Jin Xie

Personal Information

Status:	**Post-doctor**
Program:	**Computer Science and Engineering**
School:	**Tandon School of Engineering, New York University**
Period:	**From 2014-01 to 2017-12**

Biography

I was a post-doctor at New York University and advised by Professor Yi Fang. During my post doctoral period, I was a research assistant in NYU Multimedia and Visual Computing (MMVC) Lab. I am broadly interested in 3D Computer Vision and Deep Learning. Now I am full Professor at Nanjing University of Science and Technology, China.

Research Project: Deep nonlinear metric learning for 3-D shape retrieval

Description

Effective 3-D shape retrieval is an important problem in 3-D shape analysis. In this project, motivated by the fact that deep neural network has the good ability to model nonlinearity, we propose to learn an effective nonlinear distance metric between 3-D shape descriptors for retrieval. The proposed deep metric network minimizes a discriminative loss function that can enforce the similarity between a pair of samples from the same class to be small and the similarity between a pair of samples from different classes to be large. Finally, the distance between the outputs of the metric network is used as the similarity for shape retrieval. The proposed method is evaluated on the McGill, SHREC'10 ShapeGoogle, and SHREC'14 Human shape datasets. Experimental results on the three datasets validate the effectiveness of the proposed method.

Method

We propose a novel deep nonlinear metric learning method for 3-D shape retrieval. First, we employ the locality-constrained linear coding (LLC) method to encode each vertex of 3-D shapes to form a global 3-D shape descriptor. We then develop a deep metric network to learn a nonlinear transformation to map the global 3-D descriptors to a nonlinear feature space. The learned distance metric can minimize a discriminative loss function so that the similarities between the pairs of samples from the same class are as small as possible and the similarities between the pairs of samples from different classes are as large as possible. Furthermore, in order to make the learned distance metric to be more discriminative, we also encourage that the neurons in the hidden layers of the metric network are as close as possible to their means.

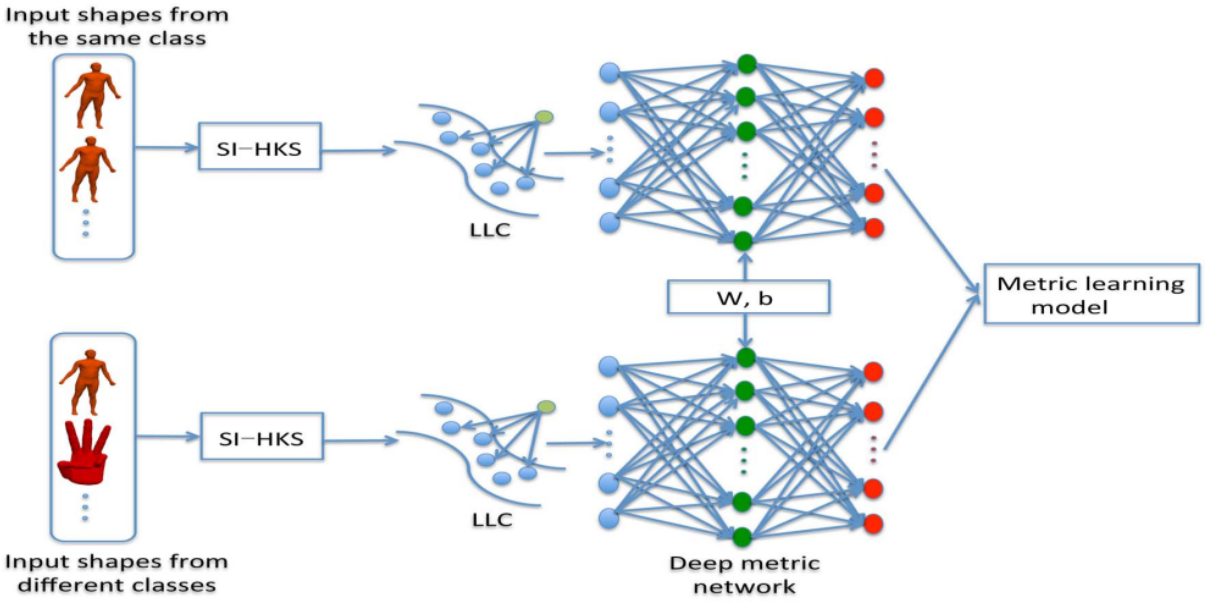


Figure 1: Proposed deep shape metric learning framework. For the input shapes, we employ the LLC method to encode the extracted SI-HKSs to form the global 3-D shape descriptors. The global shape descriptors of the input shapes from the same class and different classes are then fed into the deep metric learning model so that the similarity between the pairs of shapes from the same class are as small as possible and the similarity between the pairs of shapes from different classes are as large as possible.

Results

We first evaluate our proposed deep nonlinear metric learning based shape retrieval method, and then compare it with the state-of-the-art 3-D shape retrieval methods on three benchmark datasets, i.e., McGill shape dataset, SHREC'10 ShapeGoogle dataset, and SHREC'14 Human dataset.

Methods	NN	FT	ST	DCG
Covariance descriptor [19]	**0.977**	0.732	0.818	0.937
Graph-based method [35]	0.976	0.741	0.911	0.933
PCA based VLAT [36]	0.969	0.658	0.781	0.894
Hybrid BOW [37]	0.957	0.635	0.790	0.886
Hybrid 2D/3D [38]	0.925	0.557	0.698	0.850
Manifold ranking [23]	-	0.761	-	-
Proposed DNML	0.962	**0.906**	**0.969**	**0.967**

Figure 2: Retrieval results on the McGill dataset

Transformation	VQ [18]	UDL [20]	SDL [20]	Proposed DNML
Isometry	0.988	0.977	0.994	**1.000**
Topology	**1.000**	**1.000**	**1.000**	**1.000**
Isometry+Topology	0.933	0.934	0.956	**0.979**
Partiality	0.947	0.948	0.951	**0.983**
Triangulation	0.954	0.950	**0.955**	0.943

