}$$

where w_{ij} represents the symmetric interaction between visible unit v_i and hidden unit h_j, b_i and a_j the biases, and V and H are the numbers of visible and hidden units. The conditional probabilities can be efficiently calculated as

$$p(h_j=1|\mathbf{v};\theta)=\sigma\left(\sum_{i=1}^{V} w_{ij}v_i+a_j\right), \tag{7}$$

$$p(v_i=1|\mathbf{h};\theta)=\sigma\left(\sum_{j=1}^{H} w_{ij}h_j+b_i\right). \tag{8}$$

where $\sigma(x)=1/(1+\exp(-x))$ is a sigmoid activation function.

In principal, the RBM parameters can be optimized by performing stochastic gradient ascent on the log-likelihood of training data. Unfortunately, computing the extract gradient of the log-likelihood is intractable. Instead, the CD approximation [47] is typically used, which has been shown to work well in practice.

3.4.2. Deep belief networks

Stacking a number of the RBMs and learning layer by layer from bottom to top gives rise to a DBN. It has been shown that the layer-by-layer greedy learning strategy [10] is effective, and the greedy procedure achieves approximate maximum likelihood learning. In our work, the bottom layer RBM is trained with the input data of LGA-BoFs, and the activation probabilities of hidden units are treated as the input data for training the upper-layer RBM, and so on. We use the un-labeled 3D shape data to train the DBN layer-by-layer. After obtain the optimal parameters $\theta = \{\mathbf{w}, \mathbf{a}, \mathbf{b}\}$, the inputted LGA-BoFs are processed layer-by-layer with Eq. (7) till the final layer. And the last layer's output is used as local deep feature (LDF).

4. Experiments

For evaluating the proposed framework and the novel local deep feature, experiments on both shape correspondence and symmetry detection are performed. In the experiments we adopt the surface correspondence benchmark [48] including Watertight dataset [49], TOSCA dataset [50], and SCAPE dataset [51] that have

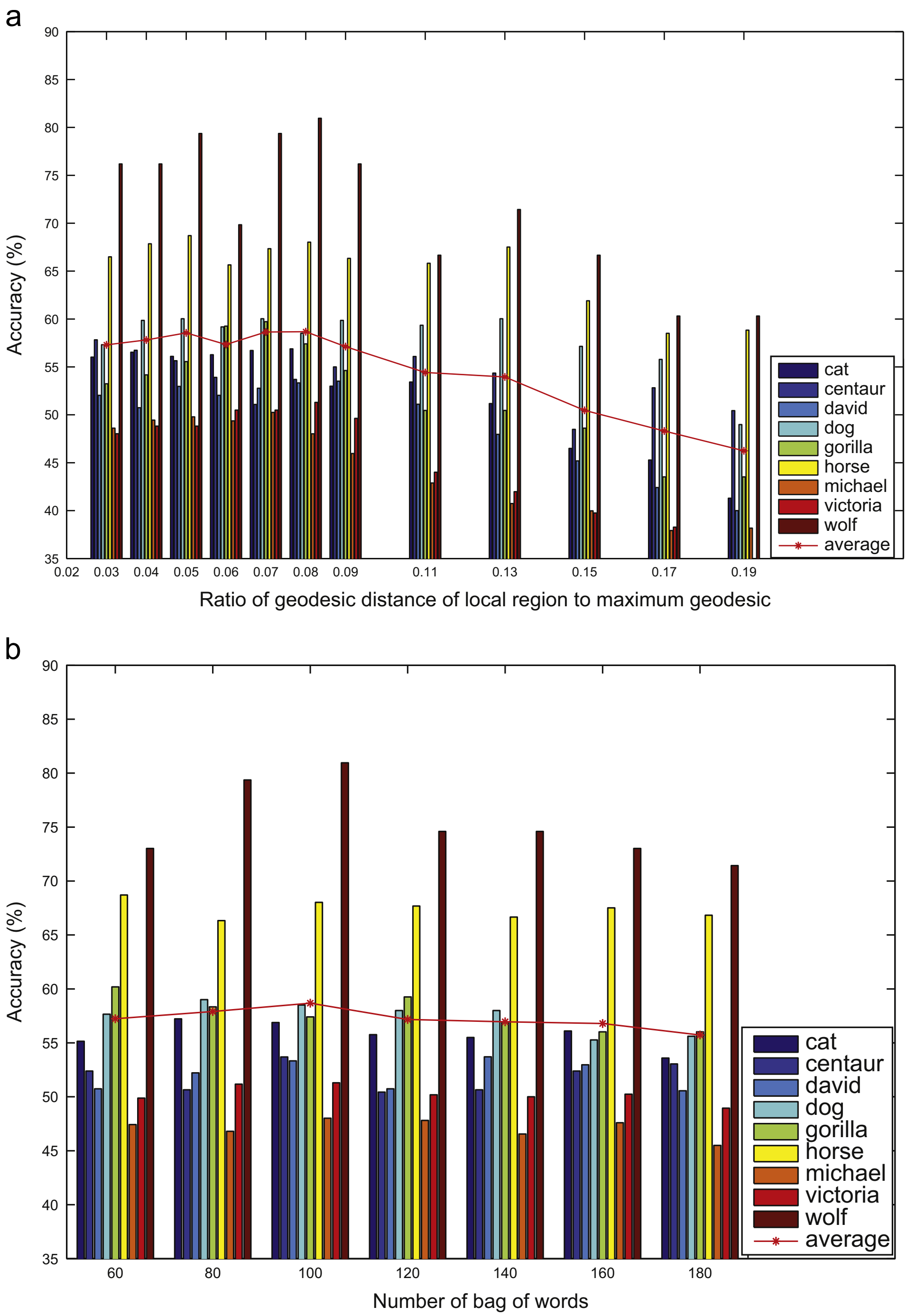

Fig. 3. Setting parameters experiment on TOSCA. (a) Averaged correspondence accuracy with different geodesic range d_l, (b) averaged correspondence accuracy with different bag-of-words number BoW_n and (c) averaged correspondence accuracy with different k_{GD}.

