

# Locality-constrained sparse patch coding for 3D shape retrieval



Zhenbao Liu, Shuhui Bu\*, Junwei Han

Northwestern Polytechnical University, Post Box 727, Youyi West Road 127, Xi'an 710072, China

## ARTICLE INFO

### Article history:

Received 16 November 2013

Received in revised form

22 March 2014

Accepted 1 June 2014

Available online 4 November 2014

### Keywords:

3D shape retrieval

Patch

Sparse coding

Locality constraint

## ABSTRACT

3D shape retrieval is a fundamental task in many domains such as multimedia, graphics, CAD, and amusement. In this paper, we propose a 3D object retrieval approach by effectively utilizing low-level patches of 3D shapes, which are similar as superpixels in images. These patches are first obtained by means of stably over-segmenting 3D shape, and then we adopt five representative geometric features including shape diameter function, average geodesic distance, and heat kernel signature, to characterize these low-level patches. A large number of patches collected from shapes in a dataset are encoded into patch words by virtue of locality-constrained sparse coding under the consideration of local smooth sparsity. Input query is compared with 3D models in the dataset through probability distribution of patch words. Experiments reveal that the proposed method achieves comparable retrieval performance to state-of-the-art methods.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

3D model as an important media contains rich 3D information preserving real object surface, color, and texture, which has been extensively applied in the domain of multimedia, graphics, virtual reality, amusement, design, and manufacturing. A huge number of publicly usable models such as in Google 3D Warehouse has been widely distributed and quickly spread, and many researchers attempt to provide content based retrieval techniques, e.g., sketch based 3D model retrieval [1], range image based retrieval [2], example shape based retrieval [3], and partial shape based retrieval [4], for accurately searching desirable objects and reusing these models.

A variety of 3D shape retrieval algorithms have been proposed, where early research on retrieval methods [5] focused mainly on global descriptors and their invariance under global Euclidean transformations. Recently significant effort has been invested on 3D interest point detection [6–8], local point description and organization [9], topological structure [10], non-rigid shape feature [11], and appearance analysis [12,13].

In this paper, we propose a 3D object retrieval approach by effectively utilizing low-level patches of 3D shapes consistent with geometric criterion, which are analogous to superpixels in images. In the novel framework, 3D shape is first over-segmented into many low-level patches, and different types of geometric features are extracted from these patches. Then we encode a large number of patches collected over a 3D model dataset via locality-

constrained sparse coding, and extract compact and representative patch words. Input query is compared with 3D models in the database by probability distribution of patch words. Several groups of experimental results show that this method improves the retrieval performance of state-of-the-art methods.

The main contributions of the paper are described as follows.

1. Compared with point descriptors based retrieval, we introduce low-level patches to represent a 3D object, and each object only requires a small number of patches to discriminate from irrelevant objects. Moreover, these patches are not randomly generated but according to geometric criterion.
2. Different from retrieval methods based on meaningful segments and graph structure, our method avoids directly generating a few meaningful parts, which possibly leads to mis-segmented parts because the techniques of semantic segmentation are not mature. Moreover, we do not adopt graph structure to organize these parts because many topologically variable objects exist, for example, vases with different number of handles. Patch based representation will make retrieval robust against topology variation.
3. The motivation of proposing sparse coding to represent 3D objects is based on our observation, that not only the same category of objects but also irrelevant objects have many visually similar patches. For example, human body has many similar patches as horse body. Therefore, we extract sparse patch words to approximate these shapes and model occurrence frequency of these common features using locality-constrained and relatively sparser coefficients than previous bag-of-features retrieval algorithms.

\* Corresponding author. Tel.: +86 29 88492344.  
E-mail address: [bushuhui@nwpu.edu.cn](mailto:bushuhui@nwpu.edu.cn) (S. Bu).

The rest of this paper is organized as follows. We first discuss related works in Sections 2 and 3 present an overview of the proposed method. Section 4 introduces how to generate low-level patches and a patch segmentation algorithm. Section 5 describes each patch via different types of 3D local descriptors. The process of encoding these patches via locality-constrained sparse coding is presented in detail in Section 6. Experimental results of 3D shape retrieval are reported on two representative 3D shape data sets in Section 7, and we conclude this work in Section 8.

## 2. Related works

In the past decade, 3D model retrieval has become an active research topic in various fields such as multimedia, computer graphics, computer vision, and computer aided design. In contrast with early works [5] which take into account only the whole shape characteristics of 3D model, recent works mainly focus on local point description and organization, topological structure, non-rigid shape feature, and appearance analysis. Local point descriptor can encode rich local context while keeping rotation or bending invariant. Local descriptor based retrieval methods are commonly composed of three steps. (1) Detecting salient points (or local regions). This step can be omitted if all the points are uniformly sampled on the surface of 3D model. (2) Describing each salient point (or local region) with one feature vector. If this step is removed, the retrieval problem is converted into shape registration, which can be solved by Iterative Closet Points (ICP) and its invariants. (3) Comparing two sets of point (region) features. A simple way is to sum the distance of each pair of nearest neighbors in feature space. It is also feasible to compare histograms of these original features, or match template features of a training set. Representative local descriptors applied in 3D shape retrieval include global point signature [14], Laplace–Beltrami operator defined on local manifold [15], heat kernel signature [16] and scale invariant heat kernel signature [17], 3D SURF [18], 3D Harris [19], 3D SIFT [20], inner distances [21], and 3D intrinsic shape context [22]. To avoid high computational burden from combinatorial comparison between two sets of dense points, and efficiently organize these descriptors, bag of features has been borrowed from text and image processing to address correspondence and matching of 3D local descriptors [23,24].

Retrieval based on topological structure assumes that a 3D model is represented by means of a topologically connected graph consisting of nodes and edges. The problem of 3D model comparison can be converted into low-dimensional graph matching after extracting topological structures such as medial surface, curve skeleton, Reeb graph, and model graph. These simplified graphs have similar topology structures as original 3D models, and comparison between 3D models is able to be realized by virtue of checking isomorphism of simplified graphs. For example, there exist several types of graph isomorphism algorithms including tree search based algorithms, decision tree based techniques, and spectral methods. These algorithms can perform inexact computation with matching cost to measure the similarity of two simplified graphs. This type of strategies has been extensively investigated in 3D shape retrieval, for example, common undirected graph adopted in [25], Reeb graph used in [26,10] and extended Reeb graph [27], bipartite graph used in [28], skeleton graph adopted in [29] and binary tree used in [30]. Each explicitly or implicitly segmented meaningful part is identified by a single node, and edges in the graph represent adjacency relations between these segments. Therefore, shape retrieval is easily achieved by resorting to checking graph isomorphism and measuring similarity of simplified graphs of two shapes.

