International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 7, No. 6 (2014), pp. 335-344
http://dx.doi.org/10.14257/ijsip.2014.7.6.29

Sparse Polynomial Mapping for Manifold Learning

Ying Xia*, Qiang Lu and Hae-Young Bae

Research Center of Spatial Information System, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
xiaying@cqupt.edu.cn, csluqiang@hotmail.com, hybae@inha.ac.kr

Abstract

Manifold learning is an approach for nonlinear dimensionality reduction and has been a hot research topic in the field of computer science. A disadvantage of manifold learning methods is, however, that there are no explicit mappings from the high-dimensional feature space to the low-dimensional representation space. It restricts the application of manifold learning methods in many practical problems such as target detection and classification. Previously, some methods have been proposed to provide linear or nonlinear mappings for manifold learning methods. However, a disadvantage of all these methods is that the learned projective functions are combinations of all the original features, thus it is often difficult to interpret the results. Moreover, the dense projection matrices of these approaches lead to a high cost of computation and storage. In this paper, a sparse polynomial mapping approach is proposed for manifold learning. We first get the low-dimensional representations of the high-dimensional input data by using a manifold learning method, and then a l_1-based simplified polynomial regression is used to get a sparse polynomial mapping between the high-dimensional data and their low-dimensional representations. In particular, we apply this to the method of Laplacian eigenmap and derive a sparse nonlinear manifold learning algorithm, which is named sparse locality preserving polynomial embedding. Experimental results on real-world data show the effectiveness of our approach.

Keywords: *manifold learning, dimensionality reduction, sparse mapping*

1. Introduction

Large volumes of high-dimensional data are acquired in the data processing. The high dimensionality of data will cause the curse of dimensionality [1], thus effective dimensionality reduction methods are needed. Manifold learning has become a kind of important dimensionality reduction method because it could obtain a low intrinsic dimensionality of high-dimensional data [2-4]. In recent years, manifold learning methods, such as isometric mapping (ISOMAP) [3], locally linear embedding (LLE) [4], Laplacian eigenmap (LE) [5], maximum variance unfolding [6], diffusion maps [7], and Hessian eigenmap [8] achieved effective performance in experiments on both synthetic and real-world data. However, all these methods have a disadvantage that there are no explicit mappings from the high-dimensional input space to the low-dimensional output space. Thus the dimensionality of new-come high-dimensional data cannot be reduced quickly. This restricts the application of manifold learning methods in many practical problems such as target detection and classification.

In order to overcome the disadvantage of manifold learning, some linear or nonlinear methods were proposed to get approximate explicit mappings, such as locality preserving projections (LPP) [9], neighborhood preserving embedding (NPE) [10], and neighborhood

ISSN: 2005-4254 IJSIP

preserving polynomial embedding (NPPE) [11]. They preserve some intrinsic structures of high-dimensional data in the dimensionality reduction and achieve effective performance in experiments. However, a disadvantage of all these approaches is that the learned projective functions are combinations of all the original features, thus it is often difficult to interpret the results. Furthermore, these methods cannot get sparse projection matrices for explicit mappings, thus it will affect subsequent storage and computation performance. For example, if we use a linear mapping method such as LPP, it will require 20 million expensive floating-point multiplications to project a feature from 20,000 dimensionalities to 1,000 dimensionalities. Besides, storage of the projection matrix in floating-point format is 80 million. The high cost is unaffordable in many real scenarios such as mobile applications.

In recent years, some approaches have been proposed for learning sparse mappings. The sparse principal component analysis method is proposed for getting a sparse projection function in [12]. A spectral regression approach is proposed for sparse subspace learning in [13]. This approach casts the problem of learning the projective functions into a linear regression framework. It can get sparse mappings for subspace learning by using a l_1-based linear regression. These approaches get sparse mappings with the assumption that there exists a linear mapping between the high-dimensional data and their low-dimensional embedding. However, this linearity assumption may be too restrictive for manifold learning.

