J Inf Process Syst, Vol.14, No.1, pp.176~190, February 2018
https://doi.org/10.3745/JIPS.04.0058

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Convolutional Neural Network Based Multi-feature Fusion for Non-rigid 3D Model Retrieval

Hui Zeng*, Yanrong Liu*, Siqi Li*, JianYong Che**, and Xiuqing Wang***

Abstract

This paper presents a novel convolutional neural network based multi-feature fusion learning method for non-rigid 3D model retrieval, which can investigate the useful discriminative information of the heat kernel signature (HKS) descriptor and the wave kernel signature (WKS) descriptor. At first, we compute the 2D shape distributions of the two kinds of descriptors to represent the 3D model and use them as the input to the networks. Then we construct two convolutional neural networks for the HKS distribution and the WKS distribution separately, and use the multi-feature fusion layer to connect them. The fusion layer not only can exploit more discriminative characteristics of the two descriptors, but also can complement the correlated information between the two kinds of descriptors. Furthermore, to further improve the performance of the description ability, the cross-connected layer is built to combine the low-level features with high-level features. Extensive experiments have validated the effectiveness of the designed multi-feature fusion learning method.

Keywords

Convolutional Neural Network, HKS, Multi-Feature Fusion, Non-rigid 3D Model, WKS

1. Introduction

In recent years, with the rapid development of the computer technology and the multimedia technology, more and more 3D models have been used in many research fields such as face recognition, object recognition, self-driving car, virtual reality, biology and 3D game. Efficient 3D model retrieval has becomes a research hot spot in the field of computer vision. Generally, 3D model retrieval includes the following three steps: model preprocessing, feature extraction, and similarity matching [1]. Among these steps, the 3D feature extraction is the most key step and it plays a decisive role in retrieval results. So it has attracted more and more researchers' attentions.

According to the type of the 3D models, the existing 3D feature extraction methods can be divided into two categories: rigid 3D model based method and non-rigid 3D model based method. Most of the existing methods are designed for the rigid 3D model, such as 3D shape contexts descriptor [2], local surface patch (LSP) descriptor [3], THRIFT descriptor [4], spin image (SI) descriptor [5,6], normal histogram (NormHist) [7], rotational projection statistics (RoPS) descriptor [8], fast point feature

Manuscript received May 3, 2017; first revision June 22, 2017; accepted July 27, 2017.

Corresponding Author: Hui Zeng (hzeng@ustb.edu.cn)
* Beijing Engineering Research Center of Industrial Spectrum Imaging, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China (hzeng@ustb.edu.cn, {1060001224, 1969570533}@qq.com)
** The Tiantan Park Management Office, Beijing, China (67013136@163.com)
*** Vocational & Technical Institute, Hebei Normal University, Shijiazhuang, China (xqwang2013@163.com)

histogram (FPFH) [9], signature histograms of orientations (SHOT) [10,11], and so on. Although the above methods have achieved good retrieval results, they are designed for rigid 3D models and not suitable for non-rigid 3D models. All the above features are not invariant to the non-rigid deformations of the 3D models. To solve this problem, the researchers began to study the 3D feature extraction method for 3D non-rigid model. Sun et al. [12] proposed the heat kernel signature (HKS) descriptor to describe the local characteristics of the non-rigid 3D models. It is based on diffusion scale-space analysis and characterized by the heat transfer process of the 3D surface. The HKS descriptor is invariant under isometric deformations and stable under perturbations of the model. It has achieved good performance in the application of non-rigid 3D model retrieval [13-15]. However, it is sensitive to the scale changes of the 3D model. Aubry et al. [16] proposed the wave kernel signature (WKS) descriptor to describe the non-rigid 3D model, which describes the average probability of quantum mechanics at a position on a non-rigid 3D model surface. The WKS descriptor explains the relationship between the points on the different spatial scales and the rest of the model surface, and it has a better discriminative ability than the HKS descriptor. Up to now, the HKS descriptor and the WKS descriptor have been used in non-rigid 3D model retrieval separately, and the relationships between them have not been fully investigated. To further improve the retrieval performance, it is necessary to study the effective method to fusion the two kinds of descriptors.

Recently, a large number of deep learning methods have been proposed for feature extraction in the field of computer vision, such as auto-encoder, Convolutional Neural Network (CNN), restricted Boltzmann machine (RBM), Deep Belief Networks (DBN), etc. Many experiments have proved that the deep learning based feature extraction methods are more effective. The most widely used deep learning methods in image analysis field is the CNN, which was proposed in 1998 by LeCun et al. [17] to identify handwritten numbers. In 2012, the CNN's accuracy beyond the second nearly 10% in the competition ImageNet. An important feature of the CNN is that it is very similar to our visual system, and the feature is extracted hierarchically. The high-level feature is a combination of low-level features. From low to high, the features are more and more abstract, and this is more conducive to the representation of semantics. Compared with the previous manual designed features, CNN can be automatically extracted more appropriate features, which can greatly improve the recognition performance. So the CNN has been used in more and more areas. For example, the deep learning methods have been used for 3D shape retrieval. Xie et al. [18] proposed a deep shape descriptor for 3D shape retrieval. Firstly, the multiscale shape distribution features are computed. Then a set of discriminative auto-encoders are trained to extract high-level shape features at different scales. Finally, the outputs from hidden layers of the auto-encoders are concatenated to for the shape descriptor. According to our knowledge, the CNN-based 3D retrieval methods have not been fully investigated. In this paper, we proposed a CNN based multi-feature fusion learning method for non-rigid 3D model retrieval. It can combine the effective discriminative information of the HKS descriptor and the WKS descriptor, which not only includes the time domain information of the HKS descriptor but also makes full use of the frequency domain information of the WKS descriptor. Our experimental results have testified that our proposed method can improve the retrieval performance than the single descriptor based method and other state-of-the-art methods.