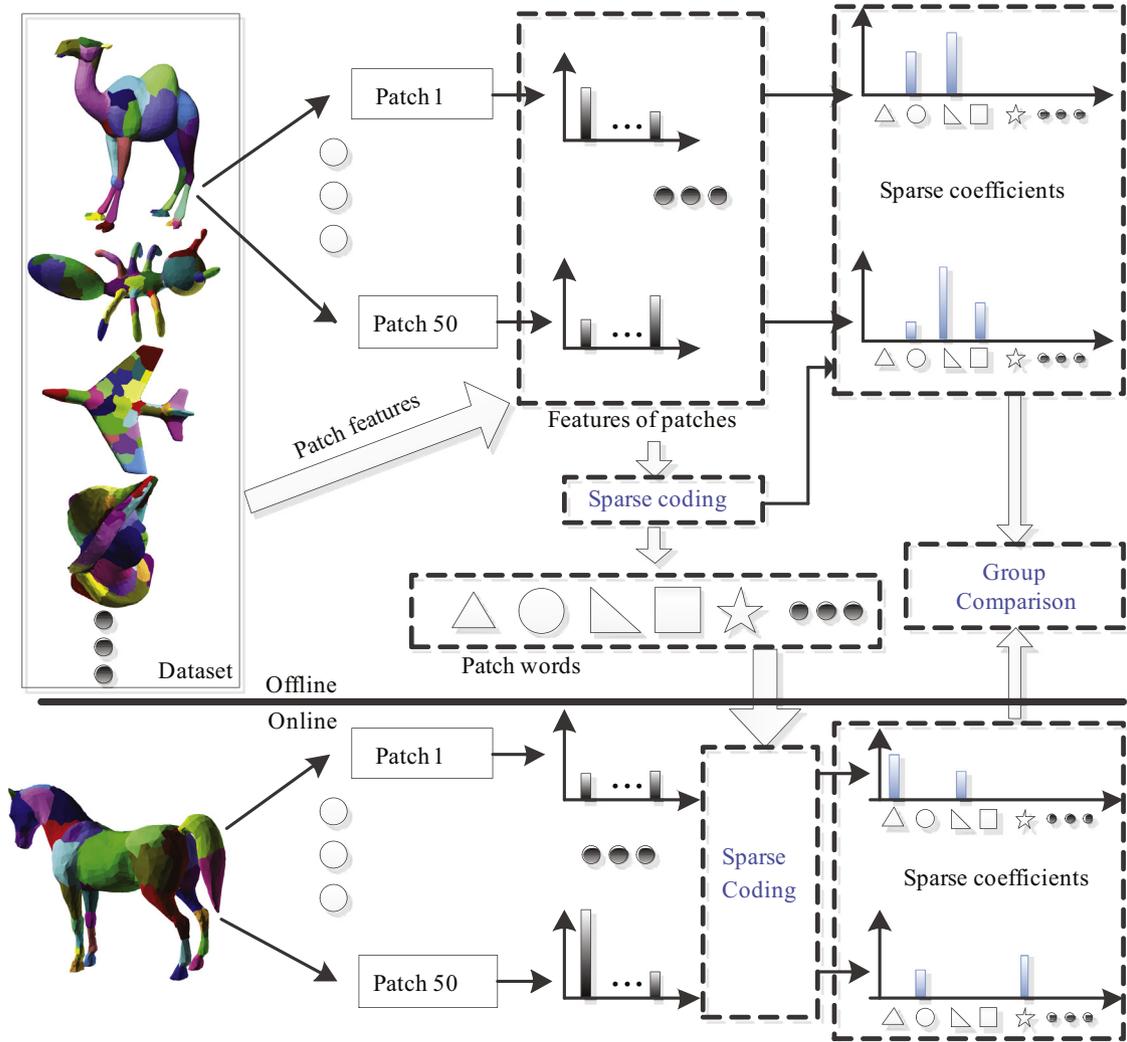
Many works try to solve the difficulty of non-rigid 3D shape retrieval [11] by means of utilizing various isometry invariant attributes. Geodesic distance measures the intrinsic distance between two arbitrary points on 3D surface, and contains rich geometric information. For example, geodesic distance commonly keeps unchanged under isometric deformations, which can assist in handling non-rigid shape deformation. A representative work is spectral method based on spectral decomposition of geodesic distances [31]. It filters geodesic distances appropriately to remove the effect of scaling and then compute a low-dimensional spectral embedding of 3D shape to obtain invariance to bending and rigid-body transformations. Spectral decomposition of an affinity matrix between geodesic distances characterizes a whole 3D shape, and its real eigenvalues are adopted to compare with other shapes. Diffusion distance based descriptor [32] inherits the isometry invariant attribute of geodesic distance, and further introduces the average of all the paths of fixed steps connecting two points on the surface, which is more robust than single geodesic distance. The diffusion distance is seen as average probability of traveling between two arbitrary points. Another isometry invariant descriptor is based on Laplace–Beltrami spectrum of 3D surface [33]. The spectrum is represented with eigenvalues of Laplace–Beltrami operator, and independent of different parametrization and spatial position of 3D shape. Additionally, the eigenvalues can be normalized so as to indirectly handle different scales of 3D shapes.

Appearance based 3D object retrieval tends to address how to effectively generate, organize, and compare many views of a 3D object. For example, several works focus on selecting query views [34], weighted bipartite graph matching of views [35], camera constrained-free view generation [12], constructing multiple hypergraphs of views [13], and panoramic views [36]. An interesting view based work employs interactive learning mechanism [37], which establishes a mapping from feature points in low-level feature space to points in high-level semantic space. The mechanism receives long-term relevance feedback from users via recorded retrieval history, and captures users' semantic information to refine retrieval results.

Different from previous works, in this paper we introduce the concept of low-level patches to represent a 3D object. This way avoids semantic segmentation used in retrieval methods, which is heavily dependent on topological structure such as skeleton, and unstable in the case of topological change. Moreover, the technique of locality-constrained sparse coding avails to extract high-level patch words from a large set of 3D shapes with many similar patches, which will be more compact and representative than low-level patches.

## 3. Overview of the proposed method

The overview of the proposed method is shown in Fig. 1. Each 3D shape is first over-segmented into a number of low-level geometric patches, and these patches are described with different types of geometric descriptors, which are adopted to characterize different geometric attributes including local, global, and topological features. After extracting these features for each patch collected from all 3D shapes in a large data set, locality-constrained sparse coding is adopted to construct a set of bases also known as visual words in a vocabulary in the domain of computer vision. Each patch is encoded by means of these bases, which are named as patch words in this paper. A number of patch words generated from a large set of 3D shapes compose a large vocabulary. Given a new object as query shape, the problem of representing it with high-level patch words is converted to optimize its coefficients via locality-constrained sparse coding.



**Fig. 1.** Overview of the proposed method based on sparse patch coding. The offline process of 3D shape retrieval is composed of three main steps, low-level patch generation, vocabulary construction, and high-level object representation. A given query is over-segmented into patches on line, and each of patches is encoded to sparse coefficients, which will be compared with coefficients stored in the database.

The input query compares with 3D models in the database by probability distribution of patch words.

#### 4. Low-level patch generation

The segmentation based retrieval methods directly segment 3D shapes into high-level semantic parts, for example, human is represented by several components, head, arms, legs, etc. A graph is used to organize these parts, and 3D model matching is simplified to a problem of graph isomorphism. However, this type of methods is sensitive to objects with different topology, for example, vases with different number of handles, because their graphs significantly vary. It results in that objects with similar semantics are easily mis-recognized into different categories. In this paper, to our best knowledge, it is the first time to introduce low-level patch into the domain of 3D shape retrieval. It is similar as the concept of superpixel in image processing, which has been successfully applied to recognition of complex scenes. However, we cannot resort to spatially uniform patch generation on 3D model like [38] because it loses low-level initial geometric information.

It is regretful, however, that up to now there are seldom segmentation methods aiming at generation of low-level patches

because most segmentation methods are designed to generate high-level meaningful parts. We attempt to introduce a classical and fully automatic segmentation method based on randomized cuts [39], which is considered as highly discriminative and robust segmentation in a recent evaluation work [40]. It generates a random set of segmentations and measures the frequencies that each edge of a mesh lies on a segmentation boundary in the randomized set. Therefore, it is less sensitive to surface noise, tessellation, pose, and intra-class shape variations, and these properties will help to accurately match shapes with same semantics but large geometrical variations and non-rigid deformations.