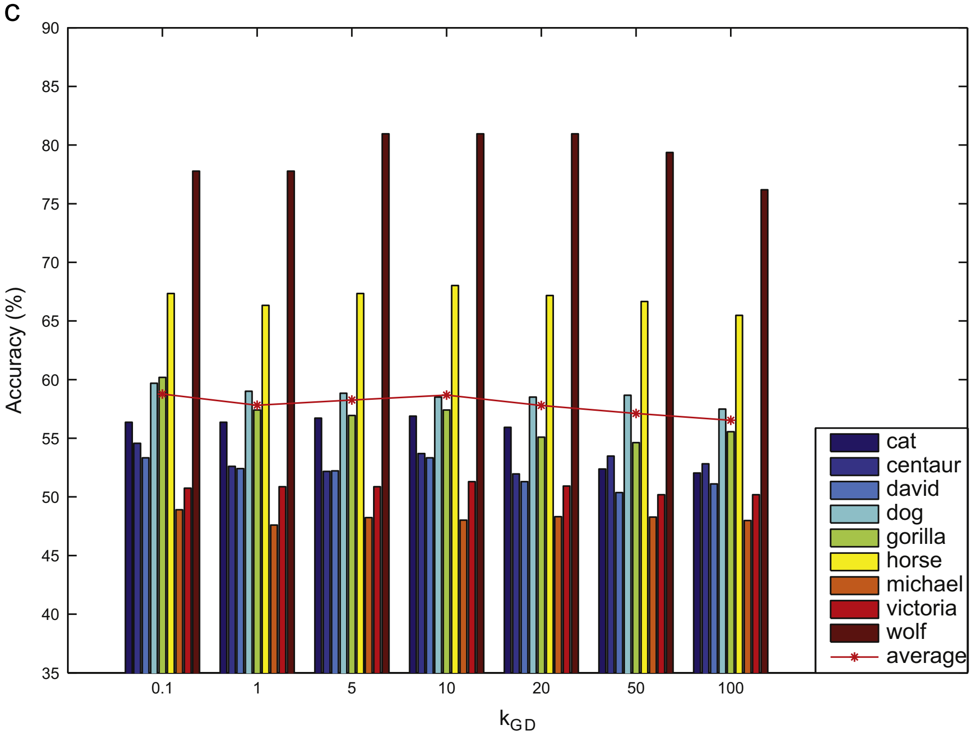

Fig. 3. (*continued*)

a variety of objects with ground-truth correspondences. Shape recognition is also conducted to show the good performance of LDF on TOSCA dataset and SHREC 2007 benchmark [52].

The Watertight dataset contain a total of 380 objects, the TOSCA dataset consist of 80 objects, and 71 objects are in the SCAPE dataset. For each shape dataset, the geometric words are calculated separately, low-level descriptors on all vertices of the randomly selected 50% shapes are used to cluster the geometric words.

For each dataset, the DBN is also trained separately. The training set is constructed by randomly selecting 50% shapes, and for each shape, 400 feature points are sampled through FPS. The procedures of generating LGA-BoF and training the DBNs are time-consuming, especially in the step of learning model parameters of DBNs. We use GPU based deep learning toolbox[1] to accelerate the learning. Compared with Matlab code, the GPU based deep learning speed up 72 times, and the total training procedure just costs several minutes.

The actual timings for different step of the proposed method are listed in Table 1. The timings are measured on a computer with Xeon 3.2 GHz CPU and 16G memory. Calculating the geodesic and SI-HKS are the most time consuming two steps, while the times of other steps are neglectable. Therefore, the proposed method has competitive performance of computation.

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 [48,53,54]. In correspondences experiments, we use above-mentioned three datasets in surface correspondence benchmark [48] as the ground-truth data. In order to assess the performance of the proposed method, we first study the performance of correspondence benchmark through setting different parameters. We use the raw correspondence, which selects pairs with minimal feature distance as estimated correspondence, to obtain the performance measures. The averaged correspondence accuracy is used as the evaluation measure, where it is calculated by averaging the percentage of correct correspondences for all pairs of shapes.

Parameters setting: At first, several experiments are conducted on TOSCA dataset to decide the optimal parameters. We investigate how the region size will affect the correspondence performance. Here we set region size d_l to 3%–19% of shape's maximum geodesic among any pair of vertices. The comparison results are shown in Fig. 3(a), while the horizontal axis represents geodesic ratio and the vertical axis represents the averaged correspondence accuracy. The results show that the performance is not satisfactory when the radius of the region is small. Below 0.08 of the maximum geodesic, the performance is increasing, however, after that geodesic range the performance decreases. According to the observation, we select 0.08 of the maximum geodesic as the radius for the following experiments.

Then, let us see how the different word number affects the correspondence accuracy. We set the number to 60, 80, 100, 120, 140, 160, and 180, respectively, and obtain different performances, which are shown in Fig. 3(b). As can be seen, generally number of bag-of-words (BoW) has little influence on the accuracy. Moreover, larger BoW number can cause the shortcoming that low computation performance results from the rapidly increasing calculation time of LGA-BoF. Therefore, according to the figure, an optimal BoW number of 100 is selected for the following experiments.

Next, we study the effects of different k_{GD}. This parameter controls the decay rate for calculating the LGA-BoF. If a small value is set, a pair of vertices with large geodesic distance still contributes to the LGA-BoF, on the contrary small contribution will be made to LGA-BoF. When its value is turned larger, the LGA-BoF will

[1] The GPU based deep learning toolbox can be downloaded from https://github.com/shaoguangcheng/DeepNet

degrade to BoF because of losing the neighborhood relationship. Fig. 3 (c) shows the correspondence accuracy under different k_{GD}. From the figure, we can see that different k_{GD} has little influence on the accuracy, and when $k_{GD}=10$, the proposed method can achieve best performance.

