

3D model classification via Principal Thickness Images



Zhenyu Shu^a, Shiqing Xin^{b,*}, Huixia Xu^c, Ladislav Kavan^d, Pengfei Wang^a, Ligang Liu^e

^a School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo, 315100, PR China

^b School of Information Science and Engineering, Ningbo University, Ningbo, 315211, PR China

^c Institute of Mathematics, Zhejiang Wanli University, Ningbo, 315100, PR China

^d School of Computing, University of Utah, Salt Lake City, 84112, USA

^e Graphics & Geometric Computing Laboratory, School of Mathematical Sciences, University of Science and Technology of China, Anhui, 230026, PR China

ARTICLE INFO

Keywords:

Non-rigid 3D model
3D model classification
Principal Thickness Images
Kernel sparse representation

ABSTRACT

With the innovation in 3D modeling software, more and more 3D models are becoming available in recent decades. To facilitate efficient retrieval and search of large 3D model databases, an effective shape classification algorithm is badly in need. In this paper, we propose a new feature descriptor named *Principal Thickness Images* (PTI) that encodes the boundary surface and the voxelized constituents of a 3D shape into three gray-scale images. With the support of PTI, we extend the kernel sparse representation-based classification from 2D case to non-rigid 3D models. Our classification algorithm inherits the robustness of kernel sparse representation and is able to achieve a high success rate and strong reliability on non-rigid models from the SHREC'11 non-rigid 3D models dataset. Numerous tests demonstrate superior performance of the proposed method over previous 3D shape classification approaches.

© 2016 Published by Elsevier Ltd.

1. Introduction

With the fast development of 3D modeling software, the number of 3D models has greatly increased in recent decades. More and more 3D model databases are created for different purposes. Meanwhile, managing or classifying 3D models in the databases becomes an important task. However, manually classifying the 3D models according to their contents into meaningful groups is very tedious and time-consuming. Effective algorithms for 3D model classification are therefore necessary.

A common scenario of 3D model classification is that there are several classes of 3D models (training models) in the database, and then a user provides another 3D model (testing model) and wants to find which class it belongs to. Many 3D model classification algorithms have been proposed in recent years. However, a robust and widely applicable 3D model classification algorithm still remains a challenge because it is difficult to measure the similarity between different 3D models.

A general outline of many 3D model classification algorithms can be described as follows. First, a feature descriptor is chosen

to capture certain geometric features of the 3D models. Second, feature vectors are extracted by applying the feature descriptor to each of the 3D models. Finally, a classifier is applied to predict the membership of the test model by comparing the test model's feature vector with the feature vectors of the training models. The key challenge of such an algorithm pipeline is to find proper 3D feature descriptors. A suitable classifier also plays an important role in 3D model classification.

Traditional view-based 3D model classification or retrieval methods [1,2] usually use projected views of 3D model to measure the similarities between different 3D models. However, they often use only the contour information of the 3D shapes while ignoring their volumetric aspects. In this paper, we extend kernel sparse representation-based classification algorithm (KSRC) [3], originally proposed for 2D face classification, to handle 3D non-rigid surface models. To adapt KSRC for the 3D shape classification purpose, we propose a new feature descriptor which we call *Principal Thickness Images* (PTI); See Fig. 1. As the name suggests, a PTI is a 2D image captured from a particular perspective and encodes not only the contour information but also the volumetric information of a 3D model. The intensity of each pixel in the PTI reflects the thickness of a 3D model along the corresponding view direction. The most significant feature of our PTI distinguishing itself from traditional view-based 3D model classification or retrieval methods is that PTIs are able to characterize the volumetric information for a 3D model. Given a test 3D model whose category is unknown and

* Correspondence to: School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo, 315100, PR China.

E-mail addresses: shuzhenyu@nit.zju.edu.cn (Z. Shu), xinshiqing@nbu.edu.cn (S. Xin).

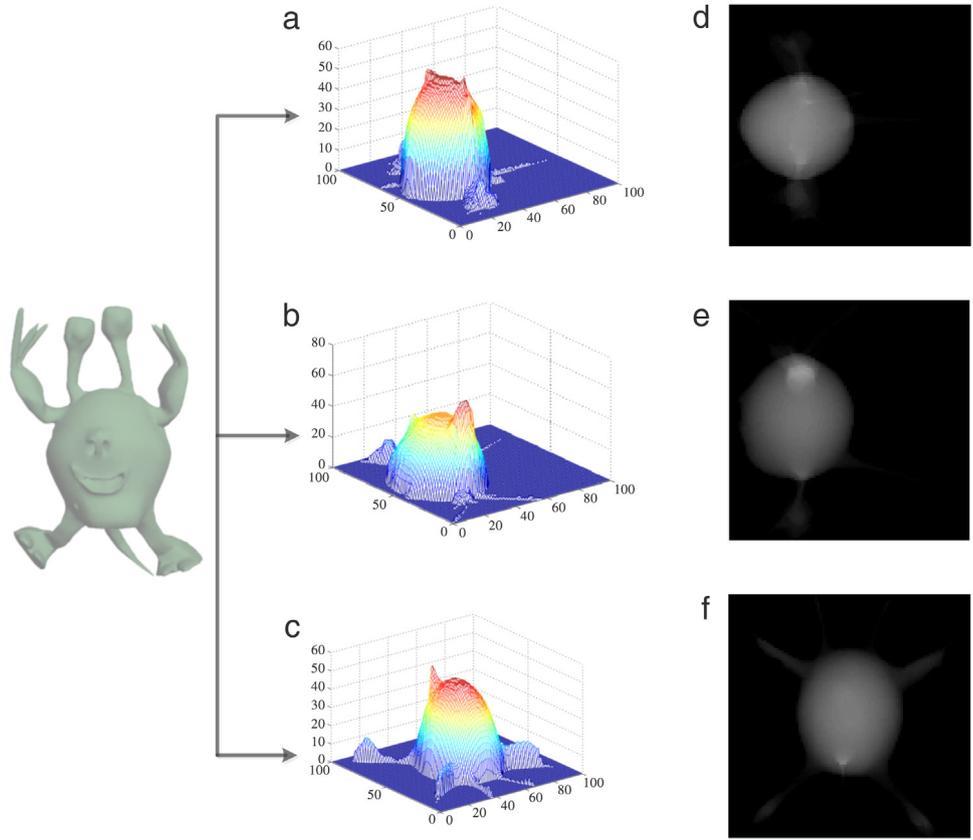


Fig. 1. The PTIs extracted from the Alien model. (a)–(c) show the thickness of the Alien model along the 3 principal axes. (d)–(f) are the corresponding PTIs.