The rest of the paper is organized as follows. In Section 2, the related works including the HKS descriptor and the WKS descriptor are reviewed. Section 3 gives our proposed CNN-based multi-feature fusion learning method. The non-rigid 3D model retrieval experimental results are presented in Section 4, and some concluding remarks are listed in Section 5.

2. Related Works

For the non-rigid 3D shape retrieval, it is extremely important to select the appropriate feature descriptor. Due to the wide existence of non-rigid deformation, the extracted features must have strong robustness. In this paper, we select the HKS descriptor and the WKS descriptor to describe the 3D local structures, which can describe the 3D local patch from different views and complement with each other.

2.1 HKS Descriptor

Sun et al. [12] proposed the HKS descriptor to describe the local structure of the 3D model. The HKS descriptor is constructed based on the heat diffusion process of the surface of the model, which abandoned the heat kernel's spatial information and only left the time domain information. Considering a 3D model M as a Riemannian manifold, there is the following heat diffusion equation:

$$\left(\Delta_M + \frac{\partial}{\partial t}\right) u(x,t) = 0 \tag{1}$$

where, Δ_M is the positive semi-definite Laplace-Beltrami operator of M and t is time parameter. The solution $u(x,t)$ of Eq. (1) is the amount of heat on the surface at point x in time t. When the $u(x,0) = \delta(x-y)$ is defined as the initial conditions, the solution set of the heat diffusion equation is called heat kernel $k_t(x,y)$, which can be written as follows:

$$k_t(x,y) = \sum_{k=1}^{\infty} e^{-\lambda_i t} \phi_i(x) \phi_i(y) \tag{2}$$

where, x and y are the points of the 3D model, $\lambda_i \geq 0$ is the i^{th} eigenvalue and $\phi_i(x)$ is the i^{th} eigenfunction of the Laplace-Beltrami operator Δ_M, satisfying $\Delta_M \phi_i = \lambda_i \phi_i$ .The heat kernel signature of the point x of the 3D model at time t can be expressed as [12]:

$$h(x,t) = k_t(x,x) = \sum_{k=1}^{\infty} e^{-\lambda_i t} \phi_i^2(x) \tag{3}$$

Then the HKS descriptor of the point x can be obtained by computing its corresponding heat kernel signatures at time sequence. The HKS feature has many excellent characteristics: isometric invariance, multi-scale, and time parameters can be used for representing the small distortion on the model. The heat kernel can be regarded as the transfer density function of the Brownian motion on the fluid, so the local distortion of the model surface will not cause much influence on the heat kernel. In addition, HKS features also have some shortcomings. For example, the HKS descriptor is simplified by the heat kernel.

Although the search efficiency is improved, but the experimental results show that the heat kernel can be affected by the spatial domain. So discarding the spatial domain information is a limitation of HKS characteristics. Secondly, the HKS descriptor is not invariant to the scale of the model. The HKS descriptor mainly contains the low frequency information of the 3D model, ignoring the high frequency information, so it is not suitable for high precision matching. Furthermore, the time parameter t of the HKS descriptor is not directly related to the intrinsic attributes of the 3D model itself. Therefore, for the HKS descriptor, there are some limitations for describing the 3D models.

2.2 WKS Descriptor

Aubry et al. [16] proposed the WKS descriptor for characterizing pints on non-rigid 3D shapes. Given a model M, the WKS descriptor is represented by measuring the average probability of the particles to be measured at each vertex at an energy level. The energy of the particles is related to the frequency, so the effects of different frequencies can be clearly distinguished. The difference between the WKS descriptor and the HKS descriptor is that the WKS descriptor uses the following Schrodinger equation instead of the heat diffusion equation:

$$\frac{\partial \Psi}{\partial t}(x,t) = i\Delta\Psi(x,t) \tag{4}$$

where Δ is the Laplace–Beltrami operator of the 3D model, $\Psi(x,t)$ is the wave function. Then the WKS descriptor can be defined by the following formula:

$$WKS(x,E) = \sum_{k=0}^{\infty} \phi_{k}^{2}(x) f_{E}^{2}(E_k) \tag{5}$$

where $f_{E}^{2}(E_k)$ is the energy probability distribution. In order to select the appropriate energy distribution, let f_E^2 be the Gaussian distribution, the energy scale $e = \log(E_k)$, there are following formula:

$$\begin{cases} WKS(x,\bullet) : \mathrm{R} \to R, \\ WKS(x,e) = C_e \sum_{k=0}^{\infty} \phi_k^2(x) \exp\left(\dfrac{-(e - \log E_k)^2}{2\sigma^2} \right). \end{cases} \tag{6}$$

where $C_e = \left(\sum_{k=0}^{\infty} \exp\left(\dfrac{-(e - \log E_k)^2}{2\sigma^2} \right) \right)^{-1}$.