From the results of above experiments, we note that the proposed feature is insensitive to the parameters of BoW_n and k_{GD}, which indicates our feature's robustness. Finally, we select d_l=0.08, BoW_n=100, and $k_{GD}=10$ as optimal parameters according to the experimental results, and weighted k-means to apply into the proposed method for the following experiments. We construct four layers for DBN including input and output layers. The number of nodes in each hidden layer is empirically set to 1000 and 800, and the node number of output layer is set to 400.

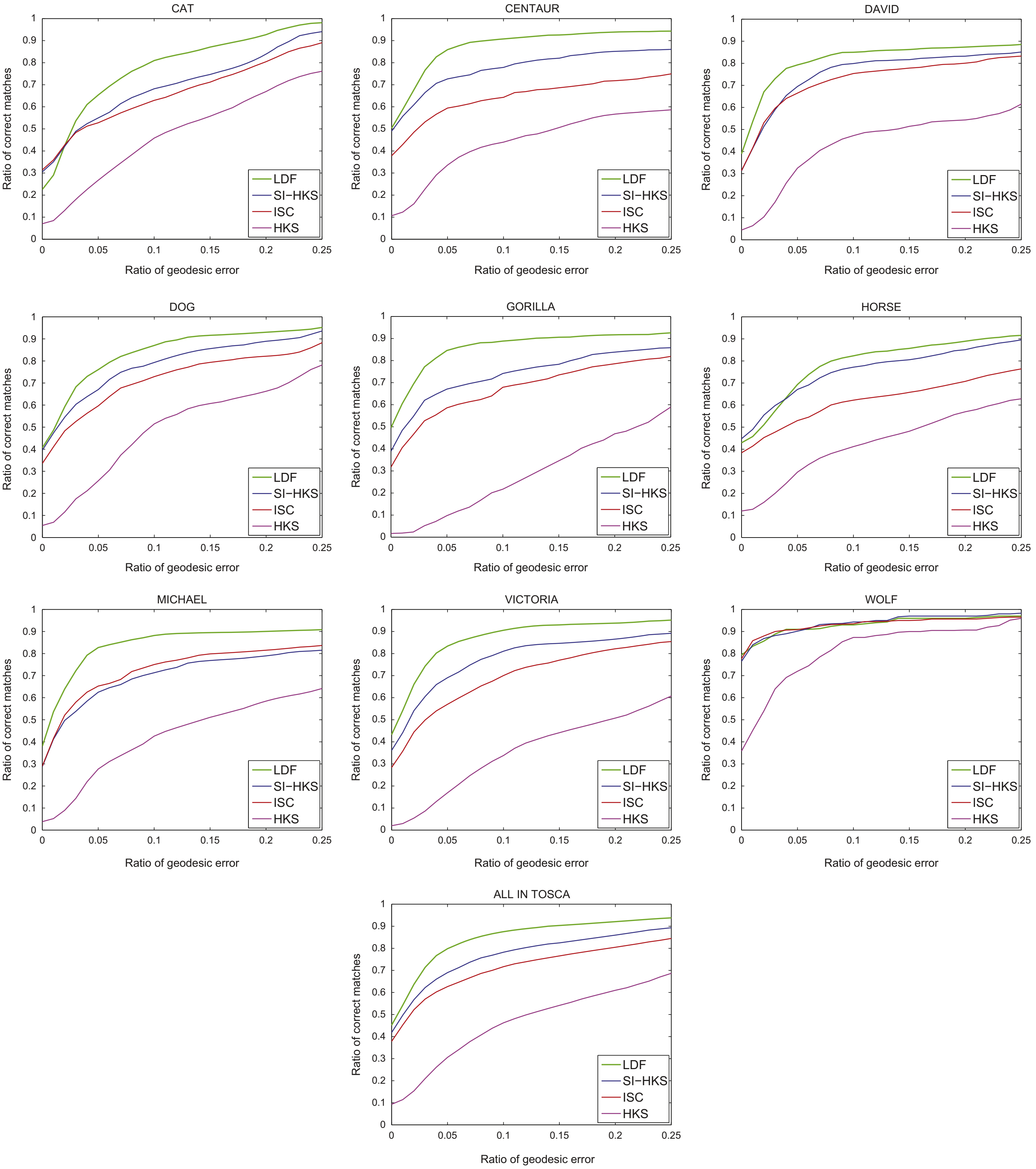

Fig. 4. Averaged correspondence accuracy with different categories, and averaged correspondence accuracy for all categories in TOSCA.

The learning rate is set to be 0.1, the momentum is 0.9, and maximum epoch is 2000.

Experiment on TOSCA. In order to demonstrate the performance of LDF, we compare it with the recent HKS [24], SI-HKS [25], and ISC [29] descriptors on TOSCA models [50], which contain nine categories: cat, centaur, david, dog, gorilla, horse, michael, victoria, and wolf. For each shape, 400 points are sampled via FPS [43], here we define the geodesic distance between ground-truth corresponding vertex and matched vertex on the target shape as geodesic error and spectral correspondence [55] is adopted here. Comparison results are shown in Fig. 4, where the horizontal axis denotes the geodesic error and the vertical axis represents the ratio of correct correspondence, implying that the proposed feature has a better performance obviously. Also, the average accuracy on the all TOSCA models is presented in bottom of Fig. 4, at the geodesic error 0.125 (ratio of geodesic error to shape's maximum geodesic distance), the accuracy can reach 90% with the proposed method, while only 78.5% via using SI-HKS, 74.5% via using ISC, and 50.5% via using HKS. Some of the comparison results are shown in Fig. 5, from which we can conclude that the proposed LDF with spectral correspondence outputs best results.