It first defines concave weights by means of a dihedral angle between two adjacent faces to form a total cut cost  $NCut_K$  as follows:

$$NCut_K = \sum_{i=1}^K \frac{w(\mathbf{S}_i, \mathbf{M} - \mathbf{S}_i)}{A(\mathbf{S}_i)}, \quad (1)$$

where  $\mathbf{S}_i$  is a segmented patch and  $\mathbf{M}$  is the whole 3D mesh.  $K$  is the number of segmented patches.  $A(\mathbf{S}_i)$  is the area of  $i$ th segmented patch, which is used to normalize the cost function to avoid dependence on mesh tessellation.  $w(\mathbf{S}_i, \mathbf{M} - \mathbf{S}_i)$  is the cut cost of a patch, defined by

$$w(\mathbf{S}_i, \mathbf{M} - \mathbf{S}_i) = \sum_e l(e) \min((\theta(e)/\pi)^\alpha, 1), \quad (2)$$

where  $e$  denotes each boundary edge of segmented patch, and  $l(e)$  is its length.  $\theta(e)$  is its dihedral angle, and here only concave boundaries are considered.  $\alpha$  is set to 10 in our experiments. And then, the difference in cut costs associated with merging each pair of smaller patches is mapped to  $[0,1]$ , and each value is raised to the power of a randomized parameter  $1/r$ .  $r$  is set to 0.02 according to the suggestion from [39]. Each pair of smaller patches is merged with the probability.

Nevertheless, we find it is sensitive to local surface fluctuation and results in extremely small patches. This problem is related to the fact that it is not robust by adopting the dihedral angle of an edge as its cut cost. To overcome this shortcoming, we adopt as its cut cost the mean of dihedral angles around each edge in a geodesic radius of 1% average geodesic distance on the 3D mesh. This performs well in practical over-segmentation. In order to provide sufficient low-level patches for next high-level processing, we set patch number to 50 for each object and generate the same number of patches for each category of objects, as illustrated in Fig. 2. It can be seen that each low-level patch has already contained rich geometric information and these small patches can sufficiently represent the whole characteristics of each object.

## 5. Patch description

After we obtain these low-level patches for each 3D model, another problem we need to resolve is how to effectively describe these patches because it is different from point features. Each patch is composed of large points, edges, and faces, and describing the patch requires enough consideration on its geometrical characteristics, topological features of connecting its neighbors, and also its relative position or function in the global shape. We tend to adopt conformal geometry signature to describe its geometrical features, for example, easily differentiating sharp or smooth

surface regions. Moreover, shape diameter function of each region is adopted to describe its relative thickness, an important geometrical attribute which can recognize the part that it belongs to. For topological features, we consider its connection relationship with local neighbor regions, and employ Laplace–Beltrami operator to describe its topological connection. In order to describe its global position or function, two descriptors, average geodesic distance and heat kernel signature, provide messages about the global relationship between this region and other regions. We first compute point descriptors and then generate patch features by estimating distributions of point descriptors. The detailed point descriptors are described as follows.

*Conformal geometry signature:* When a 3D manifold model is conformally transformed into plane, each conformal scaling factor is used to locally scale the neighborhood of a vertex in order to achieve the target curvature at the vertex [41]. High scaling factor corresponds with cone singularity so that its plane parameterization has low distortion. We use the scaling factor on each vertex as our point feature. The conformal map from a manifold mesh to a homeomorphous surface with constant Gaussian curvature is computed by solving the following sparse linear equations

$$L\phi = K^T - K^{orig}, \quad (3)$$

where,  $L$  denotes Cotangent Laplacian, and  $K^{orig}$  and  $K^T$  are Gaussian curvature at the vertex and the area-weighted mean of all the Gaussian curvatures on the mesh respectively. The solution vector  $\phi$  is composed of conformal factors of these vertices.

*Shape diameter function:* The shape diameter function (SDF) [42] is a volume-based scalar function measuring the diameters of different parts of a 3D shape. The SDF value is computed by sending 30 rays inside a small cone with angle of  $30^\circ$  to intersect with the opposite side of the boundary, and averaging these weighted ray lengths. The values remain similar on the neighborhood of the same part,



Fig. 2. Patch generation results for 3D objects from thirteen classes. Each object is partitioned into 50 patches, and we can see that these patches already have low-level geometric feature, and can be regarded as smallest meaningful patch.



Fig. 3. Shape diameter functions of points in a class of vases with different topological and geometrical variants. The red discriminates parts with larger diameters from thin parts. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

and oblivious to articulated deformation. The SDF distributions of several examples are shown in Fig. 3.

**Laplace–Beltrami descriptor:** This descriptor extends Laplace operator in Euclidean domain onto manifold to achieve the divergence of vertex gradients on a mesh. Laplace–Beltrami operator of each vertex is commonly discretized into sum of the distances with cotangent and area weights from this vertex to its one-ring neighbors. Local surface feature is extracted based on the eigen-decomposition of the Laplace–Beltrami operator defined on the local region [15]. The set of larger eigenvalues are chosen to form a feature vector for describing this point, which is isometry invariant and consequently able to tolerate non-rigid transformation.

**Average geodesic distance:** We compute approximate geodesic distances [43] on 3D mesh between each point and other points, and average these distances for the point. The average geodesic distance is used for our point feature. We compute this distance for each vertex of several examples, and visualize these vertices in Fig. 4.

**Scale invariant heat kernel signature:** Heat kernel signature is derived from a heat diffusion equation by using Laplace–Beltrami operator on surfaces. The fundamental solution of the heat equation is called the heat kernel. Specifically, the heat diffusion over vertices  $\mathbf{v}$  of a given manifold mesh  $M$  in time  $t$  is described by the following heat equation:

$$\Delta u(\mathbf{v}, t) = -\frac{\partial u(\mathbf{v}, t)}{\partial t}, \quad (4)$$

where  $\Delta$  is Laplace–Beltrami operator on the manifold mesh. Given an initial heat distribution, Dirac delta function  $\delta_{\mathbf{x}}(\mathbf{y}) = 0$ , heat kernel is understood as the amount of heat transferred from the source  $\mathbf{x}$  to the target  $\mathbf{y}$  in time  $t$  given a unit heat source at  $\mathbf{x}$ , and can be solved by eigen-decomposition of Laplace–Beltrami operator in the following equation:

$$h_t(\mathbf{x}, \mathbf{y}) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(\mathbf{x}) \phi_i(\mathbf{y}), \quad (5)$$

where  $\lambda_i$  and  $\phi_i$  are the  $i$ th eigenvalue and eigenfunction of Laplace–Beltrami operator respectively. Sun et al. [44] represent 3D shapes via the heat kernel signature  $h_t(\mathbf{x}, \mathbf{x})$  under a set of diffusion times  $\{t\}$ , and set a different diffusion time to obtain local neighborhood information at a easily controlled scale. Heat kernel signature has an intuitive physical meaning that when a quantity of heat is placed on the source point how much the heat is left while time  $t$  elapses. This descriptor is intrinsic, insensitive to isometric deformation, and robust against surface noises. Furthermore, in order to make it scale invariant, Fourier transform of heat kernel signatures is proposed in [17], and we tend to adopt this feature and then visualize point values of relevant objects in Fig. 5.

**Patch feature:** After we obtain five descriptors for each point on the low-level patch, the description of the patch is defined by constructing a histogram of all the point values for each type of descriptors. Five histograms are naturally concatenated into a

whole feature vector so as to form the final patch description. In this paper, we consider these five types of features are equally important for each patch and hence their weights are same.