Like the HKS descriptor, the WKS descriptor also can be looked as an application of a set of filters with the frequency responses $f_{E}^{2}(E_k)$. But the HKS descriptor only uses low-pass filters, and the WKS descriptor uses different frequencies to separate different scales. Compared with the HKS descriptor, the WKS descriptor describes the 3D local patch from a different view. The WKS descriptor not only contains low-frequency information, but also contains high-frequency information. The WKS descriptor is invariant to the non-rigid transformations, and it is stable under perturbations of the 3D shape. So the WKS descriptor is suitable for analyzing 3D shapes undergoing non-rigid deformations.

3. Convolutional Neural Network-Based Multi-feature Fusion

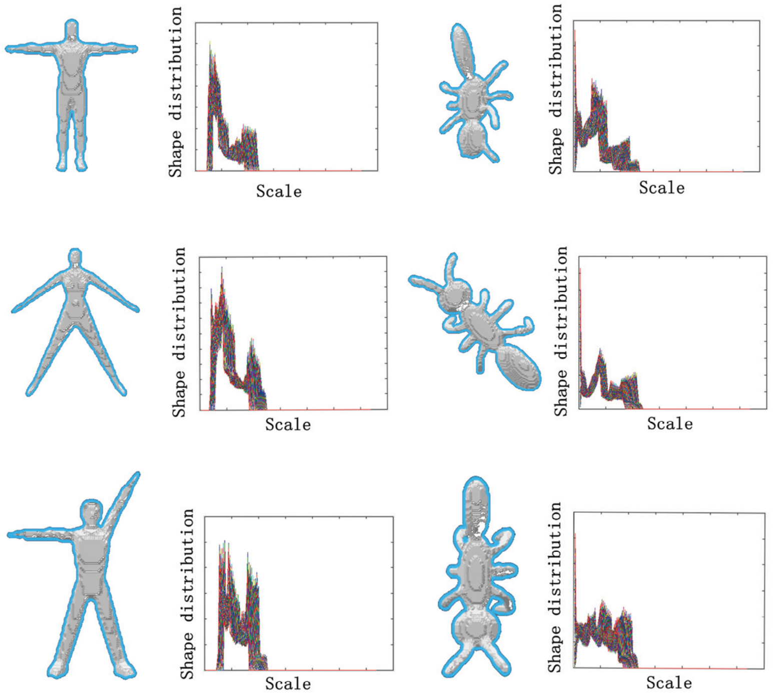
In this section, we firstly introduce the extraction method of the HKS distribution and the WKS distribution. Then the architecture of the multi-feature fusion learning networks and the network optimization are described.

3.1 The HKS Distribution and the WKS Distribution

For each vertex of the 3D model, we can calculate its corresponding HKS descriptor and corresponding WKS descriptor separately. Because the numbers of the vertices of different 3D models are different, the HKS descriptors and the WKS descriptors can't be used as the input of the CNN directly. In this paper, we use the shape distribution to describe the 3D model [19], which refers to a probability distribution sampled from a shape function and can be used as the input of the CNN. For the HKS descriptor, we use the multi-scale shape distribution proposed by Xie et al. [18] to obtain the input of the network.

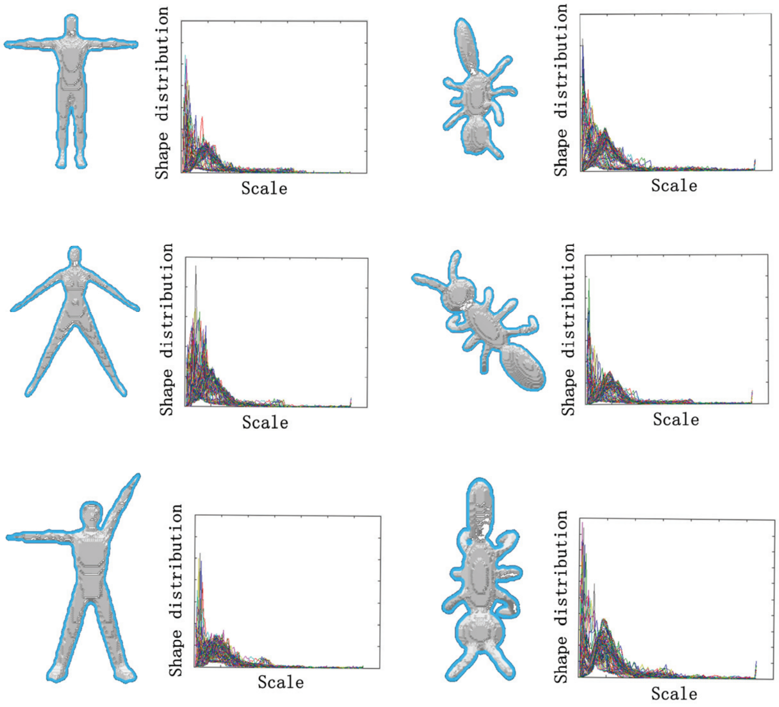
It is a statistics probability distribution of the HKS descriptor of the 3D model. For each scale, we compute the histogram of the HKS descriptor to form the shape distribution. The detailed construction method can be found in [18]. For the WKS descriptor, we can construct the multi-scale shape distribution at each energy using similar method. Here, different scale denotes different energy. In this paper, The HKS multi-scale shape distribution matrix is 128×96, and the number of the diffusion times is 96 and the number of the discrete HKS descriptor values is 128. The WKS multi-scale shape distribution matrix is also 128×96, and the number of the energy values is 96 and the number of the discrete WKS descriptor values is 128. All the above parameters are determined by experiments. Fig. 1 shows the HKS multi-scale shape distributions of the human models and the ant models with different poses. From Fig. 1 we can see that the multi-scale shape distributions are different for different classes, and the 3D models of the same class have similar multi-scale shape distributions. Although these 3D models of the same class have different postures and the details of the shape distributions are different, their main features have been captured by the multi-scale shape distributions.