Experiments on Watertight and SCAPE: In addition, we also use Watertight shapes [49] and SCAPE models [51] and ground-truth maps from [48] to conduct the comparison experiments with above correspondence methods: raw correspondence and spectral correspondence [55]. For the Watertight shapes, there are 19 categories. Experiment result with raw correspondence is plotted in Fig. 6, indicating that our feature has better performance than SI-HKS for almost all categories except the 'bearing' category. As the 'bearing' models just consist of several various of simple geometrical shapes and the LDF and SI-HKS cannot collect enough information of local region, both of them have similar performance. Since the original vertices with the same indexes between models in the same category are not corresponding, we cannot get the correspondence points in this dataset with sampling feature points and spectral correspondence experiment is not conducted on it.

For the SCAPE models, left column in Fig. 7 shows the proposed feature's good performance with raw correspondence method, from right column in Fig. 7 we note that at the geodesic error 0.08 (ratio of geodesic error to shape's maximum geodesic distance) the accuracy can reach 90% with the proposed method, while only 20.1% via using SI-HKS. The outstanding performance of our feature with spectral correspondence method is obvious.

To evaluate the robustness of the LDF, we perform cross validation experiments by randomly picking out different sets of feature points for training the DBN model. The final results of correspondence are slightly different at each time. The error bars

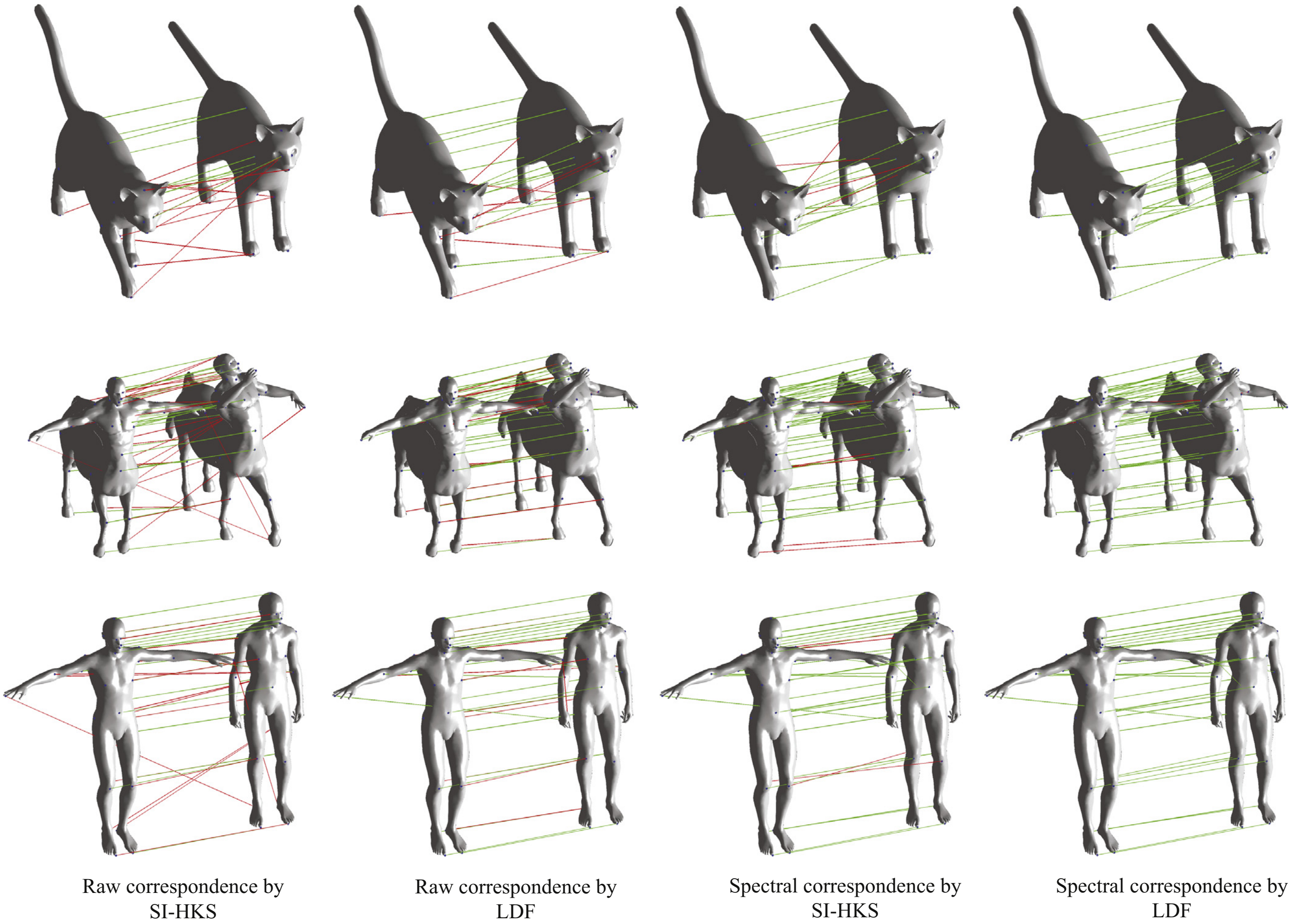
Raw correspondence by SI-HKS
Raw correspondence by LDF
Spectral correspondence by SI-HKS
Spectral correspondence by LDF

Fig. 5. Illustration of correspondences of three shapes from TOSCA by using SI-HKS and proposed LDF feature. 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 by using raw correspondence with SI-HKS and the second column demonstrates the results of raw correspondence with proposed LDF. The third column presents the results of spectral correspondence with SI-HKS and the last column shows the results with LDF. In these figures, the red line indicates wrong correspondence, while the green line indicates correct correspondence. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