## 6. High-level patch coding

Sparse coding [45] has attracted many researchers from the domain of image and vision to solve tasks of image analysis, e.g., image retrieval [46,47], classification [48], recognition [49], and segmentation [50]. The advantages of applying sparse coding into 3D shape retrieval are two-fold. The first one is that it can capture higher-level features via learning basis functions from unlabeled data, and these features contain more semantic information and are adaptable to complex recognition tasks on variable 3D shapes. Moreover, sparse coding can learn over-complete basis sets, which can more adequately represent objects than limited orthogonal basis and then capture a large number of pattern in the 3D shape dataset. In this paper we introduce the concept of sparse coding into 3D shape retrieval to improve the state-of-the-art performance.

### 6.1. Sparse coding

Given each patch feature point  $\mathbf{x} \in R^k$ , assume that it is sampled from the feature space composed of all the types of patches collected from different 3D shapes. Sparse coding aims to interpret the feature point with an over-complete set of  $n$  bases  $\{\mathbf{b}_1, \dots, \mathbf{b}_n\} \in R^k$ , which are linearly combined with a sparse weight vector  $\boldsymbol{\omega} \in R^n$  such that

$$\mathbf{x} \approx \sum_{j=1}^n \omega_j \mathbf{b}_j, \quad (6)$$

where  $n$  satisfies  $n > k$ , which means that the bases are over-complete. These bases are also known as a code dictionary  $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n] \in R^{k \times n}$ . The key of coding lies in how to discover these bases only dependent on unlabeled data, which are considered as a training set of  $m$  input vectors  $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ . The training process is realized in unsupervised way, which is converted into an optimization problem. These bases and their corresponding weight vectors  $\{\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_m\}$  are the solutions to the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{B}, \boldsymbol{\Omega}}{\text{minimize}} \quad \sum_{i=1}^m \|\mathbf{x}_i - \mathbf{B}\boldsymbol{\omega}_i\|_2^2 + \lambda \|\boldsymbol{\omega}_i\|_1, \\ & \text{s.t.} \quad \|\mathbf{b}_j\|_2 \leq 1. \end{aligned} \quad (7)$$

where  $\boldsymbol{\Omega} = \{\boldsymbol{\omega}_i\}$ . The above cost function is union of the reconstruction error approximating the input vector via linear combination of basis, and the sparsity penalty of basis. We use  $l_1$  penalty as the sparsity function, and the optimization problem is convex while holding two sets of parameters  $\mathbf{B}$  and  $\boldsymbol{\Omega}$  alternatively fixed. In order to avoid trivial solutions, the  $l_2$  norm of each basis  $\mathbf{b}_j$  is

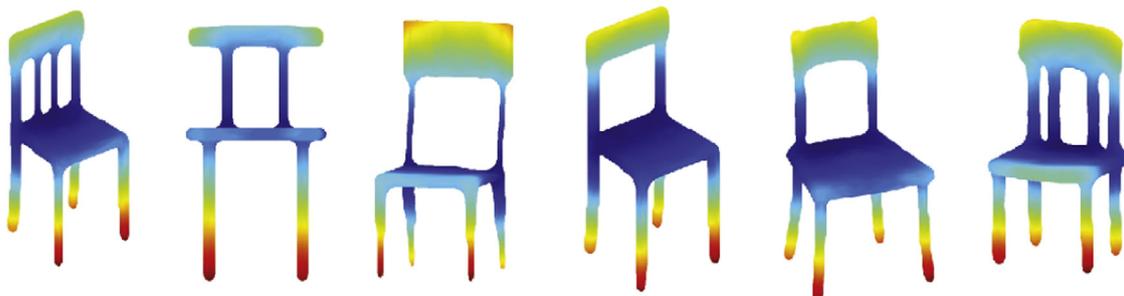
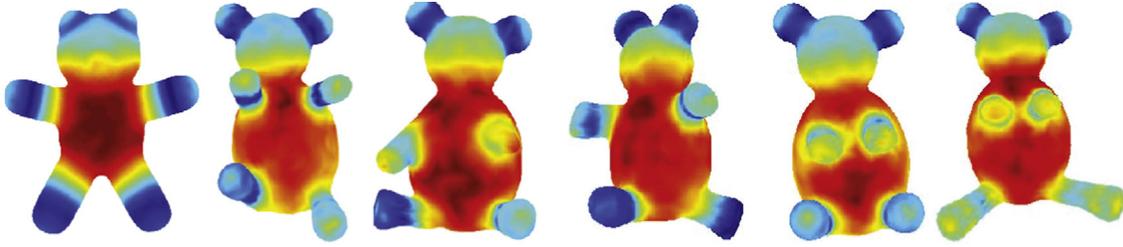


Fig. 4. Average geodesic distances of points on several chair meshes. The red shows higher values. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)



**Fig. 5.** Scale invariant heat kernel signature of points in a class of teddy bears with different poses. The red shows higher heat kernel values. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

limited to be less than or equal to 1. Although several approaches such as QCQP solver are competent for the problem, we adopt an efficient algorithm proposed in [51], where a feature-sign search strategy converges to global optimum while keeping the bases fixed and then bases are learned by the Lagrange dual given these unchangeable coefficients. We finally obtain a set of bases as our code dictionary, which contains diverse patch representatives of 3D shapes and covers most patch description.

### 6.2. Locality-constrained sparse coding

In order to obtain local smooth sparsity and make patch description more accurately represented by multiple bases, we further adopt a recent locality-constrained linear coding framework [52] to improve retrieval performance. Specifically, locality constraint instead of the sparsity constraint in Eq. (7) is integrated into the objective function using the following criteria:

$$\begin{aligned} & \underset{\mathbf{B}, \Omega}{\text{minimize}} \quad \sum_{i=1}^m \|\mathbf{x}_i - \mathbf{B}\boldsymbol{\omega}_i\|_2^2 + \lambda \|\mathbf{d}_i \odot \boldsymbol{\omega}_i\|_1, \\ & \text{s.t.} \quad \mathbf{1}^T \boldsymbol{\omega}_i = 1, \end{aligned} \quad (8)$$

where  $\odot$  denotes the element-wise multiplication. The constraint makes coding shift-invariant. The distance  $\mathbf{d}_i$  adjusts the weights of different bases by setting them proportional to the similarity of each basis to the patch descriptor  $\mathbf{x}_i$ . It is defined as follows:

$$\mathbf{d}_i = \exp\left\{\frac{\text{dist}(\mathbf{x}_i, \mathbf{B})}{\sigma}\right\}, \quad (9)$$

where  $\text{dist}(\mathbf{x}_i, \mathbf{B}) = [\text{dist}(\mathbf{x}_i, \mathbf{b}_1), \dots, \text{dist}(\mathbf{x}_i, \mathbf{b}_n)]^T$ , and each entry denotes the normalized Euclidean distance between  $\mathbf{x}_i$  and  $\mathbf{b}_j$  in the feature space.  $\sigma$  decides the weight decay speed and we empirically set it to 100 in our implementation. It should be noted that the essence of local constraint is feature selection because these distances make local bases selected for each patch description. This results in that the coding of each patch becomes relatively sparser and has the potential to reduce store requirements.

Another reason we choose the locality-constrained sparse coding is that the solution to the objective function in Eq. (8) can be accurately derived, which avoids the possibility of obtaining local optimum via iterative means. The analytical solution is expressed with the following equation:

$$\begin{aligned} \tilde{\boldsymbol{\omega}}_i &= (\mathbf{C}_i + \lambda \text{diag}(\mathbf{d}_i)) \setminus \mathbf{1}, \\ \boldsymbol{\omega}_i &= \tilde{\boldsymbol{\omega}}_i / \mathbf{1}^T \tilde{\boldsymbol{\omega}}_i, \end{aligned} \quad (10)$$

where  $\tilde{\boldsymbol{\omega}}_i$  is an intermediate solution, and  $\text{diag}(\mathbf{d}_i)$  represents a diagonal matrix in which each diagonal entry is the distance to the corresponding basis.  $\mathbf{C}_i$  denotes covariance matrix between each patch description  $\mathbf{x}_i$  and bases  $\mathbf{B}$ . It is given by

$$\mathbf{C}_i = (\mathbf{B} - \mathbf{1}\mathbf{x}_i^T)(\mathbf{B} - \mathbf{1}\mathbf{x}_i^T)^T. \quad (11)$$

### 6.3. Algorithm

There are three main steps in the retrieval algorithm based on locality-constrained sparse patch coding while comparing 3D models, which are illustrated in Fig. 1.