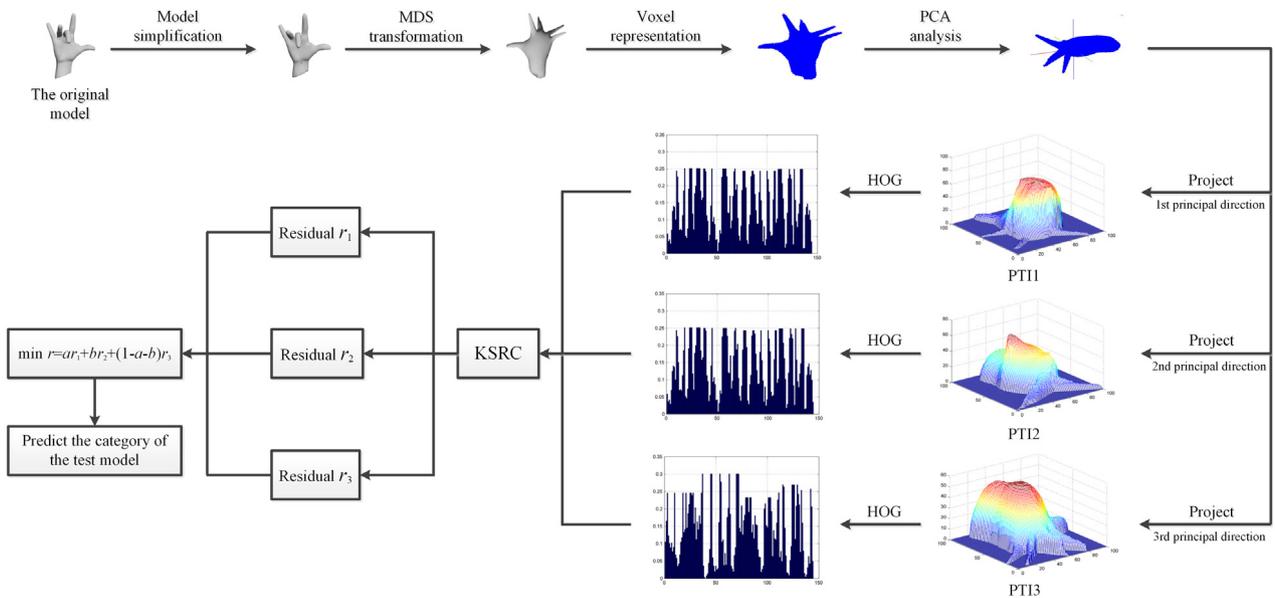


Fig. 2. An overview of the proposed approach for 3D model classification. PCA: Principal Component Analysis; HOG: Histogram of Oriented Gradient; MDS: Multi-Dimensional Scaling; KSRC: kernel sparse representation-based classification; PTI: Principal Thickness Images. The terminology will be clearly elaborated in the following sections.

a 3D model database within which each model has been labeled with a category, our algorithm predicts the class information of the test model in a supervised way by combining the PTI technique and the KSRC. By leveraging multi-dimensional scaling (MDS), our classification algorithm is oblivious to quasi-isometric deformations of 3D models and is thus applicable also to non-rigid 3D models. Experimental results show that our method can

achieve a high success rate. An overview of our algorithm is shown in Fig. 2.

The main contributions of this work are two-fold:

- We propose a new 3D model feature descriptor, named *Principal Thickness Images*. It is not only invariant to the basic geometric transformations such as translation, rotation, and scaling, but

also insensitive to different poses due to MDS. Therefore it is suitable for non-rigid 3D model classification.

- By combining the PTI technique and our extended version of the KSRC, we present a new classification algorithm for non-rigid 3D models. Comparative experiments between our algorithm and other existing methods show that our algorithm can achieve a higher success rate and stronger reliability on non-rigid models from the SHREC'11 non-rigid 3D models dataset.

The remainder of the paper is organized as follows. Section 2 reviews the related work on 3D model classification. Section 3 describes the novel feature descriptor of PTI in detail. In Section 4, we propose the complete framework of the classification algorithm. Finally in Section 5, we show extensive experimental results on the SHREC'11 non-rigid 3D models dataset, as well as statistical comparisons with other approaches.

2. Related work

2.1. Shape descriptors

Feature descriptors are central to 3D model classification. In recent years, numerous feature descriptors have been proposed, such as shape distribution [4], spherical harmonics [5], light field [6] or spectral methods [7], to name just a few.

Existing 3D model feature descriptors can be divided into two main categories. The first one is geometry based feature descriptors. This kind of feature descriptors focus on describing 3D models by using geometric measures. For example, Vranic [8] constructed a rotation invariant feature vector by using functions on concentric spheres. Kazhdan et al. [9] classified 3D models by using a symmetry descriptor, which consist of a collection of spherical functions that describes the measure of a 3D model rotational and reflective symmetry with respect to every axis passing through the center of mass. Ricard et al. [10] extended the angular radial transform to handle 3D models and used it for classification and retrieval. Knopp et al. [11] proposed a 3D model classification method by using Hough-voting for local 3D features. Bronstein and Kokkinos [12] proposed a local 3D model feature descriptor by using heat kernel signature. Their feature descriptor is defined on each vertex of a 3D model and can be integrated into bag-of-features framework for classification. By detecting and comparing salient points on the mesh, Smeets et al. [13] proposed a local 3D model feature descriptor for 3D face recognition, named meshSIFT, which is robust to expression variations, missing data and outliers. Lu et al. [14] proposed a 3D model retrieval and recognition approach, which employs both a distance histogram and 3D moment invariants as features. In their algorithm, they measured the distance of 3D models by disjoint information in the feature space in order to obtain the difference between the sets of 3D models.

The second category is vision based feature descriptors. Unlike the geometry based feature descriptors, the vision based feature descriptors try to convert 3D models to 2D images first and then apply some methods coming from computer vision to obtain the feature vectors. For example, Loffler [15] used the corresponding 2D images to compare the similarity between any two 3D models. Li et al. [16] proposed a 3D model classification method based on nonparametric discriminant analysis with kernels and geometry projection-based histogram model. Atmosukarto and Shapiro [17] presented a method for selecting salient 2D views to describe 3D objects. Lu et al. [2] proposed a view based approach that jointly learns the relevance among 3D objects using a hypergraph structure. Gao et al. [18] presented a hypergraph analysis algorithm by integrating multiple hypergraphs for 3D object retrieval and recognition. Their method does not need any training 3D models,

which is very useful when appropriate databases are not available. Lu et al. [19] proposed a distance metric learning algorithm for bipartite graph match-based 3D model retrieval framework. They estimated object relevance by employing a graph structure with semi-supervised learning; the learned metric can further enhance the 3D model retrieval performance. Gao et al. [20] presented a 3D model retrieval algorithm by using a learned view-level Mahalanobis metric, which estimates and improves the effectiveness of Hausdorff distance measure for comparing the similarity between any two 3D models. Su et al. [1] proposed a 3D model classification algorithm, which uses multiple views and convolutional neural network simultaneously to achieve good performance.

2.2. Classification approaches

After extracting feature vectors from 3D models, a classifier is always required for 3D model classification. Commonly used classifiers include decision trees [21], naive Bayesian classification [22], support vector machine [23], or probabilistic neural networks [24]. It is worth pointing out that combining feature descriptors and suitable classifiers is important for improving the success rates of classification. Specifically for 3D model classification, Leng et al. [25] used deep belief networks to extract feature and classify 3D models, while Liu et al. [26] used spectral embedding and Adaboost. Nian et al. [27] achieved 3D model classification by combining geometrical topology hypotheses and extreme learning machine. Au et al. [28] found that establishing the sparse correspondence between two given shapes can be efficiently solved by considering the underlying skeletons augmented with intrinsic surface information. The algorithm can be further used for 3D model classification or retrieval based on the correspondence matching between 3D models.

In [3], the authors proposed a new classification algorithm for well-aligned 2D human face based on sparse representation with great success. In this paper, we extend their approach to support non-rigid 3D mesh models. However, this is not straightforward. One possible idea would be to convert 3D models from the boundary representation to voxel representation, where each 3D model is represented as $L \times L \times L$ binary image and each voxel is marked as *inside* (1) or *outside* (0). The vectorized 3D image can be directly integrated with kernel sparse representation classifier. However, this approach is not very practical, because it is very hard to find a suitable L . If L is small, the voxel representation is not sufficient to capture the volumetric details of 3D models. If L is large, the dimensionality will increase sharply, unnecessarily leading to effective classification due to the curse of dimensionality.

One possibility to tackle this issue is to project 3D models to 2D images and use the resultant images for classification purpose. The rationale behind this is that a 3D model can be reconstructed from a sufficiently large number of 2D images. For example, Brubaker et al. [29] presented an efficient framework for 3D molecular reconstruction from electron cryomicroscopy images. However, if we simply project 3D models to 2D images and directly use the contour images for classification, the results may be not desirable because (1) the images should be obtained from the same perspectives and thus the input 3D models should be aligned well and (2) the volumetric information is missing in the transformation process. Therefore, in this paper we propose a new 3D model feature descriptor named PTIs that consist of three axis-aligned 2D images using PCA and encode contours and volumetric information at the same time. After PTIs are extracted, we transform them into feature vectors and then feed them into kernel sparse representation-based classifier to perform classification.