To address the above problems, we propose a two-step approach to obtain sparse polynomial mappings for manifold learning methods. In the first step, a manifold learning method can be applied to get low-dimensional representations of high-dimensional input data. In the second step, l_1-based simplified polynomial regression is applied to get a sparse polynomial mapping between the high-dimensional data and their low-dimensional representations. Through this approach, we can get a sparse polynomial mapping for a manifold learning method (*e.g.*, LE, ISOMAP, or LLE). In this paper, we concentrate on the LE manifold learning method and propose a sparse nonlinear manifold learning algorithm called sparse locality preserving polynomial embedding (SLPPE). Experiments on real-world data have been conducted to demonstrate the validity and effectiveness of the proposed approach.

2. Sparse Polynomial Mapping

In this section, we propose a novel approach to learn a sparse polynomial mapping which maps high-dimensional data samples to a low-dimensional subspace with a much lower computation and storage cost.

As shown in Figure 1, our approach can be divided into two steps. In the first step, we adopt a manifold learning method to get the low-dimensional representations of the high-dimensional input data. In the second step, we learn a sparse polynomial mapping which directly maps high-dimensional data to their low-dimensional representations by using the l_1-based simplified polynomial regression which is introduced below. In the training phase, the sparse polynomial mapping is learned by this two-step approach. In the testing phase, we get the low-dimensional representations by directly projecting the high-dimensional data using the learned sparse polynomial mapping.

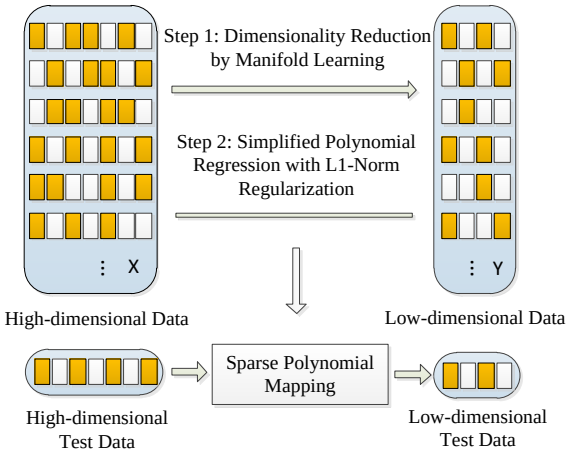


Figure 1. Sparse Polynomial Mapping Framework

We denote high-dimensional input data by $X=[x_1,x_2,...,x_N]$, $x_i \in R^m$, and their low-dimensional representations are given by the $N \times d$ matrix $Y=[y_1,y_2,...,y_d]$, where the low-dimensional representation $y^{(i)} \in R^d (d \ll m)$ is the transpose of the ith row of Y.

It has been proven in [14] that most manifold learning methods, including ISOMAP, LLE, and LE, can be cast into the framework of graph embedding. Through this framework, finding the low-dimensional representations of the high-dimensional input data is reduced to solving the following optimization problem:

$$\begin{aligned} &\min_{y^{(i)}} \ \frac{1}{2}\sum_{i,j=1}^{N} W_{ij} \| y^{(i)} - y^{(j)} \|_2^2 \\ &\text{s.t.} \ \sum_{i=1}^{N} D_i y^{(i)} y^{(i)^T} = I \end{aligned} \tag{1}$$

where $W_{ij}(i,j=1,2,...,N)$ are weights which can be defined by the input data samples, and $D_i=\sum_{j=1}^{N} W_{ij}$, and I is an identity matrix. With some simple algebraic calculation, (1) is equivalent to

$$\begin{aligned} &\min_{Y} \ \mathrm{tr}(Y^T(D-W)Y) \\ &\text{s.t.} \quad Y^T DY = I \end{aligned} \tag{2}$$

where D is a diagonal matrix whose diagonal entity is D_i, and W is a symmetrical matrix whose entity is W_{ij}. The optimal solutions $y_k \in R^N (k=1,2,...,d)$ are the eigenvectors of the following generalized eigenvalue problem corresponding to the d lowest eigenvalues:

$$(D-W)y = \lambda Dy. \tag{3}$$

Once $y_k(k=1,2,...,d)$ are computed, the dimensional representation $y^{(i)}=(y_1(i),...,y_d(i))^T$.