*Low-level patch generation:* Each object is partitioned into 50 different patches, and each low-level patch is characterized via utilizing histograms of five types of descriptors, that is, conformal geometry signature, shape diameter function, Laplace–Beltrami descriptor, average geodesic distance, and scale invariant heat kernel signature.

*Vocabulary construction:* After extracting features for each patch from all the models in a large data set, a set of bases also known as visual words in a vocabulary in the domain of computer vision is learned from large numbers of patch features. Here, we name these words as patch words so as to differentiate them from visual words in computer vision. For example, one of the data sets we adopt is composed of 400 3D models subdivided to 20 classes. Since each model is described with 50 patch features, 20 K patch features are obtained in the whole data set, which are used to construct patch words in the vocabulary. We solve the optimization mentioned above to get an optimum dictionary, which contains a set of patch words. We also studied the influence of the number of patch words while constructing vocabulary, and the size is set to 8, 16, 32, 64, 128, 256, respectively. The overall performance improved with the increase of the vocabulary size from 8, however, if the size exceeds 128, retrieval error rate has turned to rise and furthermore the run time becomes longer because of optimization of a large number of parameters.

*High-level object representation:* Given a new object as the input shape, the problem of representing it with high-level patch words is converted to optimize its coefficients by means of locality-constrained sparse coding. In patch level, the coefficients which are linearly combined with patch words are considered as a distribution of occurrences of these words. Dissimilarity between a pair of 3D objects  $P$  and  $Q$  is defined by comparing two groups of patch coefficients in the following equation.

$$\text{Dis}(P, Q) = \sum_{i=1}^K \|\boldsymbol{\omega}_i^P - \boldsymbol{\omega}_i^Q\| \quad (12)$$

where  $K$  is the patch number, and  $\boldsymbol{\omega}_i^Q$  identifies the closet match of the sparse coefficient  $\boldsymbol{\omega}_i^P$ . The distance between two sparse coefficients can be  $l_1$  norm, Kullback–Leibler Divergence (KLD), and Earth Mover's Distance. In this paper we adopt KLD to measure the difference among the distributions of coefficients by considering relative entropy.

## 7. Experiments

In this section we adopt two common 3D shape datasets and evaluation criteria to investigate the proposed retrieval algorithm,

Locality-constrained Sparse Patch Coding (LSPC), and also compare its retrieval performance to that of representative methods.

### 7.1. Evaluation criteria

To evaluate the retrieval performance of the proposed method, we adopt recall precision values, two fundamental measures. Recall is the ratio of the number of retrieved relevant objects to the total number of relevant objects in the database. Precision is the ratio of the number of retrieved relevant objects to the number of returned objects. Each object is selected as query, and compared against all the objects in the database. A retrieval list with length equal to database size is returned. For each query, the number of relevant objects in the retrieval list is same as the size of its class. The desired retrieval result is that all the relevant objects lie in the front of the list. The recall and precision values are finally averaged on each class of objects, and then the whole data set.

### 7.2. Retrieval performance on SHREC 2007 dataset

A common data collection, composed of SHREC 2007 watertight models [53], is adopted to test the retrieval performance of the proposed method. This collection is made up of 400 watertight mesh models, subdivided into 20 classes, each of which contains 20 objects with different geometrical variations and also articulated deformations. The data set contains not only natural objects, for example, human, ant, octopus, teddy, four leg animals, but also man-made objects such as cup, glasses, chair, plier, and bearing, vase. Shapes in each class contain sufficient and diverse variation from pose change to shape variability in the same semantic group. Moreover, some shapes with redundant parts exist, for example, sculpture head model with extra base, and there are also shapes with different geometric genus, in another example, vases with different number of handles. It is considered as a challenging data set. We investigate the retrieval performance of the proposed method on this dataset, and compare LSPC against the following state-of-the-art methods. These methods have also been evaluated on the same data set.

- *Augmented multi-resolution Reeb graph (AMRG)* [54,55]: They define Reeb graph to describe a contour relationship mapping vertices of a 3D shape to a geodesic space. Each contour level is represented as an edge of the Reeb graph, and each region between contour levels in a regular interval is coded into a node of Reeb graph, which is seen as a semantically segmented surface. The multi-resolution graph structure is then formed by hierarchically linking these nodes of connected regions. Topological, geometrical and visual information is attached to each graph node for enhancing graph matching and model comparison.
- *Spherical trace transform (STT)* [56]: This method first scales and places a 3D model into an unit sphere, and defines a set of planes tangential to several concentric spheres. Each plane intersects with the object, and intersection area is analyzed via 2D Krawtchouk moments, 2D Zernike Moments, and Polar Fourier Transform. Spherical Fourier transform is applied on intersection functions in order to generate rotation invariant descriptors.
- *Depth line encoding (DLE)* [57]: In their method, each 3D model is represented by a set of depth lines, extracted from depth buffer images projected onto the six faces of its bounding box after being normalized and scaled. Each depth image is considered as a collection of horizontal and vertical depth lines, each of which is encoded in a set of sequence states. 3D models in a database are compared to a query object and retrieved via

introducing a dynamic programming distance between their depth line descriptors.

Numerical values for the averaged recall and precision on all the models in the dataset are reported in Fig. 6. The average recall precision curves of some state-of-the-art methods, including AMRG, STT, and DLE, are plotted in the same figure as our reference. These results show that our method is comparable to the representative methods evaluated on the same dataset. LSPC obviously outperforms STT and DLE, and its retrieval precision is slightly better than that of AMRG. Although they are very close in the retrieval performance, we know that AMRG depends on the topological connections of different parts so that it is susceptible to topological variations in the same semantic group. Differently, LSPC directly decomposes a 3D mesh into smaller patches, and effectively avoids the influence from the change of part connections.

### 7.3. Retrieval performance on McGill 3D shape dataset

In this subsection, we report the results of our retrieval experiments conducted on the McGill 3D shape dataset [58]. This whole dataset is composed of 455 models and the exemplars span 19 basic level object classes such as human, hands, teddy bears, ants, octopuses, crabs, and so on. These classes are divided into two categories, 255 natural models with significant part articulation in 10 classes, and remaining man-made objects without clear part bending. A distinct characteristic of McGill 3D shape dataset lies in that it particularly contains a large number of articulated 3D models, which is commonly adopted to observe whether retrieval algorithms handle non-rigid deformations. We test our algorithm

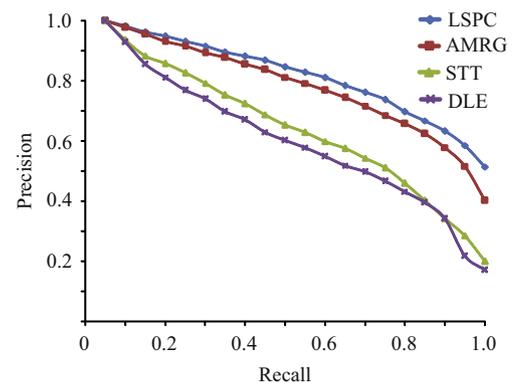


Fig. 6. Recall–precision curves of some state-of-the-art methods and the proposed method on SHREC 2007 watertight dataset.