Figure 3: Retrieval results on the SHREC'10 ShapeGoogle dataset

Method	Synthetic model	Scanned model
HAPT[39]	0.817	0.637
ISPM[40]	0.92	0.258
RBiHDM[41]	0.642	0.640
DBN[34]	0.842	0.304
VQ [18]	0.813	0.514
UDL [20]	0.842	0.523
SDL [20]	0.95.1	0.791
Proposed DNML	**0.973**	**0.801**

Figure 4: Retrieval Results on The SHREC'14 Human Dataset

Research Project: Progressive shape-distribution-encoder for learning 3D shape representation

Description

Since there are complex geometric variations with 3D shapes, extracting efficient 3D shape features is one of the most challenging tasks in shape matching and retrieval. In this project, we propose a deep shape descriptor by learning shape distributions at different diffusion time via a progressive shape-distribution-encoder (PSDE). First, we develop a shape distribution representation with the kernel density estimator to characterize the intrinsic geometry structures of 3D shapes. Then, we propose to learn a deep shape feature through an unsupervised PSDE. Finally, we concatenate all neurons in the middle hidden layers of the unsupervised PSDE network to form an unsupervised shape descriptor for retrieval. The proposed method is evaluated on three benchmark 3D shape data sets with large geometric variations, i.e., McGill, SHREC'10 ShapeGoogle, and SHREC'14 Human data sets, and the experimental results demonstrate the superiority of the proposed method to the existing approaches.

Method

We propose a deep shape descriptor for retrieval by learning shape distributions between consecutive diffusion time. First, based on the heat kernel, we develop a shape distribution representation with the kernel density estimation method. We model the complex non-linear change of the shape distributions between consecutive diffusion time through a deep network. Particularly, we restore the denoising auto-encoder to propose an unsupervised progressive shape-distribution-encoder (PSDE) to achieve this goal. Finally, we concatenate all neurons in the middle hidden layers of the unsupervised PSDE network, i.e., the discriminative shape distributions, to form an unsupervised deep shape descriptor. Furthermore, in order to better exploit the discriminative information from the hidden layers of the unsupervised PSDE, we impose a constraint on all hidden layers to propose a supervised PSDE so that for each hidden layer the outputs from the same class are as similar as possible while the outputs from

different classes are as dissimilar as possible. The neurons in the middle hidden layers of the supervised PSDE are concatenated to form a supervised shape descriptor.

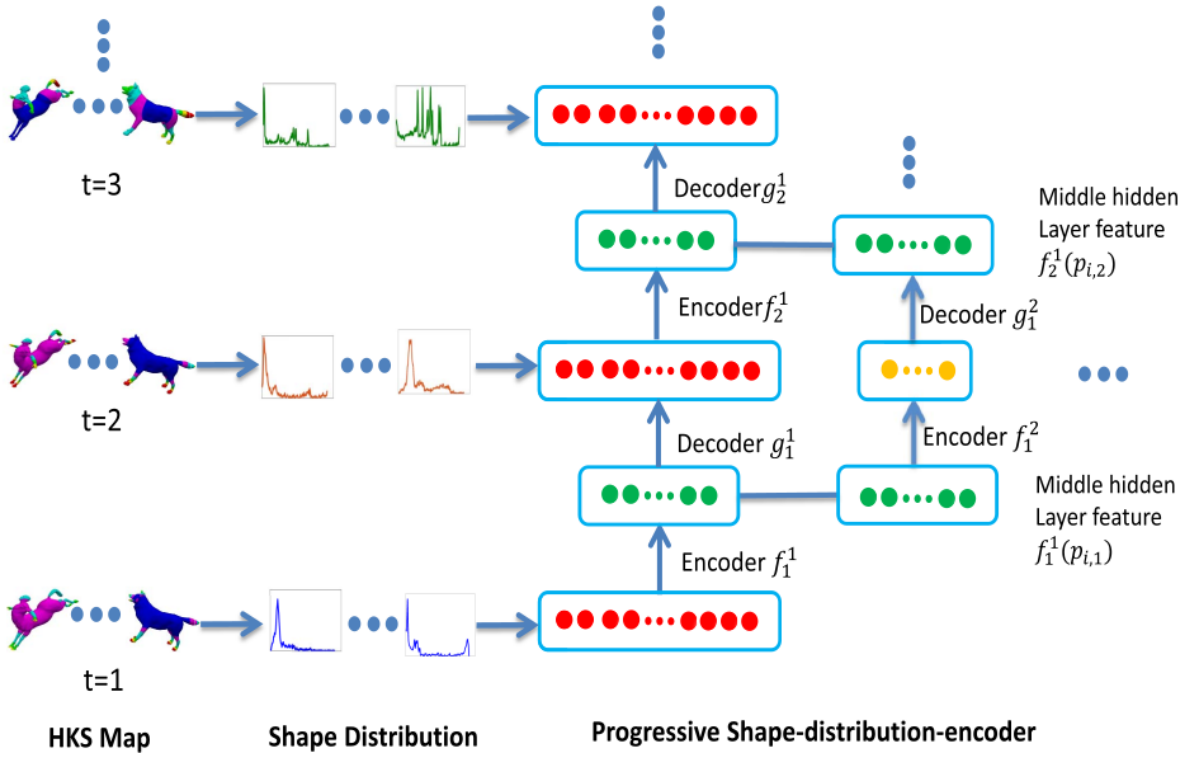


Figure 5: The framework of the proposed unsupervised PSDE. The shape distributions at t = 1 and t = 2 are fed into the fist unsupervised PSDE in the first level while the shape distributions at t = 2 and t = 3 are fed into the second unsupervised PSDE. Then, the learned middle hidden layer features and are used as the input and output of the unsupervised PSDE in the second level. Thus, the middle hidden layer features of a set of the PSDEs in level l are recursively fed into the unsupervised PSDEs in level l + 1 to form an unsupervised deep representation.