In the computing stage of a polynomial, as the polynomial degree increases, the computational complexity exponentially increases with the degree. Thus we define a simplified polynomial in the following by removing the crosswise items of the polynomial.

We assume that the *k*th component y_i^k of $y^{(i)}$ is a polynomial of degree p in x_i in the following manner:

$$y_i^k = v_k^T X_p^{(i)} \tag{4}$$

where v_k is the vector of polynomial coefficients, and $X_p^{(i)}$ is defined by

$$X_p^{(i)} = \begin{pmatrix} \overbrace{x_i \circ x_i \circ \cdots \circ x_i}^{p} \\ \vdots \\ x_i \circ x_i \\ x_i \end{pmatrix}$$

where $\circ$ stands for the Hadamard product which refers to entrywise matrix multiplication.

We can get the sparse polynomial mapping used in Figure 1 by the following simplified polynomial regression with the L_1 penalty. Due to the nature of the L_1 penalty, some coefficients will be shrunk to exact zero if the penalty parameter is large enough [15]. Thus it produces a sparse model, which is exactly what we want.

We compute the low-dimensional representations $y^{(i)}(i = 1, 2, ..., N)$ of the high-dimensional data samples $x_i(i = 1, 2, ..., N)$ by solving the generalized eigenvalue problem (3). With a L_1 penalty on v_k which is the vector of polynomial coefficients, we have

$$\min_{v_k} \sum_{k=1}^{d} \sum_{i=1}^{N} \{(y_i^k - v_k^T X_p^{(i)})^2 + \lambda \| v_k \|_1\} \tag{5}$$

which is called the l_1-based simplified polynomial regression, where λ is a penalty parameter, and $\| \cdot \|_1$ is the L_1 norm.

The optimization problem (5) can be solved by using an efficient coordinate descent algorithm [16] which is initialized by the value obtained in a previous iteration. Since the L_1 penalty is applied, we can get the sparse coefficient vector of the polynomial mapping. By tuning the parameter λ, we can control the sparsity of the vector of polynomial coefficients. The larger the value of λ is, the higher the sparsity of the coefficients. The sparse coefficients shall bring great convenience for subsequent storage and computation.

3. Sparse Locality Preserving Polynomial Embedding

It has been proven in [14] that most manifold learning methods, including LE, LLE, and ISOMAP, can be cast into the framework of graph embedding with different weights. Thus by using the sparse polynomial mapping approach proposed in Section 2 with different weights W_{ij}, $i, j = 1, 2, ..., N$, we can get different sparse polynomial mapping algorithms. In this section, we propose a new manifold learning algorithm with a sparse polynomial

mapping, named sparse locality preserving polynomial embedding (SLPPE), which is obtained by defining the weights $W_{ij}, i, j = 1,2,...,N$, in a way same to the LE method and combining them with the above sparse polynomial mapping. We also develop a supervised version of SLPPE in the end.

Given N data samples $x_1,...,x_N$ in R^m, we construct a weighted graph with N nodes, one for each sample, and a set of edges connecting neighboring samples. The sparse polynomial mapping is now provided by solving the generalized eigenvalue problem (3) and the optimization problem (5). The algorithmic procedure of SLPPE is formally stated below.

1. Step 1. **Constructing the adjacency graph**: We put an edge between nodes i and j if x_i and x_j are "close." There are two variations:

   (a) ε-neighborhoods (parameter $\varepsilon \in R$). Nodes i and j are connected by an edge if $\| x_i - x_j \|_2^2 < \varepsilon$.

   (b) k nearest neighbors (parameter $k \in N$). Nodes i and j are connected by an edge if i among k nearest neighbors of j or j is among k nearest neighbors of i.

2. Step 2. **Choosing the weights**: There are two variations as well.

   (a) Heat kernel (parameter $t \in R$).

   $W_{ij} = e^{-\frac{\|x_i - x_j\|_2^2}{t}}$ if nodes i and j are connected;

   $W_{ij} = 0$, otherwise.