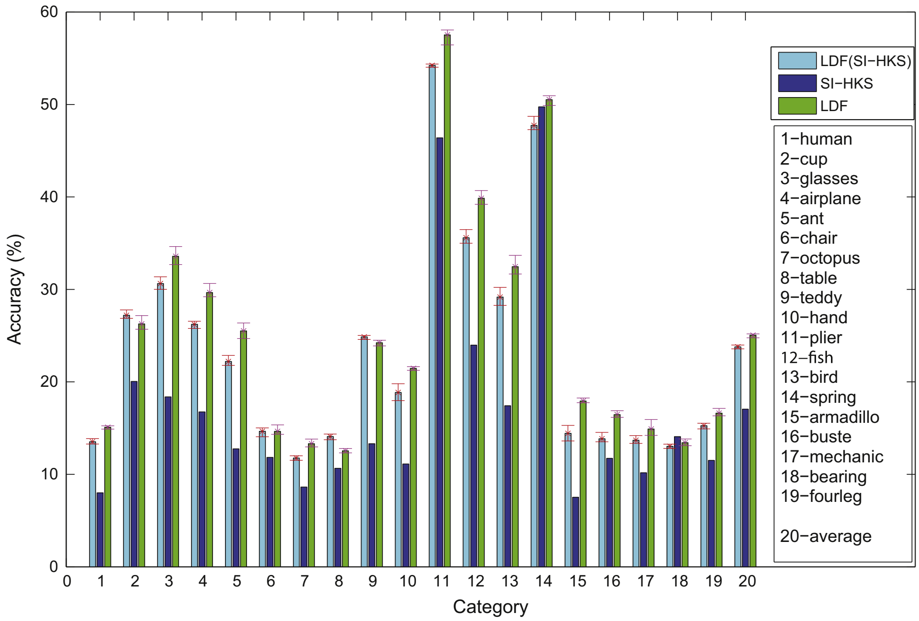

Fig. 6. Comparison of averaged correspondence accuracies on different models from Watertight dataset. In this dataset we use raw correspondence method to conduct the comparison experiments.

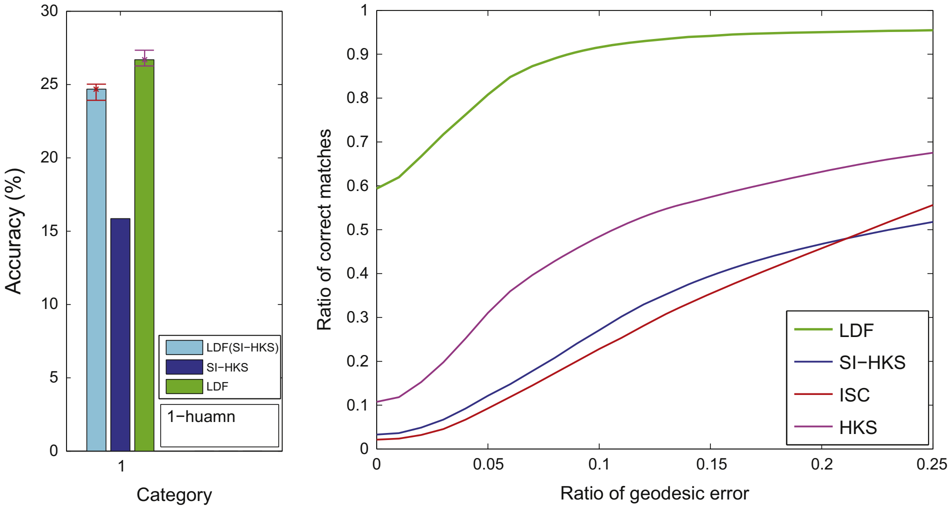

Fig. 7. Averaged correspondence accuracy on SCAPE dataset. Left column shows the experimental result with raw correspondence method; right column shows the experiment result with spectral correspondence method.

about correspondence accuracies are plotted in Figs. 6 and 7. From these two figures, we find that the range of accuracy fluctuation for every category does not exceed 2%. This demonstrates the robustness of the proposed method.

In addition, we only use SI-HKS as the low-level descriptor to validate the effectiveness of our framework. Comparison results are plotted in Figs. 6 and 7. The performances of LDF (only generated with SI-HKS) have improved much compared with SI-HKS on correspondence experiments, which shows that the proposed framework can boost the discriminability of original feature.

From the comparisons of experiments on three datasets, it is worthwhile to note that the boosting performance from SI-HKS to LDF on SCAPE and Watertight models is apparently higher than that on TOSCA. We also find that models in TOSCA are regular and smooth, while shapes from Watertight and SCAPE datasets consist of irregular and rough elements. Therefore, we consider that LDF contains more discriminative information on complex topological mesh than SI-HKS, which also demonstrates the robustness of our LDF generated from the novel framework.

4.2. Experiments on symmetry detection

Besides the correspondence experiments, shape symmetry detection experiment is conducted for evaluating whether the feature is suitable for the task of detecting shape symmetric properties. In this experiment, we implement the symmetry detection in SCAPE models [50] with two methods: the one is raw symmetry detection method and the other one is based on spectral correspondence [55]. The averaged accuracy defined with the ratio of detected symmetric pairs to ground-truth is used as the evaluation measure.