Fig. 2 shows the WKS multi-scale shape distributions of the human models and the ant models with different poses. From Fig. 2 we can see that the distributions have not clear differences for different models. This is because that the WKS descriptor is represented by the average probabilities of quantum particles of different energy levels. The scale of the WKS multi-scale shape distribution denotes the energy, and the 3D models with different postures of the same class usually correspond different energies. So it is necessary to extract more discriminative features from the HKS multi-scale shape distributions and the WKS multi-scale shape distributions, and study the effective fusion method to further improve the retrieval performance.



(a) (b)

Fig. 1. The HKS multi-scale shape distributions of human models (a) and ant models (b).



(a) (b)

Fig. 2. The WKS multi-scale shape distributions of human models (a) and ant models (b).

3.2 The Architecture of the Multi-feature Fusion Learning

Inspired by the multi-modal deep learning method for RGB-D object recognition [20], we proposed a conventional neural network based multi-feature fusion learning architecture for non-rigid 3D model retrieval. Our proposed multi-feature fusion learning architecture includes two CNNs and a multi-feature fusion layer. At first, two CNNs are built to learn the HKS and WKS features, and the inputs of them are the HKS and WKS distributions. As shown in Fig. 3, each CNN consists of three convolutional layers (C1, C2, C3), three pooling layers (S1, S2, S3), one cross-connected layer and two fully-connected layers. Then the last pool layer of each CNNs is combined with the penultimate pool layer into a cross-connected layer, which can fully utilize the characteristics of hidden layer to further improve the retrieval performance. The cross-connected layer is composed of S2 layer and S3 layer. Finally, the multi-feature fusion layer is used to fuse the two descriptor of the fully-connected layers. It combines the two kinds of features with two feature transformation matrix Q_1 and Q_2 .

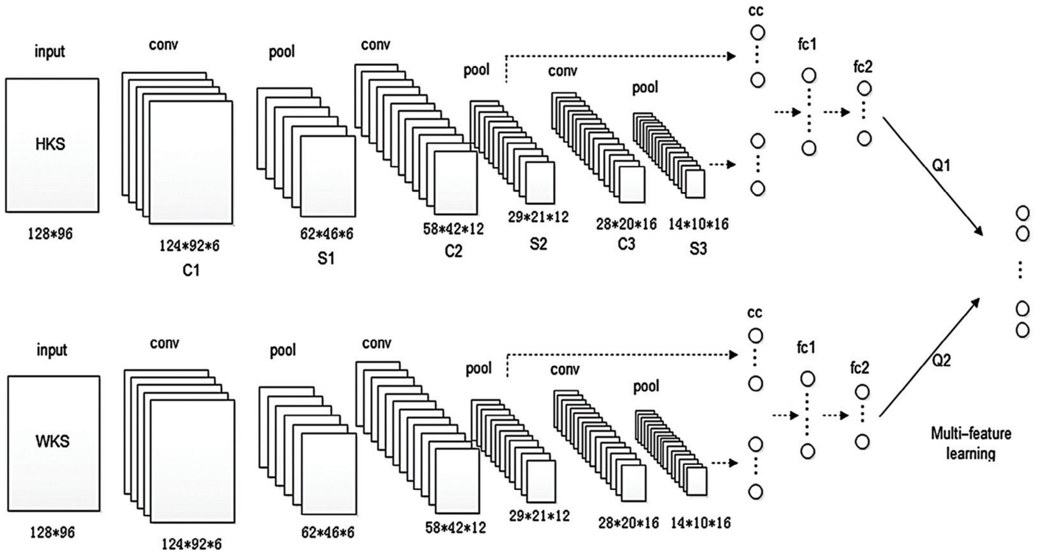


Fig. 3. Our multi-feature fusion network, where “conv”, ”pool”, “cc”, ”fc1” and “fc2” represent “convolutional”, “mean-pooling”, “cross-connected layers”, “first fully-connected layer of the HKS/WKS feature” and “second fully-connected layer of the HKS/WKS feature” respectively.