Results

We compare our proposed shape descriptor with the state-of-the-art methods on three benchmark datasets, i.e., McGill shape dataset, SHREC'10 Shape-Google dataset and SHREC'14 Human dataset. As evaluated, experimental results demonstrate that the proposed shape descriptors can yield good performance and be robust to noise.

Methods	NN	1-Tier	2-Tier	DCG
Covariance descriptor [22]	0.977	0.732	0.818	0.937
PCA based VLAT [31]	0.969	0.658	0.781	0.894
Hybrid BOW [30]	0.957	0.635	0.790	0.886
Hybrid 2D/3D [32]	0.925	0.557	0.698	0.850
UPSDE	0.984	0.783	0.841	0.941
SPSDE	**0.986**	**0.883**	**0.911**	**0.952**

Figure 6: Retrieval results on the McGill dataset

Transformation	VQ [21]	UDL [19]	SDL[19]	UPSDE	SPSDE
Isometry	0.988	0.977	0.994	**1.000**	**1.000**
Topology	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Isometry+Topology	0.933	0.934	0.956	**0.998**	0.991
Partiality	0.947	0.948	0.951	**0.983**	**0.983**
Triangulation	0.954	0.950	**0.955**	0.943	0.950

Figure 7: Retrieval results on the SHREC'10 Shape Google dataset

Method	Synthetic model	Scanned model
HAPT[33]	0.817	0.637
ISPM[34]	0.92	0.258
RBiHDM[35]	0.642	0.640
DBN[29]	0.842	0.304
VQ [21]	0.813	0.514
UDL [19]	0.842	0.523
SDL [19]	0.951	0.791
UPSDE	0.810	0.651
SPSDE	**0.970**	**0.811**

Figure 8: Retrieval results on the SHREC'14 Human dataset

Research Project: Learning barycentric representations of 3d shapes for sketch-based 3d shape retrieval

Description

Retrieving 3D shapes with sketches is a challenging problem since 2D sketches and 3D shapes are from two heterogeneous domains, which results in large discrepancy between them. In this project, we propose to learn barycenters of 2D projections of 3D shapes for sketch-based 3D shape retrieval. Specifically, we first use two deep convolutional neural networks (CNNs) to extract deep features of sketches and 2D projections of 3D shapes. For 3D shapes, we then compute the Wasserstein barycenters of deep features of multiple projections to form a barycentric representation. Finally, by constructing a metric network, a discriminative loss is formulated on the Wasserstein barycenters of 3D shapes and sketches in the deep feature space to learn discriminative and compact 3D shape and sketch features for retrieval. The proposed method is evaluated on the SHREC'13 and SHREC'14 sketch track benchmark datasets. Compared to the state-of-the-art methods, our proposed method can significantly improve the retrieval performance.

Method

First, we project 3D shapes to a set of rendered views. We employ two deep CNNs to extract the CNN features of sketches and 2D projections. The Wasserstein barycenters of CNN features of 2D projections can then be computed to characterize 3D shapes. Consequently, with a metric network, a discriminative loss is defined on the barycenters of 3D shapes and sketches in the feature space, which can maximize the within-class similarity and minimize the between-class similarity across the sketch and view domains, simultaneously.

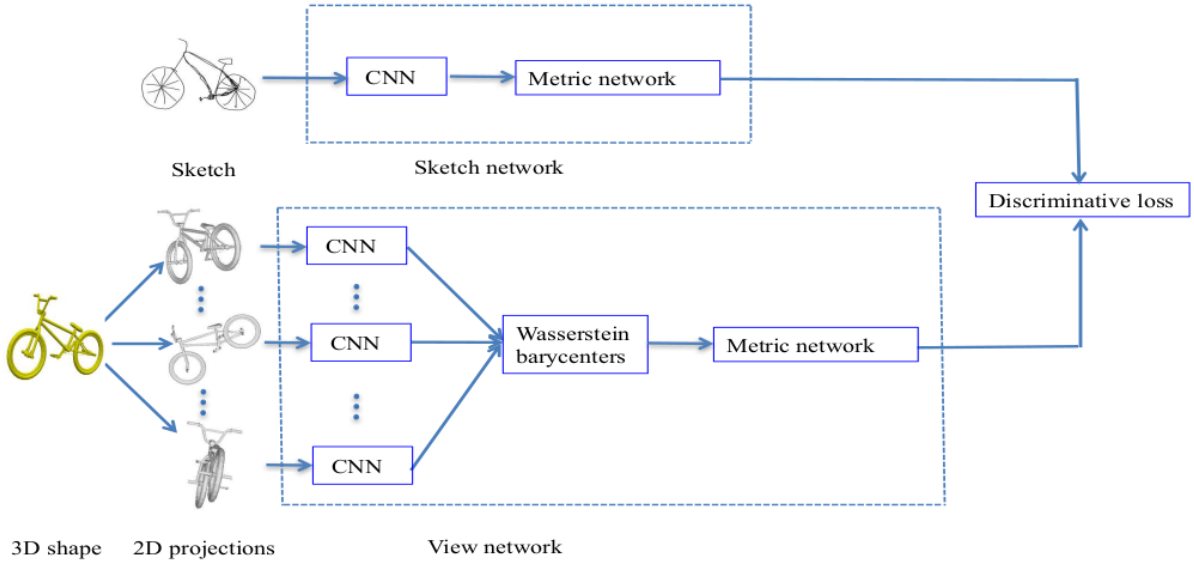


Figure 9: The cross-domain matching framework for sketch-based 3D shape retrieval. By rendering 3D shapes at multiple views, we extract deep CNN features of 2D projections. The Wasserstein barycenters of the deep CNN features are computed to represent 3D shapes. With the metric network of fully connected layers, we then formulate a discriminative loss to learn sketch and shape features for cross-domain retrieval.