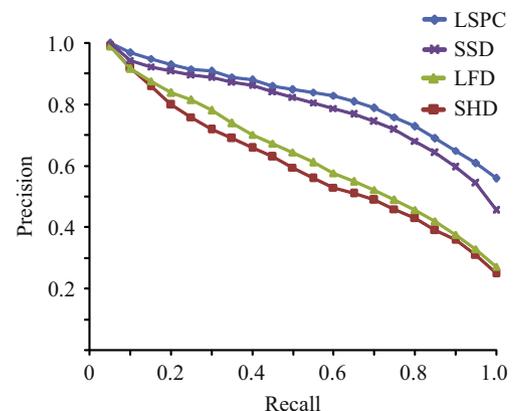


Fig. 7. Recall–precision curves of some state-of-the-art methods and the proposed method on McGill 3D shape dataset.

on these articulated models of McGill 3D shape dataset, and compare it with the following state-of-the-art methods.

- **Spherical Harmonics Descriptor (SHD)** [59]: Global spherical harmonics descriptor is constructed on several concentric spheres centering at the mass center of the analyzed 3D shape. These spheres with different radii intersect with mesh surface, and a binary identified function is used to evaluate intersection and non-intersection relationship between each sphere and mesh surface. The spherical function is decomposed into harmonic functions with different frequency vectors, which are used to characterize the global surface of the analyzed shape. The descriptor is robust to rotation and translation transformation, nevertheless, it is susceptible to non-rigid deformation.
- **Light Field Descriptor (LFD)** [60]: For each light field descriptor, 10 orthogonal silhouette projections of 3D model are generated by placing 20 virtual cameras on 20 vertices of a regular dodecahedron. And then the camera system is rotated 60 times along 3 connected edges between two vertices so as to be switched onto different vertices. The similarity between two 3D models is measured via the minimum visual similarity under different rotations. In order to enhance robustness against rotation and distribute all the cameras uniformly, they further create a set of light field descriptors by rotating the dodecahedron 91 times on a sphere. Zernike moments and Fourier descriptor are adopted to describe the features of these projected images.
- **Spectral Shape Descriptor (SSD)** [31]: In this method, an affinity matrix of 3D mesh is first formed, each entry of which represents the affinity between two mesh vertices. They define a Gaussian affinity matrix between each pair of vertices, and their affinity is inversely related to the geodesic distance. Moreover, the use of a Gaussian effectively reduces the influence from vertices geodesically far away. Spectral decomposition of the affinity matrix forms a spectral embedding, which achieves normalization against rigid transformation, uniform

scaling, and bending. The eigenvalues specify the variation of the shape along the axes given by the corresponding eigenvectors. Hence they adopt these eigenvalues as spectral shape descriptor.

We plot the recall precision curves for the three previous retrieval methods mentioned above and also our method LSPC in Fig. 7. Clearly, LSPC shows significant improvements for shape retrieval in these non-rigid objects, compared to LFD and SHD designed for searching rigid objects. Moreover, LSPC performs better than SSD, a classical spectral methods, although our method is not applied in the spectral domain. It benefits from several types of patch features against bending and their sparse representations.

#### 7.4. Study on the effect of low-level patch number

We investigate the effect of different number of low-level patches used for the proposed LSPC. We continue to test retrieval performance on McGill 3D shape dataset, and the number of

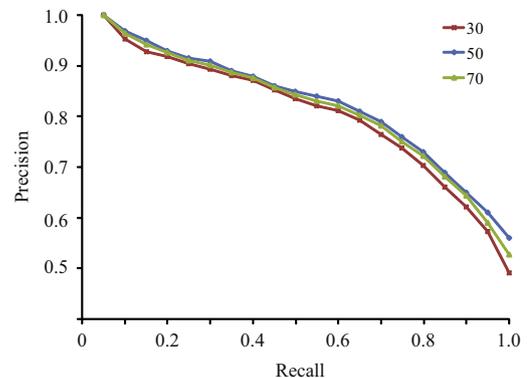


Fig. 9. Recall precision curves on McGill 3D shape dataset in the cases of setting different patch numbers to 30, 50, and 70 respectively.

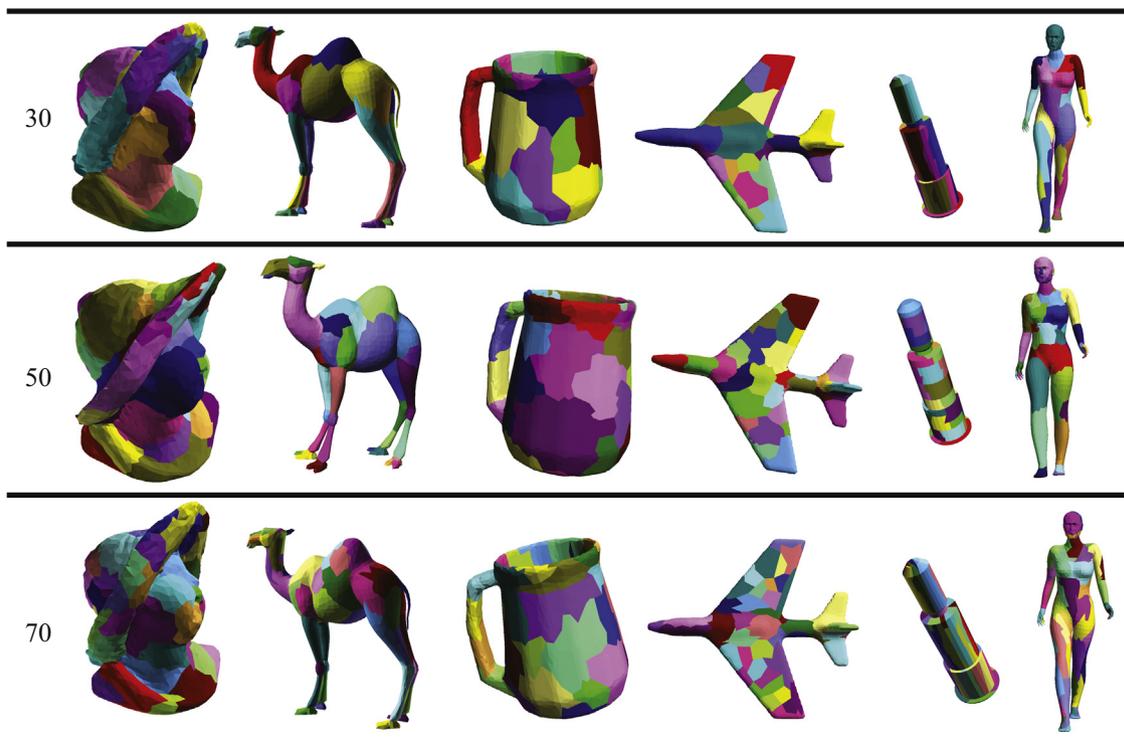


Fig. 8. Visual effects of different low-level patch numbers for several 3D shapes from different categories. The number of low-level patches is set to 30, 50, and 70 respectively.

low-level patches is set to 30, 50, and 70 respectively. Fig. 8 illustrates visual effects of over-segmented patches for several 3D shapes from different categories. The retrieval performance is evaluated using recall precision values, and higher values indicate better retrieval accuracy. The recall precision curves for individual patch numbers are reported in Fig. 9. As can be seen, the number of patches changes from 30 to 50, which leads to better retrieval accuracy. However, if the patch number continues to increase to 70, the retrieval accuracy starts to drop. Based on the experimental study, 50 is finally selected as the optimal low-level patch number.