Let A be the output of the second fully-connected layer of the HKS feature, where $A=\left[a_1,a_2,\cdots,a_N\right]\in R^{M\times N}$, where M is the dimension of the activations and N is the number of the training samples. Similarly, B is the output of the second fully-connected layer of the WKS feature, and $B=\left[b_1,b_2,\cdots,b_N\right]\in R^{M\times N}$. The transformation matrixes of the two kinds of features are Q_1 and Q_2, where $a_i'=Q_1a_i$, $b_i'=Q_2b_i$, and a_i' and b_i' are weighted by k_1 and k_2 to form the final learned features. In the process of learning, we simultaneously optimize the matrixes Q_1 , Q_2 and the weights k_1 and k_2 .

3.3 Network Optimization

In order to learn Q_1 and Q_2 for the two kinds of features to obtain better representations, we consider the discriminative information and related information of the two kinds of features. Our objective is to learn the fusion features to minimize the distances between same-class samples and maximize the distances between different-class samples. So the objective function is defined as follows:

$$\min_{\{Q_1,Q_2,k_1,k_2\}} F = k_1 D_1(Q_1) + k_2 D_2(Q_2) + \lambda \mathrm{C}(Q_1,Q_2) \qquad (7)$$
$$\text{subject to } k_1 + k_2 = 1, k_1 \geq 0, k_2 \geq 0, \lambda > 0$$

where D_1 and D_2 represent the discriminative terms, C denotes the correlated term, and λ is the weight between the discriminative terms and the related term. Here the discriminative term D_1 is defined as:

$$D_1\left(Q_1\right) = \sum_{ij} h\left(t_1 - y_{ij}\left(u_1 - d_{Q_1}\left(a_i,a_j\right)\right)\right) \qquad (8)$$

where $h\left(x\right) = \max\left(0,x\right)$, and d_{Q_1} represents the distance of transformed feature $Q_1 a_i$ and $Q_1 a_j$. The distance d_{Q_1} can be computed as:

$$d_{Q_1}(a_i,a_j)=(Q_1 a_i - Q_1 a_j)^T \bullet (Q_1 a_i - Q_1 a_j) \qquad (9)$$

If a_i and a_j are the features from the same class, the distance d_{Q_1} should be smaller than a given threshold $u_1 - t_1(u_1 > t_1 > 0)$. If a_i and a_j are the features from different classes, the distance d_{Q_1} should be larger than a given threshold $u_1 + t_1$. So we can conclude that the distances of the same and different class have the following constraint:

$$y_{ij}(u_1 - d_{Q_1}(a_i,a_j) > t_1 \qquad (10)$$

For the label y_{ij}, we define it as follows: when a_i and a_j are from the same class, $y_{ij} = 1$. when a_i and a_j are from different classes, $y_{ij} = -1$. For the WKS feature b_i, we have similar definitions and constraints.

The correlated term is used to exploit the complementary information of the two kinds of features, and it is defined by the difference of pairwise distances between the transformed HKS features and the transformed WKS features. The correlated term can be described as:

$$C\left(Q_1,Q_2\right) = \sum_{ij}\left(\sqrt{d_{Q_1}(a_i,a_j)} - \sqrt{d_{Q_2}(b_i,b_j)}\right)^2 \qquad (11)$$

For the two 3D models from the same class, their corresponding distance d_{Q_1} is small and their corresponding distance d_{Q_2} is also small. For the two 3D models from different classes, their corresponding distance d_{Q_1} is large and their corresponding distance d_{Q_2} is also large. So we can minimize the correlated term $C\left(Q_1,Q_2\right)$ to optimize the transform matrixes Q_1 and Q_2.

In this paper, we use an alternating optimization approach to obtain the optimal solution for Equation (7). At first, Q_1 and Q_2 are fixed while k_1 and k_2 are optimized. Secondly, k_1 and k_2 are fixed while Q_1 and Q_2 are optimized. Furthermore, in order to avoid suboptimal results and increase the non-linearity, we adopt the strategy proposed in [20] and modify the objective function to:

$$\begin{gathered}\min_{\{Q_1,Q_2,k_1,k_2\}} F = {k_1}^p D_1(Q_1) + {k_2}^p D_2(Q_2) + \lambda C(Q_1,Q_2) \\ \text{subject to } k_1 + k_2 = 1, k_1 \geq 0, k_2 \geq 0, \lambda > 0\end{gathered} \tag{12}$$

where $p>1$. In Eq. (12), we use k_1^p and k_2^p instead of k_1 and k_2, this will balance the weights of the two kinds of features. Then we can construct the following Lagrangian function:

$$L\left(k,\eta\right) = {k_1}^p D_1 + {k_2}^p D_2 + \lambda C - \eta(k_1 + k_2 - 1) \tag{13}$$

By setting $\dfrac{\partial L(k,\eta)}{\partial k}$ and $\dfrac{\partial L(k,\eta)}{\partial \eta}$ to 0, the weight $k_m (m=1,2)$ can be updated as:

$$k_m = \frac{\left(\dfrac{1}{D_m}\right)^{\frac{1}{p-1}}}{\sum_{m=1}^{2}\left(\dfrac{1}{D_m}\right)^{\frac{1}{p-1}}} \tag{14}$$

And then use the back propagation algorithm to update the transform matrix, Q_1 and Q_2. The derivative of Q_1 can be calculated as follows:

$$\frac{\partial F}{\partial Q_1} = 2Q_1\left[{k_1}^p \sum_{ij} y_{ij} h'\left(t_1 - y_{ij}(u_1 - d_{Q_1}(a_i,a_j))\right)A_{ij}^1 + \lambda \sum_{ij}\left(1 - \sqrt{\frac{d_{Q_2}(b_i,b_j)}{d_{Q_1}(a_i,a_j)}}\right)A_{ij}^1\right] \tag{15}$$

where $A_{ij}^1 = \left(a_i - a_j\right)\left(a_i - a_j\right)^T$. So the transform matrix Q_1 can be updated using the following equation:

$$Q_1 = Q_1 - \beta\frac{\partial F}{\partial Q_1} \tag{16}$$

The updating method of the transform matrix Q_2 is similar to Q_1.