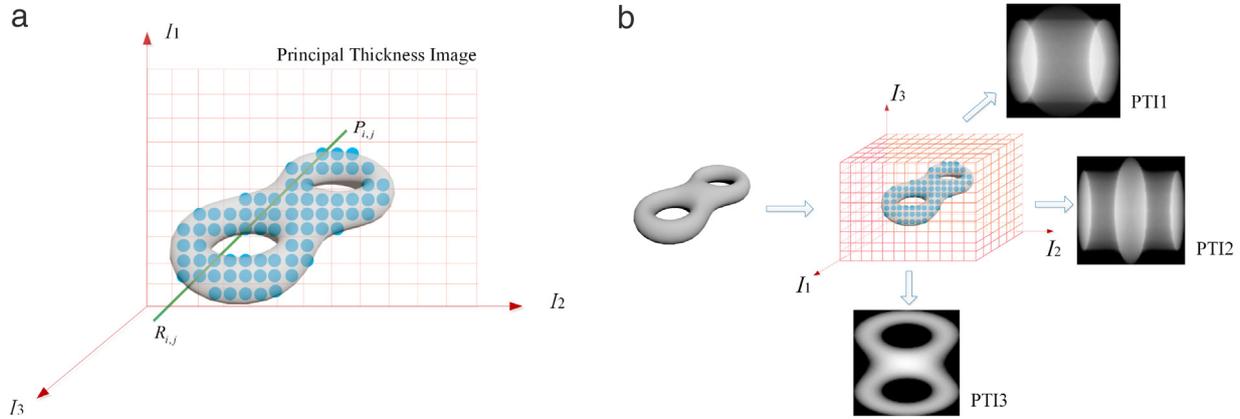


Fig. 3. The mapping process and the result of constructing the PTIs for the Eight model. (a) shows the mapping process, where the intensity of each pixel on the PTI is determined by counting the voxels inside the model along the ray $R_{i,j}$. (b) shows the result PTIs extracted from the model.

3. Principal thickness images

3.1. Preprocessing

For non-rigid 3D model classification, a family of models with different poses should be deemed to be in the same class. Therefore, it is better to perform a step of pose removal to guarantee the pose-oblivious property. For such a purpose, we first apply multi-dimensional scaling [30] (MDS) to transform the models from the same class into visually quite similar models; the resultant models can be well aligned up to a certain rigid transformation. Let M be the original 3D model, M' the transformed version by applying MDS, $\delta_{i,j} = g(v_i, v_j)$ the geodesic distance between any two vertices v_i and v_j on M , and $\delta'_{i,j} = d(v'_i, v'_j)$ the Euclidean distance between v'_i and v'_j on the transformed model. The function of MDS here is stretching the points outward while keeping $\delta_{i,j}$ and $\delta'_{i,j}$ to be as close as possible, i.e.,

$$M' = \arg \min \sum_{v'_i, v'_j} \|\delta_{i,j} - \delta'_{i,j}\|_2^2. \quad (1)$$

In implementation, we use Xin and Wang's approach [31] to efficiently compute the geodesic distance between each pair of vertices of the 3D models. In order to solve Eq. (1), we first transform the distance matrix into a cross-product matrix and then find its eigen-decomposition.

Ideally, the geodesic distance between any two vertices on the surface of a 3D shape remains unchanged with regard to pose changes. It is worth pointing out that, however, in real situations the geodesic distances may change slightly upon the MDS transformation. Fortunately, we will use the PTI technique that is able to extract the dominate thickness features and thus the slight difference does not affect the classification performance.

Besides, MDS has a side effect—stretching the input model outward as far as possible. This is helpful turning invisible parts to be visible from some perspective. Thus nearly all the constituents of the input model can be encoded into the final PTIs.

When the input model consists of hundreds of thousands of triangles, solving the above MDS will be very inefficient. The reasons are two-fold. First, MDS, in its own form, requires solving a linear system. The coefficient matrix encoding all-pair distances is dense and of a size $n \times n$, where n is the number of vertices. Second, querying the geodesic distances between all pairs of vertices is of a high computational complexity. Therefore, it is reasonable to simplify the input models while keeping the geometric shapes almost unchanged. In this paper, we employ the famous simplification algorithm, called *QEM*, proposed by Garland

and Heckbert [32] to simplify each 3D model to contain 1000 vertices before applying MDS. We shall discuss how the sampling rate affects the classification in Section 5.2.

3.2. Principal axes computation

Our PTI feature descriptor measures the thicknesses of 3D models along certain directions and maps 3D models to 2D images, where the thickness information is encoded, along the directions. The directions are computed by extracting the principal axes of a 3D model. We first convert 3D models from the boundary representation to the volumetric representation (described in more detail below), then use the principal component analysis [33] (PCA) to compute the principal directions. For a given point set, the first principal direction describes the direction of largest variance of the point set projected to this direction. The third principal direction describes the direction of least variance. If we map a 3D model to a 2D image along the third principal direction, the resulting 2D image should best preserve the contour information of the 3D model.

For the input 3D models which are represented by 3D meshes, we first convert them from the boundary representation to a volumetric representation by dividing the bounding cube of the 3D models, which has equal edge lengths, into $L \times L \times L$ parts uniformly. (Different choices of L are discussed in Section 5.2.) Then we apply PCA to get three principal axes for the volumetric representation. Let c_i ($i = 1, 2, \dots, n$) be the centers of the voxels inside our 3D model M' . We apply the PCA to the point set $\{c_i\}$ and denote the three principal directions as I_1, I_2, I_3 .

3.3. Principal thickness images extraction

We measure the thicknesses of 3D models along the three principal axes separately and encode the thickness information to 2D images. Therefore, the 3D model M' is first re-voxelized along the principal axes, and then the thickness is measured by counting the number of voxels along the directions of the principal axes. To encode the thickness information to 2D images, we map M' from the voxel-based representation to three 2D images PTI1, PTI2, PTI3 along the principal axes I_1, I_2, I_3 respectively. The intensity of each pixel $P_{i,j}$ in PTI1, PTI2, PTI3 is the count of the voxels inside M' along the ray $R_{i,j}$ emitting from $P_{i,j}$, where $R_{i,j}$ is in parallel with the corresponding principal axis (see Fig. 3(a)). We say that PTI1, PTI2, PTI3 are the *Principal Thickness Images* (PTIs) for the 3D model M . Fig. 3 describes the mapping process and the result PTIs for the Eight model.

From the construction of our PTIs, we can see that each PTI is a gray scale image and each pixel in PTI actually describes the

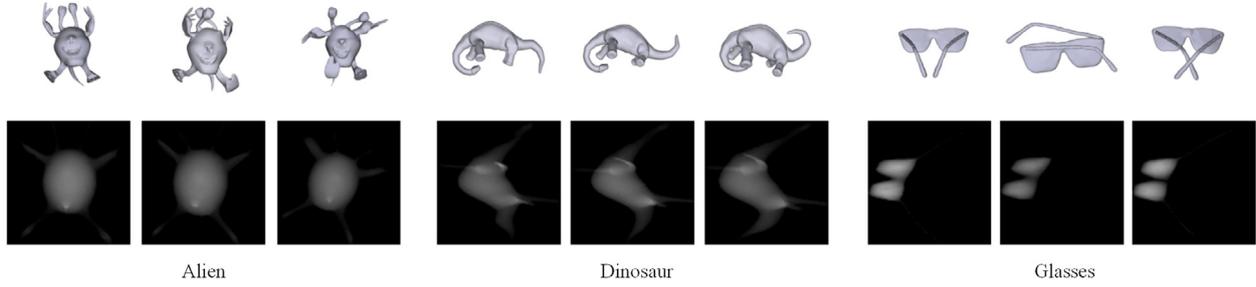


Fig. 4. Examples of PTI3 (captured along the least significant principal direction) extracted from 3 Alien models, 3 Dinosaur models, and 3 Glasses models. At the top are the 3D models. At the bottom are the corresponding PTIs.

thickness of the 3D model along the corresponding ray. If we project a 3D model directly to a 2D image without encoding the thickness information of the 3D model, the volumetric information of the 3D model will be lost, although the contour information will remain. Unsatisfactory results may be obtained if we use the 2D image without thickness information for the 3D model classification because too much information is lost in the mapping process. By extracting the PTIs from a 3D model, not only the contour information, but also the volumetric information are encoded. Note that the PTIs extracted from 3D models are well aligned because the mapping directions are the principal axes of 3D models. Therefore, high success rates may be obtained by using our PTIs for 3D model classification. In the following, we summarize the procedure of PTI extraction into Algorithm 1.