## 8. Conclusion

In this paper, we introduced the concept of low-level patch to represent a 3D object, which reduces the number of point descriptors, and makes retrieval robust against topology variation. These patches were further encoded into patch words in sparse and locality-constrained form, and the distribution of patch words was used to improve retrieval performance. Experimental results have indicated that our method is comparable to the state-of-the-art methods.

*Limitations and future works:* In this work, we only obtain combinatorial optimum when comparing similarity between two sets of patches after sparse coding, and overlook the structural similarity of patches. This will fail in recognizing some irrelevant objects with similar distributions of patch words, for example, surface of cup and vase body. It will be interesting to investigate similar concepts as spatial pyramid representation in image analysis, and combine with sparse coding for improving 3D shape retrieval performance.

In addition, we find that it is difficult to obtain visually suitable patches for all 3D models while only depending on single geometric criterion, because the over-segmentation generates some trivial patches on the surface for many models. These trivial patches actually impact on retrieval performance, and only prominent patches play a more important role in representing 3D shapes. It would be highly desirable to survey more geometric attributes to generate low-level patches suitable for effective shape retrieval.

## Acknowledgments

The work is partially supported by NSFC (No. 61003137, 61202185, 61473231, and 91120005), NWPU Basic Research Fund 310201401-(JCQ01009, JCQ01012), Fund of National Engineering Center for Commercial Aircraft Manufacturing (201410), and Open Fund of State Key Lab of CAD & CG in Zhejiang University.

## References

- [1] M. Eitz, R. Richter, T. Boubekeur, K. Hildebrand, M. Alexa, Sketch-based shape retrieval, *ACM Trans. Graph. (SIGGRAPH)* 31 (4) (2012), Article 31.
- [2] G. Stavropoulos, P. Moschonas, K. Moustakas, D. Tzovaras, M.G. Strintzis, 3-D model search and retrieval from range images using salient features, *IEEE Trans. Multimed.* 12 (7) (2010) 692–704.
- [3] T. Schreck, B. Bustos, M. Walter, A query-by-example concept and user interface for global and partial 3D object retrieval, in: *Eurographics Workshop on 3D Object Retrieval*, 2009.
- [4] S. Tao, Z. Huang, B. Zuo, Y. Peng, W. Kang, Partial retrieval of CAD models based on the gradient flows in lie group, *Pattern Recognit.* 45 (4) (2012) 1721–1738.
- [5] Y. Yang, H. Lin, Y. Zhang, Content-based 3D model retrieval: a survey, *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* 37 (6) (2007) 1081–1573.
- [6] H. Dutagaci, C.P. Cheung, A. Godil, Evaluation of 3D interest point detection techniques via human-generated ground truth, *Vis. Comput.* 28 (9) (2012) 901–917. <http://dx.doi.org/10.1007/s00371-012-0746-4>.
- [7] F. Tombari, S. Salti, L.D. Stefano, Performance evaluation of 3D keypoint detectors, *Int. J. Comput. Vis.* 102 (1–3) (2013) 198–220. <http://dx.doi.org/10.1007/s11263-012-0545-4>.
- [8] T.-H. Yu, O.J. Woodford, R. Cipolla, A performance evaluation of volumetric 3D interest point detectors, *Int. J. Comput. Vis.* 102 (1–3) (2013) 180–197. <http://dx.doi.org/10.1007/s11263-012-0563-2>.
- [9] P. Heider, A. Pierre-Pierre, R. Li, C. Grimm, Local shape descriptors, a survey and evaluation, in: *Eurographics Workshop on 3D Object Retrieval*, 2011, pp. 49–56.
- [10] P. Li, H. Ma, A. Ming, Combining topological and view-based features for 3D model retrieval, *Multimed. Tools Appl.* 65 (3) (2013) 335–361. <http://dx.doi.org/10.1007/s11042-012-1000-9>.
- [11] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoué, H.V. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, D. Vandermeulen, A comparison of methods for non-rigid 3D shape retrieval, *Pattern Recognit.* 46 (1) (2013) 449–461.
- [12] Y. Gao, J. Tang, R. Hong, S. Yan, Q. Dai, N. Zhang, T.-S. Chua, Camera constraint-free view-based 3D object retrieval, *IEEE Trans. Image Process.* 21 (4) (2012) 2269–2281.
- [13] Y. Gao, M. Wang, D. Tao, R. Ji, Q. Dai, 3-D object retrieval and recognition with hypergraph analysis, *IEEE Trans. Image Process.* 21 (9) (2012) 4290–4303.
- [14] R.M. Rustamov, Laplace–Beltrami eigenfunctions for deformation invariant shape representation, in: *Proceedings of Eurographics Symposium on Geometric Processing*, 2007, pp. 225–233.
- [15] H. Wu, H. Zha, T. Luo, X. Wang, S. Ma, Global and local isometry-invariant descriptor for 3D shape comparison and partial matching, in: *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2010, pp. 438–445.
- [16] A. Kovnatsky, M.M. Bronstein, A.M. Bronstein, D. Raviv, R. Kimmel, Affine-invariant photometric heat kernel signatures, in: *Proceedings of Eurographics Conference on 3D Object Retrieval*, 2012, pp. 39–46.
- [17] M.M. Bronstein, I. Kokkinos, Scale-invariant heat kernel signatures for non-rigid shape recognition, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1704–1711.
- [18] J. Knopp, M. Prasad, G. Willems, R. Timofte, L. van Gool, Hough transform and 3D SURF for robust three dimensional classification, in: *Proceedings of European Conference on Computer Vision*, 2010, pp. 589–602.
- [19] I. Sipiran, Local features for partial shape matching and retrieval, in: *Proceedings of ACM Multimedia*, 2011, pp. 853–856.
- [20] S. Berretti, A.D. Bimbo, P. Pala, Partial match of 3D faces using facial curves between SIFT keypoints, in: *Proceedings of Eurographics Workshop on 3D Object Retrieval*, 2011, pp. 117–120.
- [21] Y.-S. Liu, K. Ramani, M. Liu, Computing the inner distances of volumetric models for articulated shape description with a visibility graph, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (12) (2011) 2538–2544. <http://dx.doi.org/10.1109/TPAMI.2011.116>.
- [22] I. Kokkinos, M.M. Bronstein, R. Litman, A.M. Bronstein, Intrinsic shape context descriptors for deformable shapes, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 159–166.
- [23] A.M. Bronstein, M.M. Bronstein, L.J. Guibas, M. Ovsjanikov, Shape google: geometric words and expressions for invariant shape retrieval, *ACM Trans. Graph.* 30 (1) (2011) 1–20.
- [24] G. Lavoué, Combination of bag-of-words descriptors for robust partial shape retrieval, *Vis. Comput.* 28 (9) (2012) 931–942.
- [25] A. Mademlis, P. Daras, A. Axenopoulos, D. Tzovaras, M.G. Strintzis, Combining topological and geometrical features for global and partial 3-D shape retrieval, *IEEE Trans. Multimed.* 10 (5) (2008) 819–831.
- [26] S. Biasotti, S. Marini, M. Spagnuolo, B. Falcidieno, Sub-part correspondence by structural descriptors of 3D shapes, *Comput. Aided Des.* 38 (9) (2006) 1002–1019.
- [27] V. Barra, S. Biasotti, 3D shape retrieval using kernels on extended reeb graphs, *Pattern Recognit.* 46 (11) (2013) 2985–2999.
- [28] L. Shapira, S. Shalom, A. Shamir, D. Cohen-Or, H. Zhang, Contextual part analogies in 3D objects, *Int. J. Comput. Vis.* 89 (2–3) (2010) 309–326.
- [29] N.D. Cornea, M.F. Demirci, D.E. Silver, A.C. Shokoufandeh, S.J. Dickinson, P.B. Kantor, 3D object retrieval using many-to-many matching of curve skeletons, in: *Proceedings of Shape Modeling International*, 2005, pp. 366–371.
- [30] Z. Liu, K. Zhou, S. Bu, X. Sun, Geometrically attributed binary tree for 3D shape matching, in: *Computer Graphics International Conference*, 2011.
- [31] V. Jain, H. Zhang, A spectral approach to shape-based retrieval of articulated 3D models, *Comput. Aided Des.* 39 (5) (2007) 398–407.
- [32] M. Mahmoudi, G. Sapiro, Three-dimensional point cloud recognition via distributions of geometric distances, *Graph. Models* 71 (1) (2009) 22–31.
- [33] M. Reuter, F.-E. Wolter, N. Peinecke, Laplace–Beltrami spectra as ‘Shape-DNA’ of surfaces and solids, *Comput. Aided Des.* 38 (4) (2006) 342–366.
- [34] Y. Gao, M. Wang, Z. Zha, Q. Tian, Q. Dai, N. Zhang, Less is more: efficient 3D object retrieval with query view selection, *IEEE Trans. Multimed.* 11 (5) (2011) 1007–1018.
- [35] Y. Gao, Q. Dai, M. Wang, N. Zhang, 3d model retrieval using weighted bipartite graph matching, *Signal Process. Image Commun.* 26 (1) (2011) 39–47.
- [36] P. Papadakis, I. Pratikakis, T. Theoharis, S. Perantonis, PANORAMA—a 3D shape descriptor based on panoramic views for unsupervised 3D object retrieval, *Int. J. Comput. Vis.* 89 (2–3) (2010) 177–192.
- [37] B. Leng, Z. Xiong, Modelseek: an effective 3D model retrieval system, *Multimed. Tools Appl.* 51 (3) (2011) 935–962.
- [38] Y. Shan, H.S. Sawhney, B. Matei, R. Kumar, Shapeme histogram projection and matching for partial object recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (4) (2006) 568–577.
- [39] A. Golovinskiy, T. Funkhouser, Randomized cuts for 3D mesh analysis, *ACM Trans. Graph.* 27 (5) (2008), Article No. 145.
- [40] Z. Liu, S. Tang, S. Bu, H. Zhang, New evaluation metrics for mesh segmentation, *Comput. Graph. (SMI)* 37 (6) (2013) 553–564.