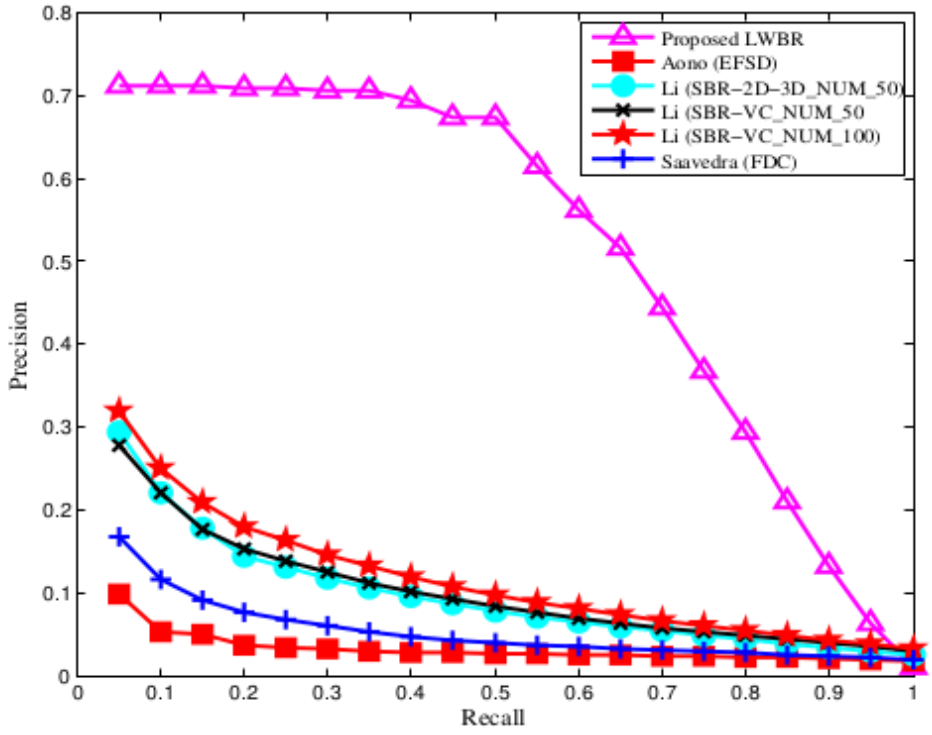


Figure 10: The precision-recall curves for the FDC, EFSD, SBR-VC and proposed LWBR methods on the SHREC'13 benchmark dataset.

Results

We first evaluate our learned Wasserstein barycentric representation method for sketch-based 3D shape retrieval, and then compare it to the state-of-the-art sketch-based 3D shape retrieval methods on two bench-mark datasets, i.e., SHREC'13 and SHREC'14 sketch track benchmark datasets.

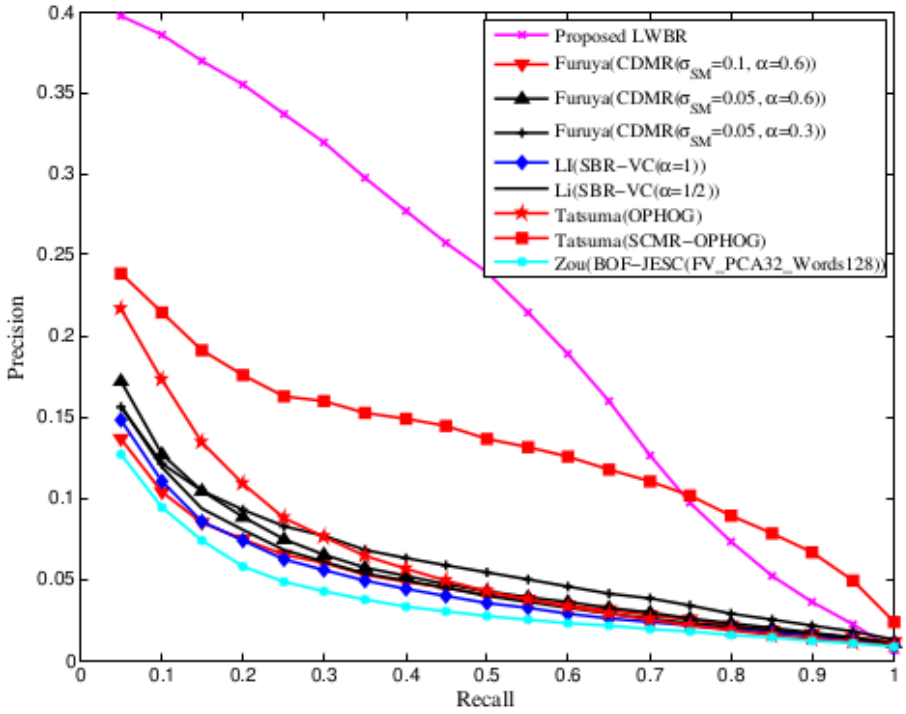


Figure 11: The precision-recall curves for the CDMR, SBR-VC, SCMR-OPHOG, BOF-JESC and proposed LWBR methods on the SHREC'14 benchmark dataset.

Research Project: Deepshape: Deep-learned shape descriptor for 3D shape retrieval

Description

Complex geometric variations of 3D models usually pose great challenges in 3D shape matching and retrieval. In this project, we propose a novel 3D shape feature learning method to extract high-level shape features that are insensitive to geometric deformations of shapes. Our method uses a discriminative deep auto-encoder to learn deformation-invariant shape features. The proposed method is evaluated on four benchmark datasets that contain 3D models with large geometric variations: McGill, SHREC'10 ShapeGoogle, SHREC'14 Human and SHREC'14 Large Scale Comprehensive Retrieval Track Benchmark datasets. Experimental results on the benchmark datasets demonstrate the effectiveness of the proposed method for 3D shape retrieval.

Method

We propose a novel discriminative auto-encoder to learn a shape descriptor for shape retrieval. In the proposed discriminative auto-encoder, we impose the Fisher discrimination criterion on the hidden layer so that the neurons in the hidden layer have small within-class scatter but large between-class scatter. To effectively represent shape, we use a multiscale shape distribution as input to the discriminative auto-encoder. We then train a discriminative auto-encoder at each scale and concatenate the outputs of the hidden layers from different scales as the shape descriptor.