Algorithm 1 PTI extraction algorithm

Step 1: Apply the MDS to the given 3D model M to get a new 3D model M' .

Step 2: Voxelize M' and compute the principal axes I_1, I_2, I_3 of M' by applying PCA to the centers of the voxels inside M' .

Step 3: Re-voxelize M' along I_1, I_2, I_3 and compute the 3 PTIs by measuring the thickness of M' along I_1, I_2, I_3 respectively.

Fig. 1 shows PTIs extracted from the Alien model along the 3 principal axes respectively. It can be seen that the PTIs are able to describe the thickness information of the Alien model (see Fig. 1(a), (b), (c)). Fig. 4 shows examples of the PTI3 extracted from 9 different 3D models from 3 different classes. We can see that the PTIs extracted from the same non-rigid class are quite similar in spite of their much different poses.

Our PTIs have several desirable geometric properties. First, PTIs are invariant to translation and rotation because the input model has to experience a PCA step before PTI generation. Second, PTIs are also invariant to scaling. The main reason is that in our algorithm a bounding box is built first and then always divided uniformly into $L \times L \times L$ parts for a 3D model and its any scaled version. The resolution of PTIs is therefore fixed to $L \times L$.

4. 3D model classification using PTI

4.1. Feature vector extraction from PTI

To facilitate 3D model classification using PTIs, feature vectors need to be extracted in advance from these images. For this purpose, we use the histogram of oriented gradients [34] (HOG) to extract a feature vector for each PTI. The HOG is a type of commonly used feature descriptors used in computer vision for object detection. The spirit is to count occurrences of gradient orientation in localized portions of an image. It is often performed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

In our experiments, the size of HOG cells is set to 32×32 . Fig. 5 shows an example of extracting feature vectors from the PTIs using the HOG technique. In Fig. 5, it can be seen that the feature vectors of the two Woman models with different poses are quite similar to each other although the two poses are visually much different.

4.2. Kernel sparse representation-based classification

Our 3D model classification algorithm combines the PTIs and an extended version of the kernel sparse representation-based classifier. Given a test model T and a set of 3D mesh models M_i (training models), which are divided into k classes according to their content in advance, the goal of our algorithm is to predict which class the model T belongs to. First, the PTIs are computed for the test model T and each model in the training model set M_i . Second, three feature vectors are extracted from the PTIs for each 3D model. Finally, our extended version of kernel sparse representation-based classifier [3] (KSRC) is employed and the class which the test model T belongs to is predicted. The KSRC is the kernel version of the sparse representation-based classifier [35] (SRC). For brevity, we give a brief introduction to SRC and KSRC in Appendix.

4.3. Extended KSRC

The original KSRC determines the category with only a single feature vector as the input. However, for each 3D model, we generate three PTIs and thus three feature vectors. Therefore, we need to extend the original KSRC to support 3 feature vectors as the input simultaneously. The technique is to optimize the mixed residual:

$$m = \arg \min_i (a \cdot r_{i,1} + b \cdot r_{i,2} + c \cdot r_{i,3}),$$

where $r_{i,1}, r_{i,2}, r_{i,3}$ are the residues of the PTIs produced by the i th 3D model, and a, b, c are non-negative weights following $a + b + c = 1$. Our extended KSRC is described in Algorithm 2.

In Algorithm 2, the parameters a, b in the optimization problem (2) can be specified by the user or learned from the training data automatically. The learning process is as follows. First, we randomly divide the given training 3D models into 2 groups. The numbers of the models from the same class are the same in each group. Second, one group is regarded as the training model set and the other group is regarded as the testing model set. Finally, we iterate the possible discrete values of the parameters a, b by setting the step to 0.1 and running our classification algorithm. The values of the parameters to achieve the highest success rate are chosen as optimal for subsequent classification. Experimental results are shown in the following section.

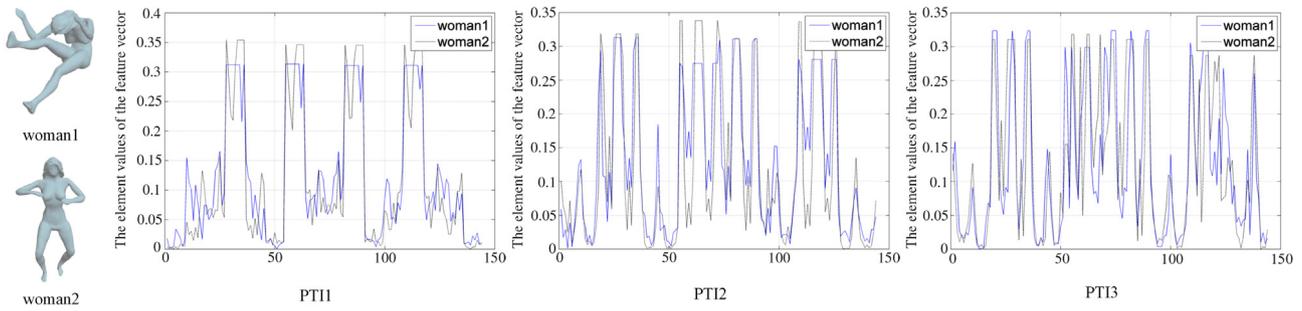


Fig. 5. The feature vectors extracted from the PTIs of two Woman models with different poses. It can be seen that the feature vectors are very close to each other.

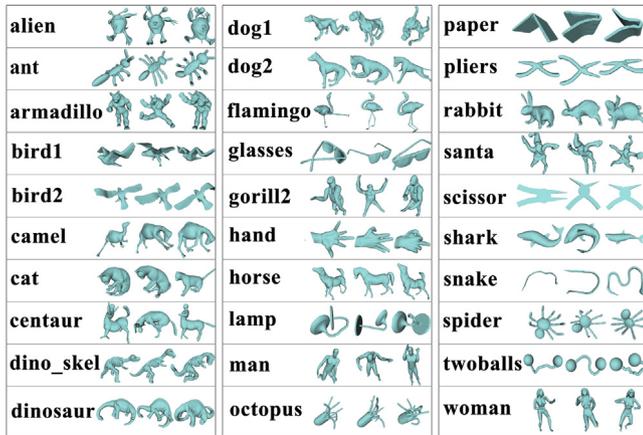


Fig. 6. Some example models of the SHREC'11 non-rigid 3D models dataset.

Algorithm 2 Extended kernel sparse representation-based classification algorithm

Given feature vectors Y_1, Y_2, Y_3 for the test 3D model T and feature vectors $a_{i,1}, a_{i,2}, a_{i,3}$ for each training 3D model M_i ,

Step 1: Solve the optimization problem (A.4) for feature vectors Y_j and coefficients $a_{i,j}$ when $j = 1, 2, 3$ respectively.

Step 2: Predict the membership m of the test 3D model T by minimizing the following mixed residuals:

$$m = \arg \min_i (a \cdot r_{i,1} + b \cdot r_{i,2} + (1 - a - b) \cdot r_{i,3}), \quad (2)$$

where $r_{i,1}, r_{i,2}, r_{i,3}$ are the corresponding residuals getting from Equation (A.7) for each feature vector and $a, b, 1 - a - b \in [0, 1]$ are weights specified by users or learned from the training data.