   (b) Simple-minded (no parameters). $W_{ij} = 1$ if nodes i and j are connected by an edge, and $W_{ij} = 0$ otherwise.

3. Step 3. **Computing the low-dimensional representations**: The low-dimensional representations $y^{(i)} (i = 1,2,...,N)$ of the high-dimensional data samples can be computed by solving the generalized eigenvalue problem (3).

4. Step 4. **Solving the optimization problem**: The optimal solutions $v_k (k = 1,2,...,d)$ of the optimization problem (5) can be computed by the pathwise coordinate descent algorithm in [16].

For supervised learning, we can put an edge between nodes i and j if x_i and x_j are same class in Step 1 of the above SLPPE algorithm. This derives a supervised version of SLPPE, which is called supervised SLPPE. The supervised SLPPE may be more discriminate in a classification task.

4. Experiments

To evaluate the proposed sparse locality preserving polynomial embedding (SLPPE) algorithm and its supervised version, we conduct experiments on two face databases using them. The polynomial degree of these two algorithms is set as 2 in the experiments.

4.1. Datasets and Experimental Settings

Two face databases were used in the experiment. The first one is the AR database [17], and the second one is the PIE database [18]. The AR database contains over 4000 color images corresponding to 126 people's faces (70 men and 56 women). These face images include front view of faces with different expressions, illumination conditions, and occlusions (sun glasses and scarf). In the implementation, we use a subset of the AR database, which contains 1400 face images corresponding to 100 people (50 men and 50 women), where each person has 14 different images. The PIE database contains more than 40,000 facial images of 68 individuals. These images were acquired across different poses, with different expressions, and under variable illumination conditions. In the experiment, we choose near frontal poses and use all the images under different illumination conditions and facial expressions. Finally we get a subset including 2040 images with 30 images per individual. All the face images in the experiments are manually aligned and cropped. These images are resized to a size of 32×32 pixels, and the gray level values are rescaled to [0, 1].

For these two databases, we randomly choose half of the images per class for training (*i.e.*, 7 and 15 images per individual for AR and PIE, respectively), and the remaining for test. The training samples are used to learn mappings. By using these mappings, the testing images can be mapped into lower dimensional subspace. For simplicity, recognitions are carried out by using nearest neighbor classifier in the subspace. 5-fold cross validation has been used in the experiments for choosing the best penalty parameter. As a baseline, we give the recognition results of the classifier directly using the raw data without dimensionality reduction. In practice, 10 training/test splits are randomly generated and the average recognition accuracies over these splits are illustrated below.

4.2. Experimental Results

The proposed SLPPE is an unsupervised dimensionality reduction method. We compare it with LPP, NPPE, PCA, sparse PCA [12] and LE. The recognition rates are shown in Figure 2. The left subfigure is recognition rate on AR, and the right one is recognition rate on PIE. We demonstrate the best results together with the standard deviations obtained by them in Table 1. The sparsity of the projection matrix is also shown, and it is calculated as the ratio of the number of zero entries and the total number of entries. As can be seen, the performances of the proposed method overtake the compared methods.

The proposed supervised SLPPE is a supervised dimensionality reduction method. We compare it with LDA [19] and sparse LDA proposed in [13]. Note that, the upper bounds of the dimensions of LDA and sparse LDA are c-1 where c is the number of individuals [19, 13]. Thus the upper bounds of the dimensions of them on the selected datasets are 99 and 67, respectively. The recognition rates are shown in Figure 3. The left subfigure is recognition rate on AR, and the right one is recognition rate on PIE. We also show the best results together with the standard deviations obtained by them in Table 2. As can be seen, our proposed method surpasses the competitive methods.