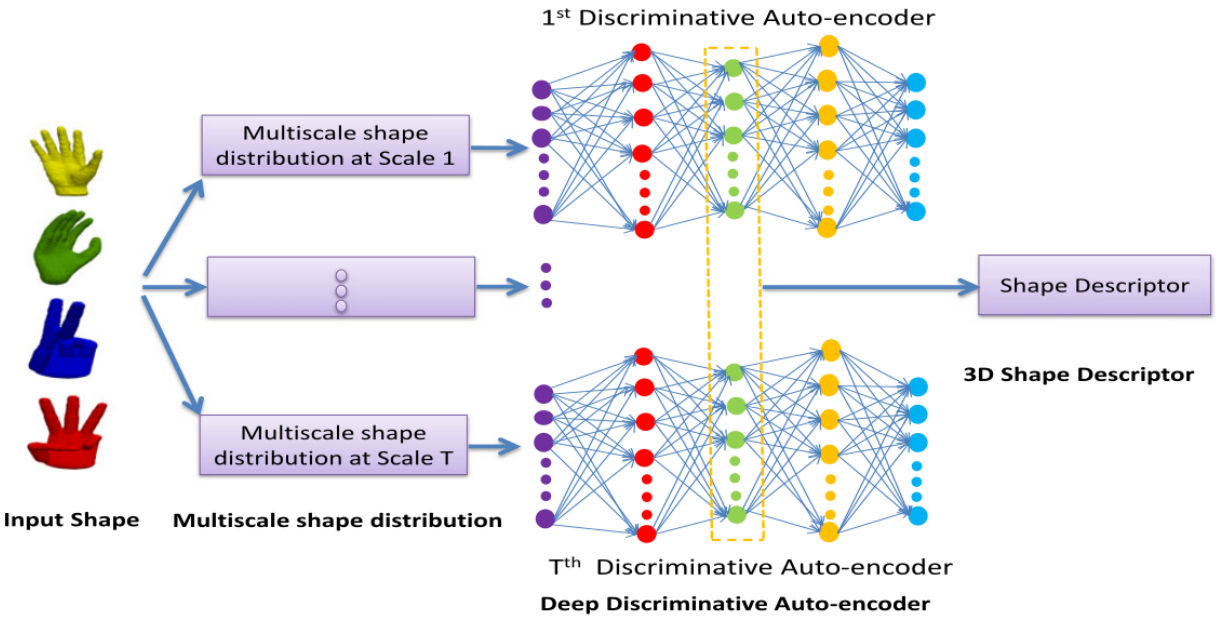


Figure 12: The framework of the proposed discriminative auto-encoder based shape descriptor.

Results

We evaluate our proposed shape descriptor and compare it to state-of-the-art methods on four benchmark datasets: McGill shape dataset, SHREC'10 Shape-Google dataset, SHREC'14 Human dataset and SHREC'14 Large Scale Comprehensive Retrieval Track Benchmark (SHREC'14 LSCRTB) dataset. Experimental results demonstrated the superior performance of our proposed descriptor.

Methods	NN	1-Tier	2-Tier	DCG
Covariance method [30]	0.977	0.732	0.818	0.937
Graph-based method [29]	0.976	0.741	**0.911**	0.933
PCA-based VLAT [28]	0.969	0.658	0.781	0.894
Hybrid BOW [27]	0.957	0.635	0.790	0.886
Hybrid 2D/3D [14]	0.925	0.557	0.698	0.850
CBoFHKS [15]	0.901	0.778	0.876	0.891
DASD	**0.988**	**0.782**	0.834	**0.955**

Figure 13: Retrieval results on the McGill dataset.

Transformation	VQ [13]	UDL [16]	SDL [16]	CBoFHKS [15]	DASD
Isometry	0.988	0.977	0.994	0.966	**0.998**
Topology	**1.000**	**1.000**	**1.000**	**1.000**	0.996
Isometry+Topology	0.933	0.934	0.956	0.915	**0.982**
Partiality	0.947	0.948	0.951	0.968	**0.973**
Triangulation	0.954	0.950	**0.955**	0.891	**0.955**

Figure 14: Retrieval results (mean average precision) on the SHREC'10 ShapeGoogle dataset

Method	Synthetic model	Scanned model
HAPT [31]	0.817	0.637
ISPM [32]	**0.92**	0.258
RBiHDM [33]	0.642	0.640
DBN [25]	0.842	0.304
VQ [13]	0.813	0.514
UDL [16]	0.842	0.523
DASD	0.823	**0.657**

Figure 15: Retrieval results (mean average precision) on the SHREC'14 Human dataset

Method	NN	1-Tier	2-Tier	E	DCG
CSLBP [26]	0.840	0.353	0.452	0.197	0.736
HSR-DE [26]	0.837	0.381	0.490	0.203	0.752
KVLAD [26]	0.605	0.413	0.546	0.214	0.746
DBNAA_DERE [26]	0.817	0.355	0.464	0.188	0.731
BF-DSIFT [26]	0.824	0.378	0.492	0.201	0.756
VM-1SIFT [26]	0.732	0.282	0.380	0.158	0.688
ZFDR [26]	0.838	0.386	0.501	0.209	0.757
DBSVC [26]	0.868	**0.438**	**0.563**	0.234	**0.790**
DASD	**0.897**	0.401	0.503	**0.243**	0.774

Figure 16: Retrieval results on the SHREC'14 LSCRTB dataset

Research Project: Deep multimetric learning for shape-based 3D model retrieval

Description

Recently, feature-learning-based 3D shape retrieval methods have been receiving more and more attention in the 3D shape analysis community. In this project, by exploring the nonlinearity of the deep neural network and the complementarity among multiple shape features, we propose a novel deep multimetric network for 3D shape retrieval. The developed multimetric network minimizes a discriminative loss function that, for each type of shape feature, the outputs of the network from the same class are encouraged to be as similar as possible and the outputs from different classes are encouraged to be as dissimilar as possible. Meanwhile, the Hilbert-Schmidt independence criterion is employed to enforce the outputs of different types of shape features to be as complementary as possible. Experimental results demonstrate that the proposed method can obtain better performance than the learned deep single metric and outperform the state-of-the-art 3D shape retrieval methods.