In the process of the back-propagation, the derivatives of D and C for a_i and b_i is used. For a_i, they can be described as follows:

$$\frac{\partial D_1}{\partial a_i} = \sum_j y_{ij}\left(Q_1^T Q_1 + Q_1 Q_1^T\right)\left(a_i - a_j\right)h'\left(t_1 - y_{ij}\left(u_1 - d_{Q_1}\left(a_i, a_j\right)\right)\right) \tag{17}$$

$$\frac{\partial C}{\partial a_i} = \sum_j \left(Q_1^T Q_1 + Q_1 Q_1^T\right)\left(a_i - a_j\right)\sqrt{\frac{dQ_1\left(a_i, a_j\right) - dQ_2\left(b_i, b_j\right)}{dQ_1\left(a_i, a_j\right)}} \tag{18}$$

Likewise for b_i to compute the derivatives of D and C . The proposed multi-feature fusion learning method can be listed as follows ：

(i) For each training sample, compute the HKS descriptor and the WKS descriptor of each vertex;

(ii) Compute the HKS multi-scale shape distribution and the WKS multi-scale shape distribution of each 3D model;

(iii) Randomly initialize each conventional neural network, and then pre-training to get pre-trained a_i and b_i .respectively.

(iv) Perform an alternating method to optimize Q_1 , Q_2 , k_1 and k_2 : Fix Q_1 and Q_2 , firstly, update k_1 and k_2 , and then fix k_1 and k_2 , update Q_1 and Q_2 .

(v) Perform back-propagation: fix Q_1 , Q_2 , k_1 and k_2 , update the parameters of the conventional neural networks with gradient descent method.

(vi) Repeat (iv)–(v) until convergence or reach the maximum number of iterations.

4. Experimental Results

In this paper, the McGill 3D shape benchmark is used for non-rigid 3D model retrieval experiments to evaluate the performance of our proposed method [20]. This benchmark contains 255 non-rigid 3D models from 10 different categories, including: ant, crab, spectacle, hand, human, octopus, plier, snake, spider and teddy-bear. Each category has 20–30 3D models. Among these models, there are rotational transformation, scale transformation and non-rigid deformation. In our non-rigid 3D model retrieval experiments, 15 3D models per class are randomly selected for training and the others for querying.

Some example 3D models are shown in Fig. 4. The experimental environment was i7-6700 CPU 3.40 GHz 12.0G memory Lenovo computer with MATLAB R2014a. To evaluate the retrieval performance, we use the following measures: nearest neighbor (NN), the first tier (FT), the second tier (ST) and the discounted cumulative gain (DCG).

In this experiments, the HKS descriptor and the WKS descriptor of each vertex are firstly computed. Secondly, the HKS multi-scale shape distribution and the WKS multi-scale shape distribution are calculated, which can be used as the inputs of the two neural networks. Then the two conventional neural networks and the fusion layer are built. Table 1 gives a detailed description of each conventional neural network, including the type of each layer, the size of the convolutional kernel, the stride and the output size of each layer. The activation function of all convolutional layers and fully-connected layers are ReLU. One of the advantages of the ReLU function is that it can learn the features faster. The other advantage is the biology of rationality, which is unilateral. Compared with sigmoid function and tanh function, the ReLU function conforms to the characteristics of biological neurons. In this paper, the pre-training methods are used to initialize the networks, which can make the network converge faster. The two conventional neural networks are trained independently with the HKS distribution and the WKS distribution. The learning rate is set to 0.05, Q_m is initialized as an identity of 1000×1000, and k is initialized as [0.5, 0.5]. The $u_1, u_2, t_1, t_2, p, \beta, \lambda$ were set as 700, 500, 70, 50, 2, 0.0003, 0.009, respectively. In the process of learning, we use the stochastic gradient descent to update the parameters.

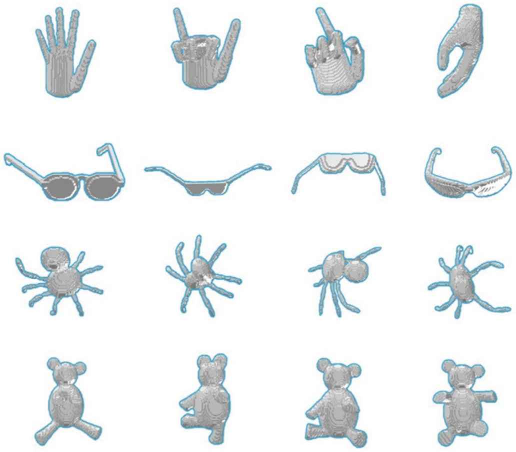


Fig. 4. Example 3D models of the McGill 3D shape benchmark.