5. Experimental results and discussion

5.1. Experimental results

We test our algorithm on the SHREC'11 non-rigid 3D models dataset [36] (<http://www.itl.nist.gov/iad/vug/sharp/contest/2011/NonRigid/>). To our knowledge, it is the largest non-rigid 3D model dataset which is publicly available. The dataset contains 600 models, which are categorized into 30 classes, 20 models in each class. Some example models are shown in Fig. 6.

To test the classification performance of our algorithm, we randomly select ten 3D models from each class in the SHREC'11 non-rigid 3D models dataset and regard them as test models. The rest of the models are regarded as training models. Totally there are 300 testing models and 300 training models. For each test model, we use 1 to denote “success” and 0 “failure”. The overall success rate of our algorithm is then computed by averaging the prediction accuracy over all the test models. Fig. 7 shows the

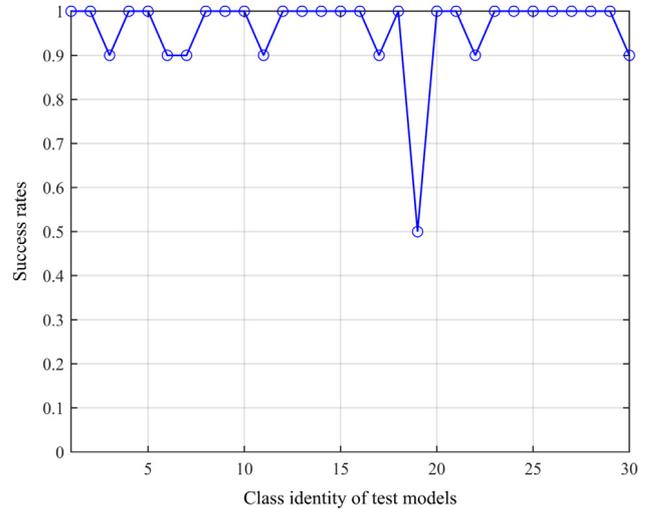


Fig. 7. Success rate plot curve of our algorithm on the SHREC'11 non-rigid 3D models dataset.

success rates of our algorithm on each class, and the average success rate amounts to 96%. It can be seen that our algorithm achieves a high classification performance for most classes.

5.2. Discussion

Robustness to different sampling rates. Fig. 8 shows the feature vectors extracted from the simplified models of the Alien model with various resolutions. The original Alien model contains 5000 vertices. We simplify it to 2500, 1000, 750, 500 vertices respectively and then compute the corresponding feature vectors. We can see that both the principal thickness images and the feature vectors are quite similar. Due to the property that PTIs are robust to different sampling rates, it is reasonable to simplify the models into a small-scale one, typically consisting of 1k vertices, in order to accelerate the computation while keeping the classification performance nearly unchanged.

Setting of parameters a, b . Note that there are 2 parameters a, b in Eq. (2), which need to be determined before the classification. The parameters are automatically learned; See details in Section 4.3, Fig. 9 shows the success rates for different settings of a and b . From the figure, we can see that when $a = 0.3$ and $b = 0.3$, our algorithm achieves the highest success rate of 96%, and thus we take (0.3, 0.3, 0.4) as the best weighting choice on the three residues. The rationale behind is that the PTI captured along the least significant principal direction encodes the richest information of the volumetric thickness.

Setting of the parameter L . When computing the PTIs for the given 3D model, we need to discretize the volume into $L \times L \times L$ regular grids. Fig. 10 shows the success rates of our experimental results

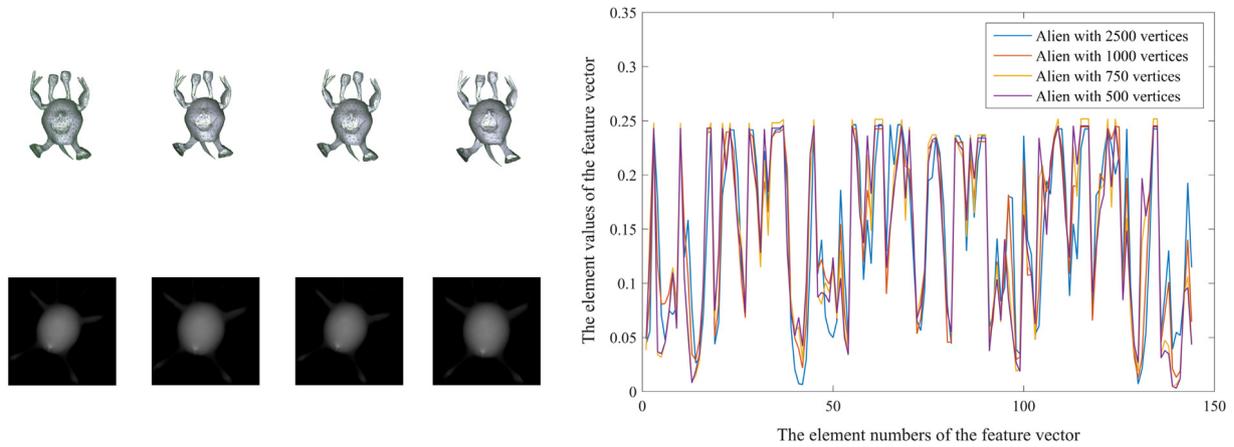


Fig. 8. The feature vectors extracted from PTIs for the Alien model with various sampling rates.

for different choices of L . Generally, a larger value of L usually leads to a higher success rate. However, a larger L will lead to a higher time cost; See Fig. 11 for the timing plot with regard to L . We observe that $L = 250$ makes a good balance between classification performance and efficiency and thus we discretize the 3D models into $250 \times 250 \times 250$ cells in our experiments.

5.3. Comparison with state-of-the-art methods

To demonstrate the advantages of our algorithm, we compare our algorithm with eight existing methods. Totally we have 9 methods, including

- PTI + KNN [37];
- PTI + SVM;
- PTI + PNN [24];
- SIHKS + BOF [12];
- SDF [38] + KSRC;
- SDF + KNN;
- Pickup et al. [39];
- Su et al. [1];
- PTI + KSRC (our algorithm in this paper).

We explain the abbreviated terminology as follows:

- KNN denotes the k -nearest algorithm. An object is classified by the majority votes of its neighbors, with the object being assigned to the class most common among its k nearest neighbors.
- SVM denotes the support vector machine. An SVM model maps the models to be classified into a high-dimensional space so that they can be divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.
- PNN denotes the probabilistic neural network. It is a feedforward neural network, typically consisting of 4 layers, and derived from the Bayesian network and a statistical algorithm called Kernel Fisher discriminant analysis.
- SIHKS denotes the scale-invariant heat kernel signature. It is a feature descriptor for use in deformable shape analysis and belongs to the group of spectral shape analysis methods. For each point in the shape, SIHKS defines its scale-invariant feature vector representing the point's local and global geometric properties.
- BOF denotes a bag of features or words. It takes a collection of high-dimensional vectors as the input and then outputs a histogram that is the occurrence counts with regard to a vocabulary of representative features.

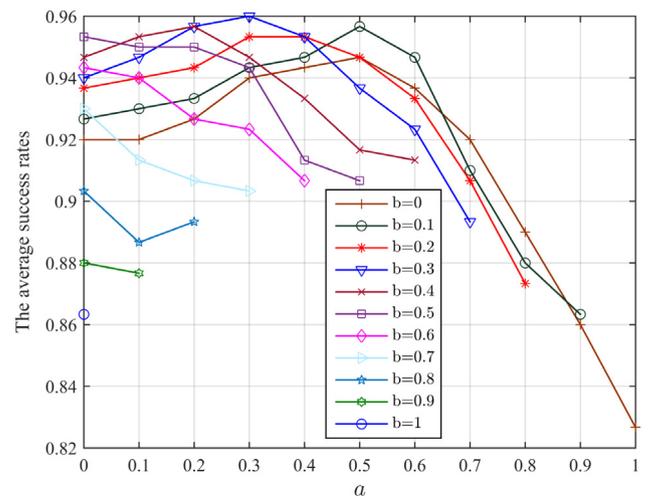


Fig. 9. The average success rates for the different values of the parameters a and b .