Method

We propose a novel deep multi-metric network to map multiple shape features to multiple non-linear feature spaces. It is expected that in the non-linear feature spaces the learned multiple deep shape features are discriminative and complementary so that they can characterize the manifold of 3D shapes well. Particularly, we construct a multi-metric network to jointly learn multiple non-linear metrics by minimizing the within-class variations of the learned shape features, maximizing the between-class variations of the learned shape features and employing the Hilbert-Schmidt independence criterion (HSIC) to minimize dependence of the learned multiple shape features, simultaneously. The learned distance metrics are fused as the similarity for shape retrieval.

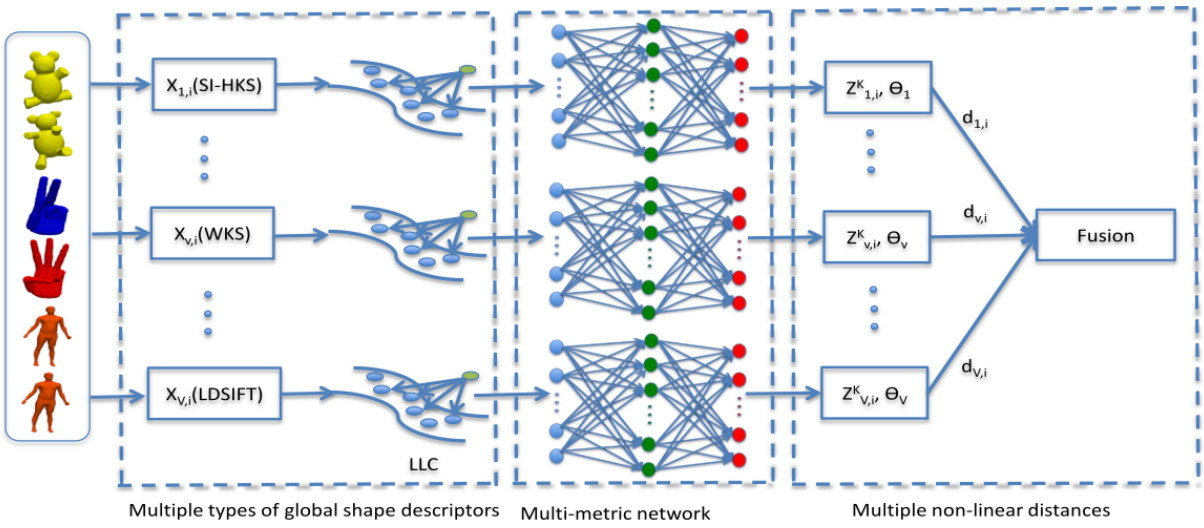


Figure 17: Proposed deep multimetric learning framework. Based on different types of 3D shape point signatures, the LLC method is employed to extract the global shape descriptors. The formed multiple types of shape features are then fed into the deep multimetric network so that the learned deep shape features are discriminative and complementary. The learned multiple nonlinear distance metrics are fused with the learned weights for retrieval.

Results

We first evaluate our proposed deep multi-metric learning model for retrieval, and then compare it with the state-of-the-art 3D shape retrieval methods on four benchmark datasets, i.e., Princeton Shape Benchmark (PSB), McGill shape dataset, SHREC'10 ShapeGoogle dataset and SHREC'14 Human dataset. Experimental results demonstrate that the proposed method can yield good retrieval performance.

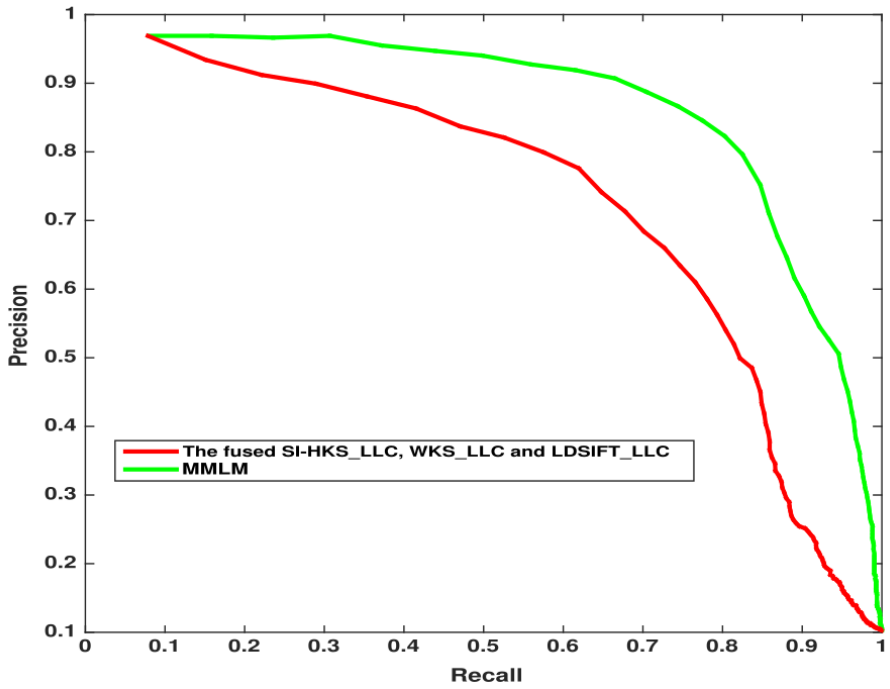


Figure 18: Precision–recall curves for the fused multiple shape features without using the deep multimetric network and the proposed MMLM method on the McGill shape dataset.