The DBNs model generated from the previous experiments is also used here to get the LDF. We simply implement the symmetry detection, the basic idea is as follows. Different from the original

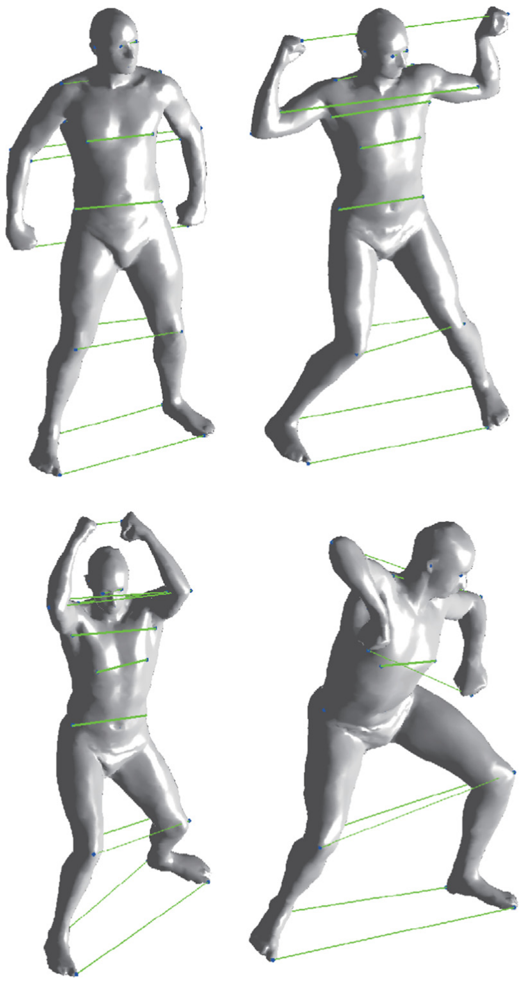

Fig. 8. Several symmetry detection demos about LDF feature with spectral correspondence method (on SCAPE dataset).

correspondence methods, we find the corresponding points between the same shape. First, for each feature point of the shape the second best match is selected as the candidate symmetric point. Second, if the candidate symmetric point of P_a is P_b, and P_a is also the candidate symmetric point of P_b, we regard them as a pair of symmetric points.

The numerical result shows that the accuracy is improved from SI-HKS (27.5%) to the proposed LDF (33.5%) with 6.0% via raw symmetry detection method, our feature achieves better performance obviously. In addition, we also use spectral correspondence method to do the symmetry experiment with the proposed feature, because parameters of the spectral correspondence are sensitive and should be tuned manually, thereby just several demos are showing in Fig. 8. Further research on the adaptive parameters tuning of this symmetry detection based on spectral correspondence will be investigated in the following study.

4.3. Experiments on recognition

Shape recognition experiment is also tested for evaluating whether the feature is qualified to correctly classify set of shapes. In this experiment, we implement a recognition method based on the idea of Shape Google [28] with the proposed LDF on different datasets including TOSCA [50] and SHREC 2007 benchmark [52].

The first step is to sample 400 points on the mesh by using the FPS [43], and then a dictionary is learned through proposed LDF. The shape's global feature is calculated by using the SS-BoF. Finally, we train a DBN model with 50% of all the samples, and test the classification performance with remained data on above two datasets separately.

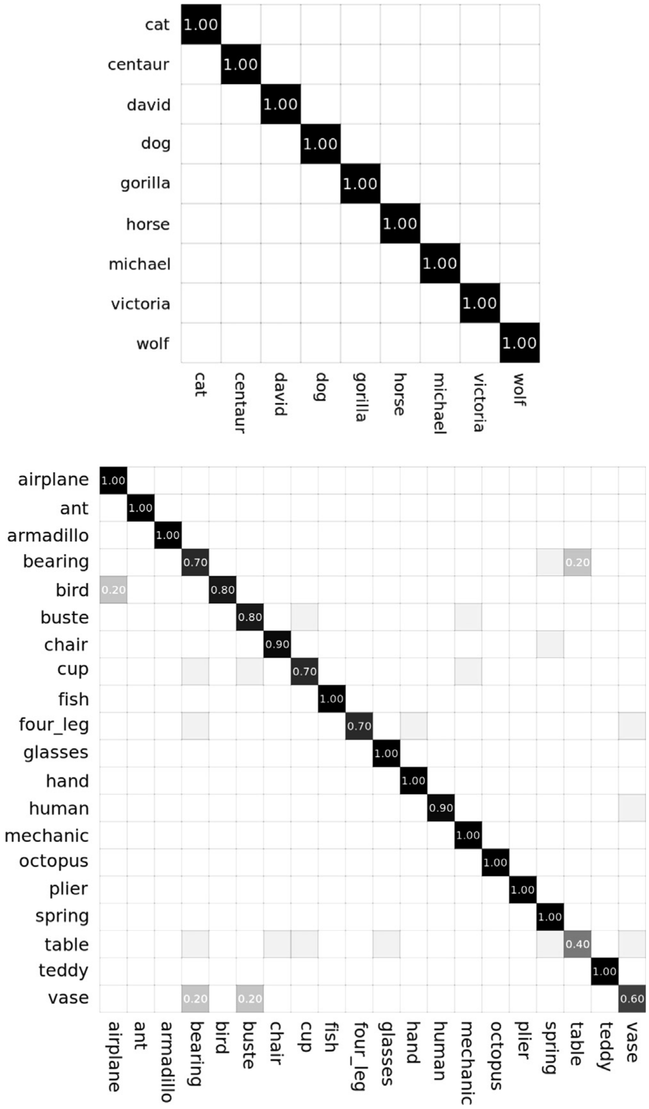

Fig. 9. Confusion matrices calculated by using the proposed LDF on TOSCA (Top) and SHREC 2007 (Bottom).

The confusion matrices of TOSCA and SHREC 2007 are plotted in Fig. 9. The average accuracy by using the proposed LDF is 100% for TOSCA data and 87.5% for SHREC 2007. From the results, we can draw the conclusion that the proposed local feature also has a promising prospect for recognition.

5. Conclusion

Extracting high-level local feature for 3D shape is still a challenging topic up to date, due to the complex structure compared with image data. In this paper, we present a novel high-level local 3D shape feature extraction framework for various applications of 3D shapes.