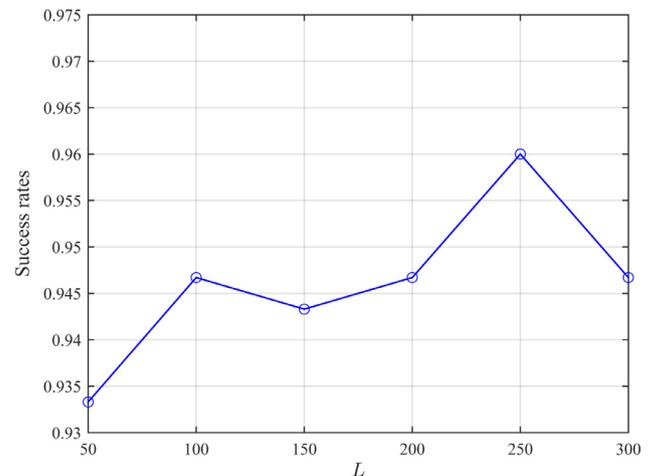


Fig. 10. The success rates for different values of the parameter L .

- SDF means the shape diameter function which is a scalar function defined on a closed manifold surface, measuring the neighborhood diameter of the object at each point. Due to its pose oblivious property, SDF is widely used in shape analysis, segmentation and retrieval.
- Pickup et al. represents the method proposed in [39];
- Su et al. represents the method proposed in [1];

Table 1

The average success rates of our algorithm, as well as those of the other algorithms on the SHREC'11 non-rigid 3D models dataset.

Algorithm	Our algorithm	PTI + KNN	PTI + SVM	PTI + PNN	SDF + KNN	SDF + KSRC	SIHKS + BOF	Pickup et al. [39]	Su et al. [1]
Success rate	96.00%	92.67%	85.33%	91.67%	87.67%	92.33%	95.00%	94.33%	92.67%

Table 2

The performance of our algorithm on the SHREC'11 non-rigid 3D models dataset. The running time is measured for each model except the KSRC step, which is measured on the whole dataset.

Step	Model simplification	Geodesic distances computation	MDS	Voxelization and PCA	HOG	KSRC on the whole dataset
Running time (s)	0.5	11.9	28.1	4.7	0.015	17.0

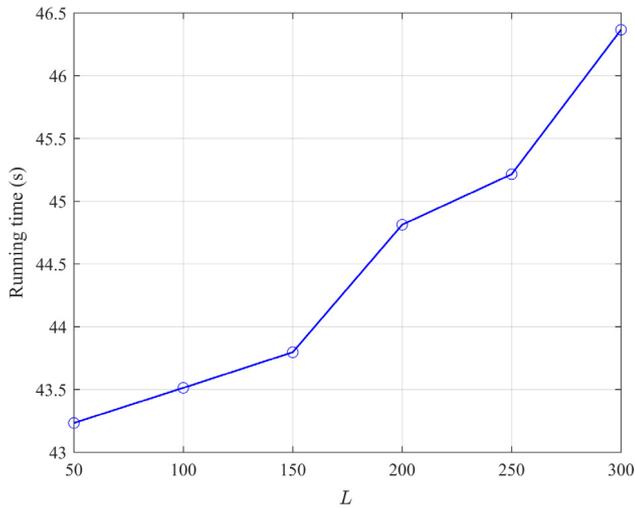


Fig. 11. The average running time of extracting PTI feature vectors for each 3D model in the SHREC'11 non-rigid 3D models dataset with regard to the different values of the parameter L .

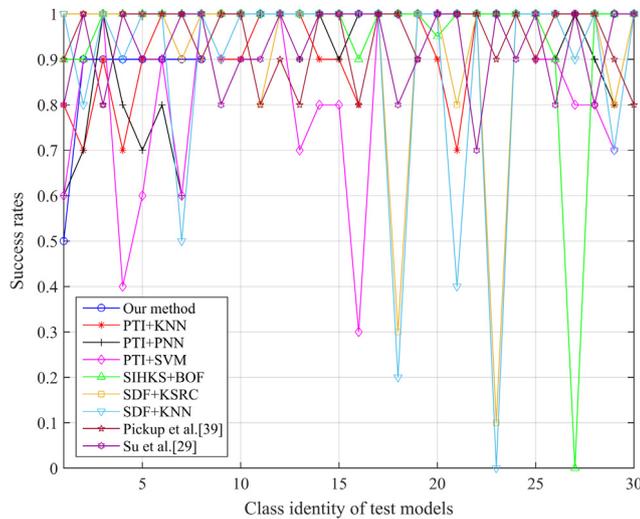


Fig. 12. Comparison of success rates between our algorithm and other algorithms for each class of the SHREC'11 non-rigid 3D models dataset.

We test these algorithms on the SHREC'11 non-rigid 3D models dataset. The success rates for each class are shown in Fig. 12. The average success rates of each algorithm are listed in Table 1. We can see that our algorithm is able to obtain a higher success rate than not only SIHKS + BOF, which is surface-based, but also SDF + KSRC and SDF + KNN, which are also volume-based.

We also show the robustness to noise of our algorithm. For each 3D model in the SHREC'11 non-rigid 3D models dataset, we add Gaussian noise, with the intensity varying from 0.1% to

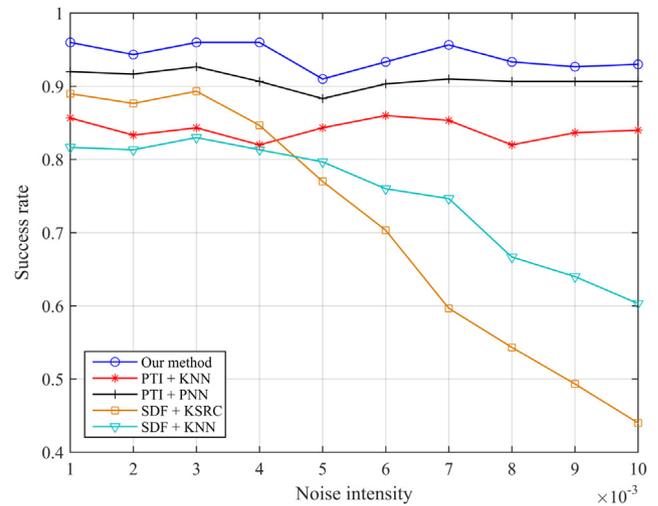


Fig. 13. Comparison of robustness between our algorithm and other algorithms on the SHREC'11 non-rigid 3D models dataset.

1%, to each vertex's coordinates of 3D models in the dataset. Then we run each classification algorithm on the noisy 3D model dataset. The classification accuracy statistics are shown in Fig. 13. Our method is robust due to the combination of partial volumetric information and the kernel sparse representation-based classification. Note that although the SDF is also a 3D model feature descriptor describing volumetric information, PTIs exhibit superior robustness when compared to SDF. We believe this is because the normal estimation, which is a key step in calculating the SDF, is easily affected by noise.

5.4. Performance

All experiments were conducted on a PC with Intel Dual-core 2.67 GHz CPU and 10 GB RAM. We implemented the algorithm using C++ and Matlab. Table 2 shows the average running time of each step for one single 3D model from the SHREC'11 non-rigid 3D models dataset, where each 3D model contains about 10000 vertices and 19000 faces.