Transformation	VQ [9]	UDL [10]	SDL [10]	DA [11]	MMLM
Isometry	0.988	0.977	0.994	0.998	**1.000**
Topology	**1.000**	**1.000**	**1.000**	0.996	**1.000**
Isometry+Topology	0.933	0.934	0.956	0.982	**0.988**
Partiality	0.947	0.948	0.951	0.973	**0.985**
Triangulation	0.954	0.950	0.955	0.955	**0.963**

Figure 19: Retrieval results on the SHREC'10 ShapeGoogle dataset

Methods	NN	FT	ST	DCG
BoVF [7]	0.481	0.253	0.345	0.527
Hybrid 2D/3D [45]	0.742	0.473	0.606	–
CMVD [46]	0.566	0.286	0.367	0.564
3D CNN [24]	0.901	0.639	**0.849**	0.841
GIFT [23]	0.849	0.712	0.830	–
MMLM	**0.911**	**0.720**	0.831	**0.863**

Figure 20: Retrieval results on the PSB dataset

Method	Synthetic model	Scanned model
HAPT [52]	0.817	0.637
ISPM [51]	0.92	0.258
RBiHDM [50]	0.642	0.640
DBN [44]	0.842	0.304
VQ [9]	0.813	0.514
UDL [10]	0.842	0.523
SDL [10]	0.951	0.791
MMLM	**0.983**	**0.815**

Figure 21: Retrieval results on the McGill dataset

Research Project: Learned binary spectral shape descriptor for 3d shape correspondence

Description

Dense 3D shape correspondence is an important problem in computer vision and computer graphics. In this paper, we propose to learn a novel binary spectral shape descriptor with the deep neural network for 3D shape correspondence. First, we construct a neural network to form a nonlinear spectral representation to characterize the shape. Then, we train the constructed neural network by minimizing the errors between the outputs and their corresponding binary descriptors. Finally, we binarize the output of the neural network to form the binary spectral shape descriptor for shape correspondence. The proposed binary spectral shape descriptor is evaluated on the SCAPE and TOSCA 3D shape datasets for shape correspondence. The experimental results demonstrate the effectiveness of the proposed binary shape descriptor for the shape correspondence task.

Method

Based on the Laplace-beltrami operator, we propose to learn the local binary spectral shape descriptor for shape correspondence. First, we construct a neural network to compute the responses of the eigenvectors of the Laplace-beltrami operator of each point to characterize the shape. For each point on the shape, we define the positive/negative points that are in/out of the neighborhoods of the point on the shape and the corresponding point on the deformed shape, respectively. We then train the constructed neural network such that the errors between the real-valued outputs of the network and their binary outputs are as small as possible. Moreover, we encourage that the variations of the outputs associated with the pairs of positive points are as small as possible and the variations of the outputs associated with the pairs of negative points are as large as possible. Finally, we binarize the outputs of the network to form a binary spectral shape descriptor for correspondence.

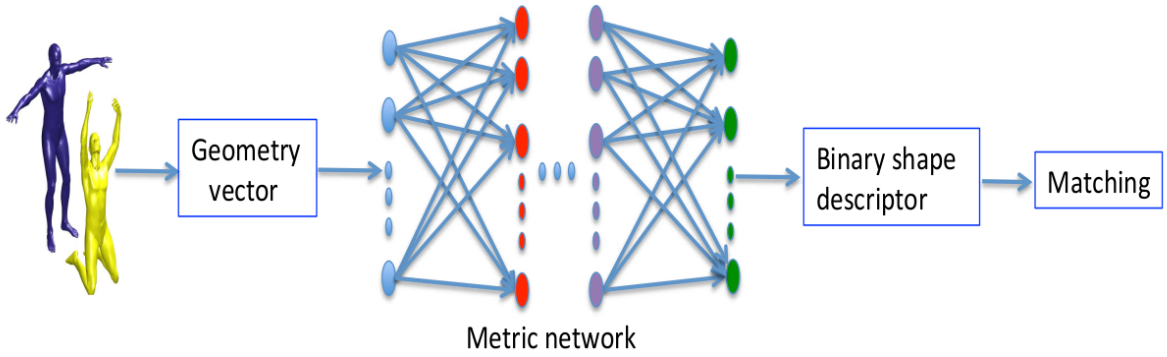


Figure 22: The shape matching framework with the proposed binary spectral shape descriptor. The geometry vectors of the points on a pair of shapes are used as the inputs to the metric network to form a nonlinear spectral representation. In the constructed metric network, the outputs of the pairs of positive points are required to be as similar as possible, the outputs of the pairs of negative points are required to be as dissimilar as possible, and the errors between the real-valued outputs of the network and their binary outputs are encouraged to be as small as possible.

Results

We test our proposed method on the SCAPE and TOSCA datasets. We evaluate our proposed learned binary spectral shape descriptor in terms of matching performance and computational time. Experimental results demonstrate its superior correspondence performance.

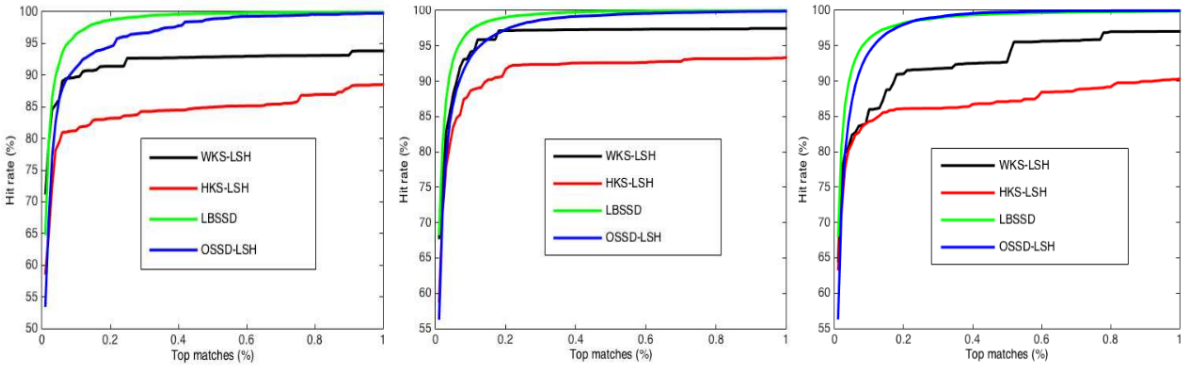


Figure 23: The CMC curves for HKS-LSH, WKS-LSH, OSSD-LSH and the proposed LBSSD on the SCAPE shape dataset: from left to right, 16 bits, 32 bits and 64 bits.