With the proposed method, in order to preserve the local geometric information of 3D shape, LGA-BoFs are calculated with the decay coefficient of geodesic distance. In previous works [30–32], they have a common limitation that the feature generation is based on extracted iso-geodesic rings and the extraction of iso-geodesic rings on complex region may fail, which might decrease the performance. Compared with these methods, the proposed method just uses the description of vertices in the region with predefined geodesic range to encode the intermediate representation, therefore, this method avoids the problem from the generation of iso-geodesic rings around feature points and can be implemented with better robustness. Also, previous methods require interpolation for obtaining equal-spaced nodes, while our method uses the original vertices on the local area as operational objectives preserving the original and abundant information of the region.

Furthermore, we introduce deep learning method to learn deep relationship between intermediate representations and encode them, which makes the feature full of high-level information and more discriminative. The experiment results demonstrate that the learned high-level feature has better performance on correspondence, symmetry detection, and recognition tasks.

Although the proposed framework achieves better performance, it has a limitation that no hierarchical information can be extracted. In fact, this information is very important to further improve the performance or implement semantic analysis. In the following works, better method will be researched to make full use of the advantages of deep learning. In addition, we use a fixed geodesic ratio to determine the local region which is employed for generating the intermediate representation, however, for different parts of 3D model the selection of the region size should be considered with its structure. To cope with the problem, a scheme which can adaptively determine the region size according to its content will be investigated in the future work.

Acknowledgments

This work is partly supported by grants from National Natural Science Foundation of China (Nos. 61202185, 61003137, 91120005, 61473231), the Fundamental Research Funds for the Central Universities (No. 310201401-(JCQ01009,JCQ01012)), Shaanxi Natural Science Fund (No. 2012JQ8037), Open Project Program of the State Key Lab of CAD&CG (A1306) Zhejiang University.

References

[1] Tangelder JW, Veltkamp RC. A survey of content based 3d shape retrieval methods. Multimed Tools Appl 2008;39(3):441–71.
[2] Bimbo AD, Pala P. Content-based retrieval of 3d models. ACM Trans Multimed Comput Commun Appl (TOMCCAP) 2006;2(1):20–43.
[3] Lian Z, Godil A, Bustos B, Daoudi M, Hermans J, Kawamura S, et al. A comparison of methods for non-rigid 3d shape retrieval. Pattern Recognit 2012;46(1):449–61.
[4] Liu Z, Bu S, Zhou K, Gao S, Han J, Wu J. A survey on partial retrieval of 3d shapes. J Comput Sci Technol 2013;28(5):836–51.
[5] Osada R, Funkhouser T, Chazelle B, Dobkin D. Matching 3d models with shape distributions. In: SMI 2001 international conference on shape modeling and applications. IEEE; 2001. p. 154–66.
[6] Shen Y-T, Chen D-Y, Tian X-P, Ouhyoung M. 3d model search engine based on lightfield descriptors. In: EUROGRAPH interactive demos, Granada, Spain; 2003. p. 1–6.
[7] Tang S, Godil A. An evaluation of local shape descriptors for 3d shape retrieval. CoRR abs/1202.2368.
[8] Heider P, Pierre-Pierre A, Li R, Grimm C. Local shape descriptors, a survey and evaluation. In: Proceedings of the 4th eurographics conference on 3D object retrieval, EG 3DOR'11. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland; 2011. p. 49–56.
[9] Bengio Y. Learning deep architectures for AI. Found Trends Mach Learn 2009;2(1):1–127.
[10] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput 2006;18(7):1527–54.
[11] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504–7.
[12] Johnson AE, Hebert M. Using spin images for efficient object recognition in cluttered 3d scenes. IEEE Trans Pattern Anal Mach Intell 1999;21(5):433–49.
[13] Darom T, Keller Y. Scale-invariant features for 3-d mesh models. IEEE Trans Image Process 2012;21(5):2758–69.
[14] Sipiran I, Bustos B, Schreck T. Data-aware 3D partitioning for generic shape retrieval. Computers & Graphics, Elsevier 2013;37(5):460–72. http://dx.doi.org/10.1016/j.cag.2013.04.002.
[15] 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. Springer; 2010. p. 589–602.
[16] López-Sastre R, García-Fuertes A, Redondo-Cabrera C, Acevedo-Rodríguez FJ, Maldonado-Bascón S. Evaluating 3d spatial pyramids for classifying 3d shapes. Computers & Graphics, Elsevier 2013;37(5):473–83. http://dx.doi.org/10.1016/j.cag.2013.04.003.
[17] Hu J, Hua J. Salient spectral geometric features for shape matching and retrieval. Vis Comput 2009;25(5–7):667–75.
[18] 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: 2010 IEEE conference on computer vision and pattern recognition (CVPR). . IEEE; 2010. p. 438–45.
[19] Dubrovina A, Kimmel R. Matching shapes by eigendecomposition of the laplace-beltrami operator. In: Proceedings of symposium on 3D data processing, vol. 2; 2010.
[20] Lavoue G. Bag of words and local spectral descriptor for 3d partial shape retrieval. In: Eurographics workshop on 3D object retrieval; 2011.
[21] Ben-Chen M, Gotsman C. Characterizing shape using conformal factors. In: Eurographics workshop on 3D object retrieval; 2008. p. 1–8.
[22] Shapira L, Shalom S, Shamir A, Cohen-Or D, Zhang H. Contextual part analogies in 3d objects. Int J Comput Vis 2010;89(2–3):309–26.
[23] Sfikas K, Theoharis T, Pratikakis I. Non-rigid 3d object retrieval using topological information guided by conformal factors. Vis Comput 2012;28(9):943–55.
[24] Sun J, Ovsjanikov M, Guibas L. A concise and provably informative multi-scale signature based on heat diffusion. In: Computer graphics forum, vol. 28. Wiley Online Library; 2009. p. 1383–92.
[25] Bronstein MM, Kokkinos I. Scale-invariant heat kernel signatures for non-rigid shape recognition. In: IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2010. p. 1704–11.
[26] Kovnatsky A, Bronstein MM, Bronstein AM, Raviv D, Kimmel R. Affine-invariant photometric heat kernel signatures. In: Proceedings of eurographics conference on 3D object retrieval, Eurographics Association. Cagliari, Italy; 2012. p. 39–46.
[27] Ovsjanikov M, Bronstein AM, Bronstein MM, Guibas LJ. Shape google: a computer vision approach to isometry invariant shape retrieval. In: IEEE 12th international conference on computer vision workshops (ICCV Workshops). IEEE; 2009. p. 320–7.
[28] Bronstein AM, Bronstein MM, Guibas LJ, Ovsjanikov M. Shape google: geometric words and expressions for invariant shape retrieval. ACM Trans Graph (TOG) 2011;30(1):1.
[29] Kokkinos I, Bronstein MM, Litman R, Bronstein AM. Intrinsic shape context descriptors for deformable shapes. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2012. p. 159–66.
[30] 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. Wiley Online Library; 2008. p. 643–52.
[31] Castellani U, Cristani M, Murino V. Statistical 3d shape analysis by local generative descriptors. IEEE Trans Pattern Anal Mach Intell 2011;33(12):2555–60.
[32] Bu S, Han P, Liu Z, Li K, Han J. Shift-invariant ring feature for 3d shape. Vis Comput 2014;30(6–8):867–76.
[33] Litman R, Bronstein AM. Learning spectral descriptors for deformable shape correspondence. IEEE Trans Pattern Anal Mach Intell 2014;36(1):171–80.
[34] Barra V, Biasotti S. Learning kernels on extended reeb graphs for 3d shape classification and retrieval. In: Eurographics workshop on 3D object retrieval. 2013; p. 25–32.
[35] Kalogerakis E, Hertzmann A, Singh K. Learning 3d mesh segmentation and labeling. ACM Trans Graph (TOG) 2010;29(4):102.
[36] Laga H, Mortara M, Spagnuolo M. Geometry and context for semantic correspondences and functionality recognition in man-made 3d shapes. ACM Trans Graph (TOG) 2013;32(5):150.
[37] Laga H. Semantics-driven approach for automatic selection of best views of 3d shapes. In: Proceedings of the 3rd eurographics conference on 3D object retrieval. Eurographics Association; 2010. p. 15–22.
[38] Tabia H, Picard D, Laga H, Gosselin P-H. Compact vectors of locally aggregated tensors for 3d shape retrieval. In: Eurographics workshop on 3D object retrieval; 2013.
[39] Secord A, Lu J, Finkelstein A, Singh M, Nealen A. Perceptual models of viewpoint preference. ACM Trans Graph (TOG) 2011;30(5):109.
[40] Gao Y, Wang M, Ji R, Wu X, Dai Q. 3d object retrieval with Hausdorff distance learning. IEEE Trans Ind Electron 2014;61(4):2088–98.
[41] Leng B, Xiong Z. Modelseek: an effective 3d model retrieval system. Multimed Tools Appl 2011;51(3):935–62.
[42] Hilaga M, Shinagawa Y, Kohmura T, Kunii TL. Topology matching for fully automatic similarity estimation of 3d shapes. In: Proceedings of the 28th annual conference on computer graphics and interactive techniques. ACM; 2001. p. 203–12.
[43] Peyré G, Cohen LD. Geodesic remeshing using front propagation. Int J Comput Vis 2006;69(1):145–56.
[44] Cordeiro de Amorim R, Mirkin B. Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering. Pattern Recognit 2012;45(3):1061–75.