- [41] M. Ben-Chen, C. Gotsman, G. Bunin, Conformal flattening by curvature prescription and metric scaling, *Comput. Graph. Forum (Eurograph.)* 28 (2) (2008) 449–458.
- [42] L. Shapira, A. Shamir, D. Cohen-Or, Consistent mesh partitioning and skeletonisation using the shape diameter function, *Vis. Comput.* 24 (4) (2008) 249–259.
- [43] V. Surazhsky, T. Surazhsky, D. Kirsanov, S.J. Gortler, H. Hoppe, Fast exact and approximate geodesics on meshes, *ACM Trans. Graph.* 25 (4) (2005) 553–560.
- [44] J. Sun, M. Ovsjanikov, L.J. Guibas, A concise and provably informative multi-scale signature based on heat diffusion, *Comput. Graph. Forum (SGP)* 28 (5) (2009) 1383–1392.
- [45] F. Bach, J. Mairal, J. Ponce, G. Sapiro, Sparse coding and dictionary learning for image analysis, in: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2010, Tutorial.
- [46] R. Ji, H. Yao, W. Liu, X. Sun, Q. Tian, Task-dependent visual-codebook compression, *IEEE Trans. Image Process.* 21 (4) (2011) 2282–2293.
- [47] R. Ji, L.-Y. Duan, J. Chen, L. Xie, H. Yao, W. Gao, Learning to distribute vocabulary indexing for scalable visual search, *IEEE Trans. Multimed.* 15 (1) (2011) 153–166.
- [48] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1794–1801. doi:10.1109/CVPR.2009.5206757.
- [49] X. Ren, D. Ramanan, Histograms of sparse codes for object detection, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3246–3253. doi:10.1109/CVPR.2013.417.
- [50] L. Shen, C. Ye, B.-S. Hua, Intrinsic image decomposition using a sparse representation of reflectance, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (12) (2013) 2904–2915. <http://dx.doi.org/10.1109/TPAMI.2013.136>.
- [51] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, in: *Proceedings of Neural Information Processing Systems*, 2007, pp. 801–808.
- [52] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3360–3367. doi:10.1109/CVPR.2010.5540018.
- [53] D. Giorgi, S. Biasotti, L. Paraboschi, Shape Retrieval Contest 2007: watertight models track, in: *Proceedings of Shape Modeling International*, 2007, pp. 5–10.
- [54] T. Tung, F. Schmitt, The augmented multiresolution reeb graph approach for content-based retrieval of 3D shapes, *Int. J. Shape Model.* 11 (1) (2005) 91–120.
- [55] T. Tung, F. Schmitt, T. Matsuyama, Topology matching for 3D video compression, in: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [56] D. Zarpalas, P. Daras, A. Axenopoulos, D. Tzovaras, M. G. Strintzis, 3D model search and retrieval using the spherical trace transform, *EURASIP J. Adv. Signal Process.* (2007) Article 23912.
- [57] M. Chaouch, A. Verroust-Blondet, 3D model retrieval based on depth line descriptor, in: *Proceedings of IEEE International Conference on Multimedia and Expo*, 2007, pp. 599–602.
- [58] K. Siddiqi, J. Zhang, D. Macrini, A. Shokoufandeh, S. Bouix, S. Dickinson, Retrieving articulated 3D models using medial surfaces, *Mach. Vis. Appl.* 19 (4) (2008) 261–274.
- [59] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz, Rotation invariant spherical harmonic representation of 3D shape descriptors, in: *Proceedings of Eurographics Symposium on Geometry Processing*, 2003, pp. 156–165.
- [60] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, M. Ouhyoung, On visual similarity based 3D model retrieval, *Comput. Graph. Forum (Eurograph.)* 22 (3) (2003) 223–232.



**Zhenbao Liu** received the Ph.D. degree in computer science from College of Systems and Information Engineering, University of Tsukuba, Japan in 2009. He is an associate professor at Northwestern Polytechnical University, China. He was a visiting scholar in the GRUVi Lab of Simon Fraser University, Canada, in 2012. His research interests include 3D shape analysis, matching, retrieval and segmentation. He has published over 40 journal and conference papers in the related areas.



**Shuhui Bu** received the Ph.D. degree in College of Systems and Information Engineering from University of Tsukuba, Japan in 2009. He is an associate professor at Northwestern Polytechnical University, China. Prior to joining Northwestern Polytechnical University, he was with Kyoto University, Japan. He was a visiting scholar in Simon Fraser University in 2012. His research includes 3D shape analysis, image processing, and computer vision. He has published over 30 journal and conference papers in the related areas.



**Junwei Han** is currently a professor with Northwestern Polytechnical University, Xi'an, China. He received his Ph.D. degree from Northwestern Polytechnical University in 2003. From 2003 to 2010, he worked as a research staff in Nanyang Technological University, Singapore, The Chinese University of Hong Kong, Dublin City University, Ireland, and University of Dundee, UK, respectively. He was also a visiting researcher in Microsoft Research Asia, China, and University of Surrey, UK. His research interests include computer vision and multimedia processing. He has published over 60 journal and conference papers in the related areas, including 6 IEEE Transactions. His paper has been cited over 350 times.