Table 1. Description of each conventional neural network

Layer	Type	Patch size	Stride	Number of feature maps	Output size
x	Input				128×96
C1	Convolution	5×5	1	6	124×92
S1	Mean pooling	2×2	2	6	62×46
C2	Convolution	5×5	1	12	58×42
S2	Mean pooling	2×2	2	12	29×21
C3	Convolution	2×2	1	16	28×20
S3	Mean pooling	2×2	2	16	14×10
cc	Cross-connected			1	9548
fc1	Fully-connected			1	3000
fc2	Fully-connected			1	1000
ff	Feature fusion layer			1	2000

To validate the effectiveness of the multi-feature fusion, we compare the retrieval results of our proposed method with single feature based method. For the HKS descriptor, we firstly compute the HKS multi-scale shape distribution. Then the conventional neural network is built and its structure is the same as the single conventional neural network used in the multi-feature fusion learning method. Likewise for the WKS descriptor. Table 2 gives the retrieval results of the single-feature based methods and the multi-feature fusion method. From Table 2 we can see that compared with the single-feature based methods, our proposed multi-feature fusion learning method has better performance with the NN, FT, ST, and DCG measures. It shows that the fusion layer can learn more useful information. So compared with single-feature based method, using the HKS descriptor and the WKS descriptor simultaneously and our proposed conventional neural network based fusion method can improve the retrieval results effectively.

Then we compare our proposed method to the covariance descriptor based method [21], the graph-based method [22], the PCA-based VLAT method [23], the hybrid BOW method [24], hybrid 2D/3D method [25], the CBoFHKS method [26] and the discriminative auto-encoder based shape descriptor (DASD) method [18]. Table 3 gives the retrieval results of our proposed multi-feature fusion learning method and other methods. From Table 3 we can see that our proposed method has the best performance with the FT, ST, and DCG measures and have comparable performance with NN measures. So our proposed method is more robust to non-rigid deformations and it has achieved very competitive results compared with other methods.

Table 2. Retrieval results compared with single-feature based methods

	NN	FT	ST	DCG
HKS	0.819	0.622	0.744	0.827
WKS	0.914	0.775	0.866	0.914
Multi-feature fusion learning method	0.971	0.905	0.981	0.963

Table 3. Retrieval results compared with other methods

Methods	NN	FT	ST	DCG
Covariance method [21]	0.977	0.732	0.818	0.937
Graph-based method [22]	0.976	0.741	0.911	0.933
PCA-based VLAT [23]	0.969	0.658	0.781	0.894
Hybrid BOW [24]	0.957	0.635	0.790	0.886
Hybrid 2D/3D [25]	0.925	0.557	0.698	0.850
CBoFHKS [26]	0.901	0.778	0.876	0.891
DASD [18]	0.988	0.782	0.834	0.955
Multi-feature fusion learning method	0.971	0.905	0.981	0.963

5. Conclusion

In this paper, we have proposed a novel multi-feature fusion learning method for non-rigid 3D model retrieval. Firstly, the HKS descriptor and the WKS descriptor are computed. Then the corresponding HKS multi-scale shape distribution and the WKS multi-scale shape distribution are constructed to be used as the inputs of the conventional neural networks. Finally the conventional neural networks based multi-feature fusion learning framework is built to obtain the fusion feature. The contribution of this paper is that we use the cross-connected layer to combine the low-level features with high-level features, and the fusion layer can learn not only the discriminative characteristics of the two kinds of descriptors but also the correlated information between them. So our proposed fusion feature can make full use of the effective information of the HKS descriptor and the WKS descriptor, and it has achieved a better retrieval performance.

Acknowledgement

This paper is supported by the National Natural Science Foundation of China (No. 61375010 and 61503224), the Fundamental Research Funds for the China Central Universities of USTB (No. FRF-BD-16-005A), Beijing Key Discipline Development Program (No. XK100080537).

References

[1] Y. Matsuda, N. Miura, A. Nagasaka, H. Kiyomizu, and T. Miyatake, “Finger-vein authentication based on deformation-tolerant feature-point matching,” *Machine Vision and Applications*, vol. 27, no. 2, pp. 237-250, 2016.

[2] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik, “Recognizing objects in range data using regional point descriptors,” in *Proceedings of the European Conference on Computer Vision*, Prague, Czech Republic, 2004, pp. 224-237.

[3] H. Chen and B. Bhanu, “3D free-form object recognition in range images using local surface patches,” *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1252-1262, 2007.

[4] A. Flint, A. Dick, and A. Van den Hengel, “Local 3D structure recognition in range images,” *IET Computer Vision*, vol. 2, no. 4, pp. 208-217, 2008

[5] A. E. Johnson and M. Hebert, “Surface matching for object recognition in complex three-dimensional scenes,” *Image and Vision Computing*, vol. 16, no. 9-10, pp. 635-651, 1998.

[6] A. E. Johnson and M. Hebert, “Using spin images for efficient object recognition in cluttered 3D scenes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 443-449, 1999.

[7] G. Hetzel, B. Leibe, P. Levi, and B. Schiele, “3D object recognition from range images using local feature histograms,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, HI, 2001, pp. 394-399.