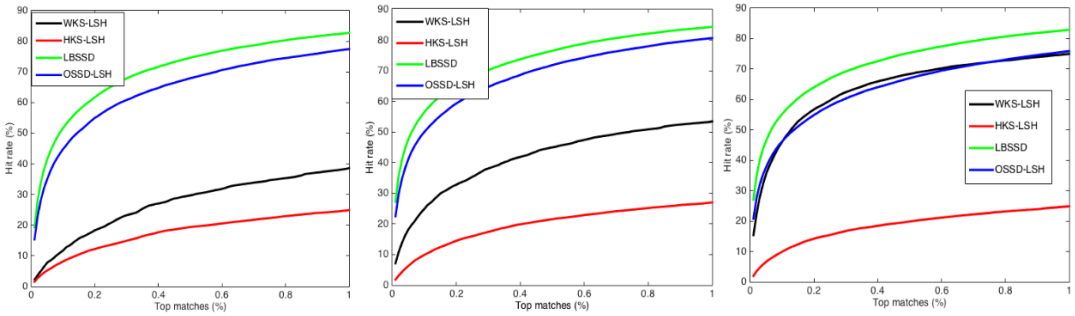
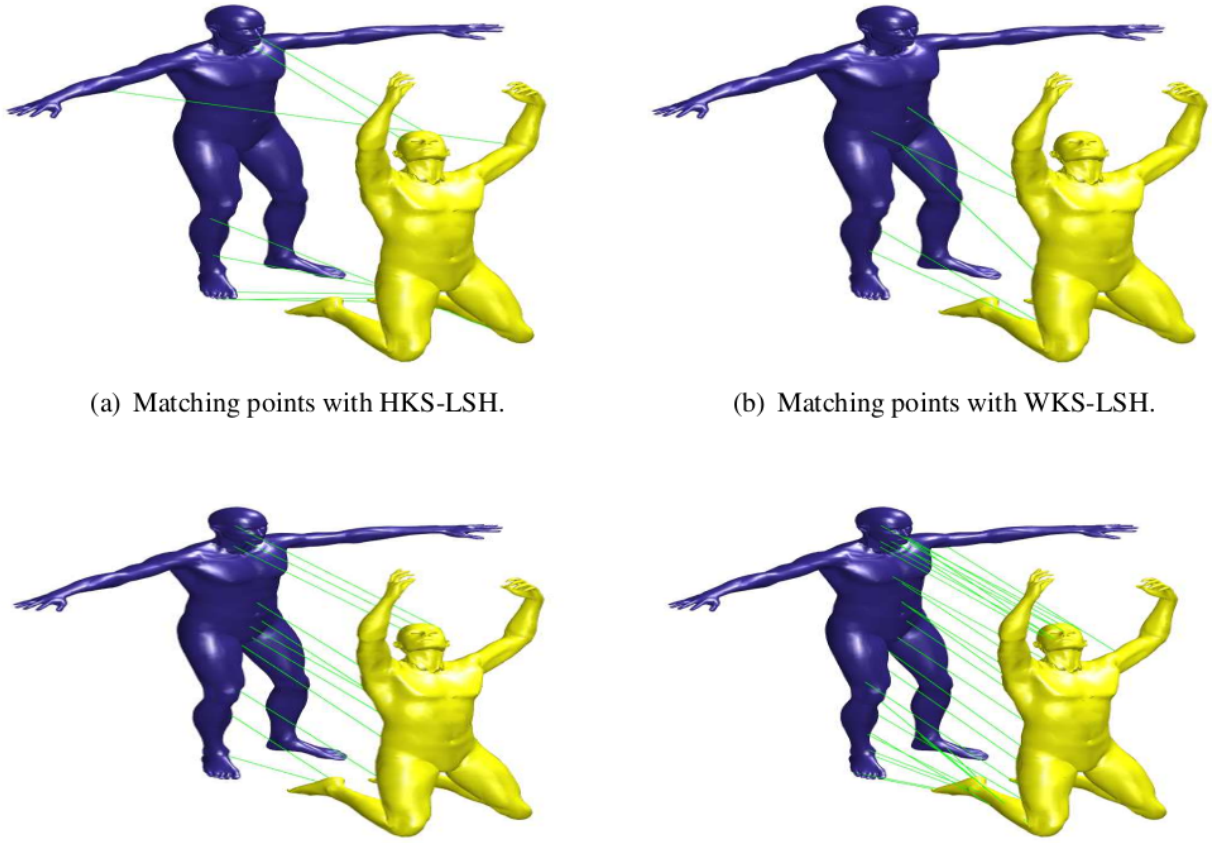


Figure 24: The CMC curves for HKS-LSH, WKS-LSH, OSSD-LSH and the proposed LBSSD on the TOSCA shape dataset: from left to right, 16 bits, 32 bits and 64 bits.



(a) Matching points with HKS-LSH.

(b) Matching points with WKS-LSH.

Figure 25: Matching points with the different 32-bit binary shape descriptors are shown with the geodesic distance distortion below 10% of the shape diameter among the sampled 100 points. (a): HKS-LSH, 8 matches; (b): WKS-LSH, 5 matches; (c): OSSD-LSH, 11 matches; (d): LBSSD, 28 matches.