5.5. Extension to 3D model retrieval

Although we claim that the PTI descriptor proposed in this paper is suitable for non-rigid 3D model classification, it can also be used in 3D model retrieval. The key idea is to mix the three feature vectors of each 3D model into a new feature vector, where the weight scheme is the same with a , b , c mentioned in Eq. (2). Using the resultant feature vectors, we can get a similarity matrix to encode the L_1 distance between the feature descriptors of any pair of models in the dataset. Based on the similarity matrix, there are 5 commonly used measures to evaluate retrieval performance:

Table 3

Retrieval performance statistics on SHREC'11 non-rigid 3D models dataset using various measures. Higher scores mean better performances.

Algorithm	PTI	GC	SDF	SIHKS	WKS
NN	96.67%	90.83%	96.67%	100%	99.83%
Tier 1	71.03%	61.69%	68.40%	97.52%	97.35%
Tier 2	80.14%	74.37%	80.29%	99.51%	98.85%
E-measure	59.58%	54.38%	58.75%	75.94%	75.60%
DCG	91.07%	87.48%	91.33%	99.69%	99.49%

- Nearest neighbor (NN): the percentage of the closest matches belonging to the same class as the query.
- First-tier and second-tier: the percentage of models in the query's class that appear within the top $K - 1$ and $2(K - 1)$ matches respectively, where K is the size of the query's class.
- E-measure: a composite measure of the precision and recall for a fixed number (32) of retrieved results.
- Discounted Cumulative Gain (DCG): a statistic that weighs correct results near the front of the ranked list more than correct results toward the end of the list.

The results are shown in Table 3. These statistics show the retrieval performance of our algorithm.

6. Conclusion

We proposed a novel descriptor called PTI. By encoding the thickness information of input 3D models, the PTI is not only invariant to the basic geometric transformation, such as translation, rotation, and scaling, but also insensitive to quasi-isometric deformations, e.g., human body poses. Benefitting from partial volumetric information, the PTI is robust to small geometric noise added to the vertices of 3D models. To apply the PTI to 3D model classification, we combined the PTI with KSRC and presented a new 3D model classification algorithm. Extensive experimental results demonstrate the high accuracy, robustness, and effectiveness of our algorithm.

However, we found that the MDS technique generally introduces shape distortion especially at extremity points although it basically eliminates pose differences. We will study if it is possible to obtain a better MDS method without shape distortion.

Acknowledgments

We thank the anonymous reviewers for their valuable comments and suggestions. This work is supported by National Natural Science Foundation of China (11226328, 61222206, 11526212, 61300168, 61303144, 31302231), National Science Foundation CAREER Award (IIS-1350330), Natural Science Foundation of Zhejiang Province (LY13F020018), Opening Foundation of Zhejiang Provincial Top Key Discipline (XKXL1406), Natural Science Foundation of Ningbo City Grant (2015A610123), and One Hundred Talent Project of the Chinese Academy of Sciences (Ligang-Liu).

Appendix. Sparse representation-based classification (SRC) and kernel sparse representation-based classification (KSRC)

The basic idea of the SRC is that the membership of a testing sample should be predicted by which class training samples can best linearly represent the test sample. Given t class training samples $A = [A^{(1)}A^{(2)} \dots A^{(t)}]$, where $A^{(i)} = [a_1^{(i)}, a_2^{(i)}, \dots, a_{n_i}^{(i)}] \in R^{s \times n_i}$ represents the training samples belong to the i th class and s is the dimension of samples. Let $Y \in R^s$ be a testing sample, the following optimization problem is used by the SRC to calculate the

sparsest coefficients \hat{X} to represent the testing sample using the training samples:

$$\begin{cases} \hat{X} = \arg \min \|X\|_0 \\ \text{s.t. } \|AX - Y\|_2 \leq \varepsilon. \end{cases} \quad (\text{A.1})$$

Here, ε is a small threshold and $\|\cdot\|_0$ denotes the ℓ_0 -norm, which counts the number of nonzero entries in a vector. However, finding the solution of optimization problem (A.1) is NP-hard [40]. Therefore, the SRC tries to seek the solution of another optimization problem:

$$\begin{cases} \hat{X} = \arg \min \|X\|_1 \\ \text{s.t. } \|AX - Y\|_2 \leq \varepsilon, \end{cases} \quad (\text{A.2})$$

where $\|\cdot\|_1$ denotes the ℓ_1 -norm, which is the sum of all entries' absolute values in a vector. Some researches [41–43] show that the solutions of the problem (A.1) and (A.2) are equal if the solution \hat{X} is sparse enough. Since the problem (A.2) can be solved efficiently by standard linear programming methods, the SRC obtains the sparsest representation $Y = A\hat{X}$ for the test sample Y by solving the problem (A.2).

After obtaining the sparsest representation $Y = A\hat{X}$, one can determine the classification of the test sample Y easily by inspecting the nonzero entries of the solution \hat{X} . Let the entries of \hat{X} be denoted by $[\hat{X}_{1,1}, \hat{X}_{1,2}, \dots, \hat{X}_{1,n_1}, \hat{X}_{2,1}, \hat{X}_{2,2}, \dots, \hat{X}_{2,n_2}, \dots, \hat{X}_{n,1}, \hat{X}_{n,2}, \dots, \hat{X}_{n,n_t}]^T$, the SRC predicts the membership m of Y by finding the minimal residual between Y and $A\hat{X}_i$ ($i = 1, 2, \dots, n$), where $\hat{X}_i = [0, \dots, 0, \hat{X}_{i,1}, \hat{X}_{i,2}, \dots, \hat{X}_{i,n_i}, 0, \dots, 0]$. That is, the SRC classifies the test sample Y by minimizing the following residuals:

$$m = \arg \min_i r_i = \arg \min_i \|Y - A\hat{X}_i\|_2. \quad (\text{A.3})$$

The SRC regards the result of optimization problem (A.3) as the membership m of Y .

Although the SRC algorithm is very successful in human face recognition, it is not always successful in our case. The SRC relies on the assumption that the feature vectors belonging to the same class belong, at least approximately, to the same linear subspace, while feature vectors belonging to different classes correspond to different linear subspaces. Therefore, the SRC will be confused if the feature vectors belonging to different classes are co-linear. The KSRC has been proposed to remedy this drawback of SRC. The key idea of KSRC is to map the samples to a higher dimensional space where the classification is performed instead. The KSRC adopts a kernelized version of the optimization problem (A.2) used in SRC, which can be formulated as follows:

$$\begin{cases} \hat{X} = \arg \min \|X\|_1 \\ \text{s.t. } \|B^T BX - B^T \phi(Y)\|_2 \leq \varepsilon, \end{cases} \quad (\text{A.4})$$

where $B = \phi(A)$ and $\phi(\cdot)$ is a function mapping the samples to a higher dimensional space. Denote $k(x, z) = \langle \phi(x), \phi(z) \rangle$ as the inner product of $\phi(x)$ and $\phi(z)$, then

$$\begin{aligned} B^T B &= [\phi(A^{(1)}), \phi(A^{(2)}), \dots, \phi(A^{(t)})]^T \\ &\quad \times [\phi(A^{(1)}), \phi(A^{(2)}), \dots, \phi(A^{(t)})] \\ &= \begin{pmatrix} k(A^{(1)}, A^{(1)}) & \dots & k(A^{(1)}, A^{(t)}) \\ k(A^{(2)}, A^{(1)}) & \dots & k(A^{(2)}, A^{(t)}) \\ \dots & \dots & \dots \\ k(A^{(t)}, A^{(1)}) & \dots & k(A^{(t)}, A^{(t)}) \end{pmatrix} \end{aligned} \quad (\text{A.5})$$

and

$$\begin{aligned} B^T \phi(Y) &= [\phi(A^{(1)}), \dots, \phi(A^{(t)})]^T \phi(Y) \\ &= [k(A^{(1)}, Y), \dots, k(A^{(t)}, Y)]^T. \end{aligned} \quad (\text{A.6})$$

The predicted membership m_Y of the testing sample Y can be obtained by minimizing the residual:

$$m_Y = \arg \min_i r_i(Y) \\ = \arg \min_i \left(\left\| B^T \phi(Y) - B^T B \hat{X}_i \right\|_2 \right). \quad (\text{A.7})$$

Note that $\phi(\cdot)$ is not necessary to be defined explicitly because the two terms $B^T B$ and $B^T \phi(Y)$, required for computing the residual, rely on the kernel function $k(\cdot, \cdot)$ only. That is, we need to find a suitable $k(\cdot, \cdot)$ to support KSRC. Experimental results show that the Gaussian kernel function $k(x, z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$ is a good choice for our testing data. We experimentally set $\sigma = 1$ and $\varepsilon = 0.01$.