[8] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, “Rotational projection statistics for 3D local surface description and object recognition,” *International Journal of Computer Vision*, vol. 105, no. 1, pp. 63-86, 2013.

[9] R. B. Rusu, N. Blodow, and M. Beetz, “Fast point feature histograms (FPFH) for 3D registration,” in *Proceedings of the IEEE International Conference on Robotics and Automation*, Kobe, Japan, 2009, pp. 3212-3217.

[10] F. Tombari, S. Salti, and L. Di Stefano, “Unique signatures of histograms for local surface description,” in *Proceedings of the European Conference on Computer Vision*, Crete, Greece, 2010, pp. 356-369.

[11] S. Salti, F. Tombari, and L. Di Stefano, “SHOT: unique signatures of histograms for surface and texture description,” *Computer Vision & Image Understanding*, vol. 125, no. 8, pp. 251-264, 2014.

[12] J. Sun, M. Ovsjanikov, and L. Guibas, “A concise and provably informative multi-scale signature based on heat diffusion,” *Computer Graphics Forum*, vol. 28, no. 5, pp. 1383-1392, 2009.

[13] A. M. Bronstein, M. M. Bronstein, R. Kimmel, M. Mohmoudi, and G. Sapiro, "A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching," *International Journal of Computer Vision*, vol. 89, no. 2-3, pp. 266-286, 2010.

[14] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, et al., "A comparison of methods for non-rigid 3D shape retrieval," *Pattern Recognition*, vol. 46, no. 1, pp. 449-461, 2013.

[15] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikow, "Shape Google: geometric words and expressions for invariant shape retrieval," *ACM Transactions on Graphics*, vol. 30, no. 1, pp. 623-636, 2011.

[16] M. Aubry, U. Schlickewei, and D. Cremers, "The wave kernel signature: a quantum mechanical approach to shape analysis," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, Barcelona, Spain, 2011, pp. 1626-1633.

[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[18] J. Xie, G. Dai, F. Zhu, E. K. Wong, and Y. Fang, "DeepShape: deep-learned shape descriptor for 3D shape retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 7, pp. 1335-1345, 2017.

[19] R. Osada, T. Funkhouser, B. Chazelle, and D. Dokin, "Shape distributions," *ACM Transactions on Graphics*, vol. 21, no. 4, pp. 807-832, 2002.

[20] A. Wang, J. Lu, J. Cai, T. J. Cham, and G. Wang "Large-margin multi-modal deep learning for RGB-D object recognition," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1887-1898, 2015.

[21] H. Tabia, H. Laga, D. Picard, and P. H. Gosselin, "Covariance descriptors for 3D shape matching and retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 4185-4192.

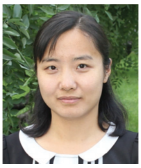
[22] A. Agathos, I. Pratikakis, P. Papadakis, S. J. Perantonis, P. N. Azariadis, and N. S. Sapidis, "Retrieval of 3D articulated objects using a graph-based representation," in *Proceedings of the Eurographics Workshop on 3D Object Retrieval*, Munich, Germany, 2009, pp. 29-36.

[23] H. Tabia, D. Picard, H. Laga, and P. H. Gosselin, "Compact vectors of locally aggregated tensors for 3D shape retrieval," in *Proceedings of the Eurographics Workshop on 3D Object Retrieval*, Girona, Spain, 2013, pp. 17-24.

[24] P. Papadakis, I. Pratikakis, T. Theoharis, G. Passalis, and S. Perantonis, "3D object retrieval using an efficient and compact hybrid shape descriptor," in *Proceedings of the Eurographics Workshop on 3D Object Retrieval*, Crete, Greece, 2008, pp. 9-16.

[25] G. Lavoue, "Combination of bag-of-words descriptors for robust partial shape retrieval," *The Visual Computer*, vol. 28, no. 9, pp. 931-942, 2012.

[26] Z. Lian, A. Godil, T. Fabry, T. Furuya, J. Hermans, R. Ohbuchi, et al., "SHREC'10 Track: non-rigid 3D shape retrieval," *Proceedings of the Eurographics Workshop on 3D Object Retrieval*, Zurich, Switzerland, 2015, pp. 107-120.



Hui Zeng https://orcid.org/0000-0002-4137-7424

She received B.S. and M.S. degrees from Shandong University in 2001 and 2004, respectively, and received the Ph.D. degree from National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences in 2007. She is currently an associate professor at School of Automation and Electrical Engineering, University of Science and Technology Beijing, China. Her main research interests include computer vision, pattern recognition and machine learning.

Yanrong Liu

He received B.S. degree from Shanxi University in 2015. His current research interests include computer vision and pattern recognition.

Siqi Li

She received B.S. degree from University of Science & Technology Beijing in 2015. Her current research direction is computer vision and pattern recognition

JianYong Che

He received B.S. degree from China University of Geosciences in 1999. He is currently a staff member at the Tiantan Park Management Office, Beijing, China. His current research direction is 3D model analysis.

Xiuqing Wang

She received the Ph.D. degree from Institute of Automation, Chinese Academy of Sciences in 2007. She is currently a professor at Vocational & Technical Institute, Hebei Normal University, China. Her main research interests include intelligence robot and machine learning.