[45] Smolensky P. Information processing in dynamical systems: Foundations of harmony theory.

[46] Freund Y, Haussler D. Unsupervised learning of distributions of binary vectors using two layer networks. Computer Research Laboratory, University of California, Santa Cruz; 1994.

[47] Hinton GE. Training products of experts by minimizing contrastive divergence. Neural Comput 2002;14(8):1771–800.

[48] Kim VG, Lipman Y, Funkhouser T. Blended intrinsic maps. In: ACM Transactions on Graphics (TOG), vol. 30. ACM; 2011. p. 79.

[49] Giorgi D, Biasotti S, Paraboschi L. Watertight models track. Technical Report; 2007.

[50] Bronstein AM, Bronstein M, Bronstein MM, Kimmel R. Numerical geometry of non-rigid shapes. New York, NY, USA: Springer Science+Business Media, LLC; 2008.

[51] Anguelov D, Srinivasan P, Koller D, Thrun S, Rodgers J, Davis J. Scape: shape completion and animation of people, in: ACM Transactions on Graphics (TOG), vol. 24. ACM; 2005. p. 408–16.

[52] Siddiqi K, Zhang J, Macrini D, Shokoufandeh A, Bouix S, Dickinson S. Retrieving articulated 3-d models using medial surfaces. Mach Vis Appl 2008;19 (4):261–75.

[53] Mykhalchuk V, Cordier F, Seo H. Landmark transfer with minimal graph. Comput Graph 2013;37(5):539–52.

[54] Tevs A, Berner A, Wand M, Ihrke I, Seidel H-P. Intrinsic shape matching by planned landmark sampling. In: Computer graphics forum, vol. 30. Wiley Online Library; 2011. p. 543–52.

[55] Leordeanu M, Hebert M. A spectral technique for correspondence problems using pairwise constraints. In: Tenth IEEE international conference on computer vision, 2005. ICCV 2005, vol. 2. IEEE; 2005. p. 1482–9.