References

- [1] Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3D shape recognition. In: Proceedings of the IEEE international conference on computer vision, 2015, pp. 945–53.
- [2] Ke L, Ning H, Jian X, Jiyang D, Ling S. Learning view-model joint relevance for 3D object retrieval. *IEEE Trans Image Process* 2015;24(5):1449–59.
- [3] Zhang L, Zhou W-D, Chang P-C, Liu J, Yan Z, Wang T, Li F-Z. Kernel sparse representation-based classifier. *IEEE Trans Signal Process* 2012;60(4):1684–95.
- [4] Osada R, Funkhouser T, Chazelle B, Dobkin D. Shape distributions. *ACM Trans Graph* 2002;21(4):807–32.
- [5] 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. Aachen, (Germany): Eurographics Association; 2003. p. 156–64.
- [6] Chen D-Y, Tian X-P, Shen Y-T, Ouhyoung M. On visual similarity based 3D model retrieval. *Comput Graph Forum* 2003;22(3):223–32.
- [7] Jain V, Zhang H. A spectral approach to shape-based retrieval of articulated 3D models. *Comput-Aided Des* 2007;39(5):398–407.
- [8] Vranic DV. An improvement of rotation invariant 3D-shape based on functions on concentric spheres. In: IEEE international conference on image processing, Vol. 3, 2003, pp. III–757.
- [9] Kazhdan M, Funkhouser T, Rusinkiewicz S. Symmetry descriptors and 3D shape matching. In: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on geometry processing. ACM; 2004. p. 115–23.
- [10] Ricard J, Coeurjolly D, Baskurt A. Generalizations of angular radial transform for 2D and 3D shape retrieval. *Pattern Recognit Lett* 2005;26(14):2174–86.
- [11] Knopp J, Prasad M, Van Gool L. Orientation invariant 3D object classification using hough transform based methods. In: Proceedings of the ACM workshop on 3D object retrieval, 2010, pp. 15–20.
- [12] Bronstein MM, Kokkinos I. Scale-invariant heat kernel signatures for non-rigid shape recognition. In: IEEE conference on computer vision and pattern recognition, 2010, pp. 1704–1711.
- [13] Smeets D, Keustermans J, Vandermeulen D, Suetens P. meshshift: Local surface features for 3D face recognition under expression variations and partial data. *Comput Vis Image Underst* 2013;117(2):158–69.
- [14] Lu K, Wang Q, Xue J, Pan W. 3D model retrieval and classification by semi-supervised learning with content-based similarity. *Inform Sci* 2014;281:703–13.
- [15] Loffler J. Content-based retrieval of 3D models in distributed web databases by visual shape information. In: IEEE international conference on information visualization, 2000, pp. 82–87.
- [16] Li J-B, Sun W-H, Wang Y-H, Tang L-L. 3D model classification based on nonparametric discriminant analysis with kernels. *Neural Comput Appl* 2013; 22(3–4):771–81.
- [17] Atmosukarto I, Shapiro IG. 3D object retrieval using salient views. *Int J Multimed Inf Retr* 2013;2(2):103–15.
- [18] Yue G, Meng W, Dacheng T, Rongrong J, Qionghai D. 3-D object retrieval and recognition with hypergraph analysis. *IEEE Trans Image Process* 2012;21(9): 4290–303.
- [19] Ke L, Ji R, Tang J, Yue G. Learning-based bipartite graph matching for view-based 3D model retrieval. *IEEE Trans Image Process* 2014;23(10):4553–63.
- [20] Gao Y, Wang M, Ji R, Wu X, Dai Q. 3-D object retrieval with hausdorff distance learning. *IEEE Trans Ind Electron* 2014;61(4):2088–98.
- [21] Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1(1):81–106.
- [22] Sebe N, Lew MS, Cohen I, Garg A, Huang TS. Emotion recognition using a Cauchy naive Bayes classifier. In: IEEE 16th international conference on pattern recognition, Vol. 1, 2002, pp. 17–20.
- [23] Liu Y, You Z, Cao L. A novel and quick SVM-based multi-class classifier. *Pattern Recognit* 2006;39(11):2258–64.
- [24] Bishop CM. *Neural networks for pattern recognition*. Oxford university press; 1995.
- [25] Leng B, Zhang X, Yao M, Xiong Z. 3D object classification using deep belief networks. In: *MultiMedia modeling*. Springer; 2014. p. 128–39.
- [26] Liu Z, Zhang F, Bu S. Spectral classification of 3D articulated shapes. In: *MultiMedia modeling*. Springer; 2014. p. 315–22.
- [27] Nian R, He B, Lendasse A. 3D object recognition based on a geometrical topology model and extreme learning machine. *Neural Comput Appl* 2013; 22(3–4):427–33.
- [28] Au OK-C, Tai C-L, Cohen-Or D, Zheng Y, Fu H. Electors voting for fast automatic shape correspondence. *Comput Graph Forum* 2010;29:645–54.
- [29] Brubaker MA, Punjani A, Fleet DJ. Building proteins in a day: Efficient 3D molecular reconstruction. In: Proceedings of IEEE conference on computer vision and pattern recognition, 2015.
- [30] Cox TF, Cox MA. *Multidimensional scaling*. CRC Press; 2010.
- [31] Xin S-Q, Wang G-J. Improving Chen and Han's algorithm on the discrete geodesic problem. *ACM Trans Graph* 2009;28(4):1–8.
- [32] Garland M, Heckbert PS. Surface simplification using quadric error metrics. In: Proceedings of the 24th annual conference on computer graphics and interactive techniques. SIGGRAPH '97, New York, (USA): ACM Press, Addison-Wesley Publishing Co.; 1997. p. 209–16.
- [33] Jolliffe I. *Principal component analysis*. Wiley Online Library; 2005.
- [34] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, 2005, Vol. 1, 2005, pp. 886–93.
- [35] Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 2009;31(2):210–27.
- [36] Lian Z, Godil A, Bustos B, Daoudi M, Hermans J, Kawamura S, Kurita Y, Lavoué G, Nguyen H, Ohbuchi R. Shrec'11 track: shape retrieval on non-rigid 3D watertight meshes. In: Eurographics workshop on 3D object retrieval, Vol. 11, 2011, pp. 79–88.
- [37] Nene S, Nayar SK. A simple algorithm for nearest neighbor search in high dimensions. *IEEE Trans Pattern Anal Mach Intell* 1997;19(9):989–1003.
- [38] Shapira L, Shamir A, Cohen-Or D. Consistent mesh partitioning and skeletonisation using the shape diameter function. *Vis Comput* 2008;24(4):249–59.
- [39] Pickup D, Sun X, Rosin PL, Martin RR. Euclidean-distance-based canonical forms for non-rigid 3D shape retrieval. *Pattern Recognit* 2015;48(8):2500–12.
- [40] Amaldi E, Kann V. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoret Comput Sci* 1998;209(1): 237–60.
- [41] Donoho DL. For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution. *Comm Pure Appl Math* 2006;59(6):797–829.
- [42] Candes EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Comm Pure Appl Math* 2006;59(8):1207–23.
- [43] Candes EJ, Tao T. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans Inf Theory* 2006;52(12):5406–25.