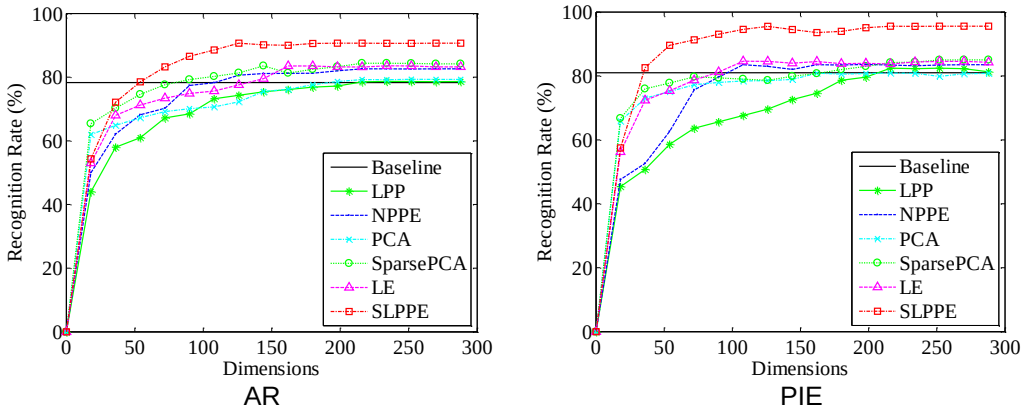


Figure 2. Recognition Rates of Unsupervised Methods

Table 1. Recognition Results of Unsupervised Methods on AR and PIE

Method	AR			PIE		
	accuracy (%)	dim	sparsity (%)	accuracy (%)	dim	sparsity (%)
Baseline	78.32±0.3	1024	--	80.98±0.7	1024	--
LPP	78.68±0.7	238	0	82.41±0.6	220	0
NPPE	82.62±0.6	185	0	83.53±0.6	115	0
PCA	79.36±0.4	246	0	80.88±0.7	230	0
SparsePCA	84.36±0.5	215	88.65±0.4	85.00±0.6	208	92.16±0.5
LE	83.54±0.3	162	0	84.61±0.5	106	0
SLPPE	90.67±0.2	136	90.68±0.3	95.59±0.7	92	94.53±0.6

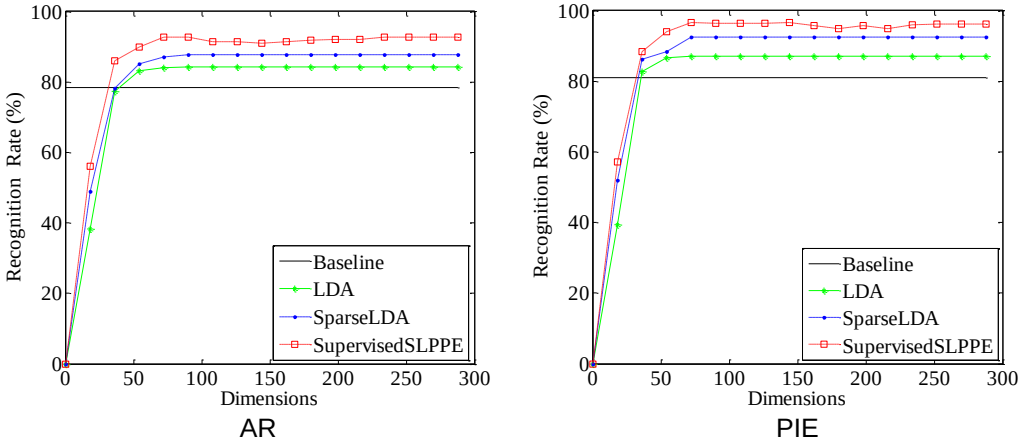


Figure 3. Recognition Rates of Supervised Methods

Table 2. Recognition Results of Supervised Methods on AR and PIE

Method	AR			PIE		
	accuracy (%)	dim	sparsity (%)	accuracy (%)	dim	sparsity (%)
Baseline	78.32 ± 0.3	1024	--	80.98 ± 0.7	1024	--
LDA	84.36 ± 0.2	99	0	87.16 ± 0.6	67	0
SparseLDA	87.65 ± 0.4	99	93.62 ± 0.6	92.50 ± 0.4	67	90.30 ± 0.6
SupervisedSLPPE	92.74 ± 0.6	96	92.35 ± 0.3	96.62 ± 0.5	62	90.60 ± 0.2

5. Conclusion

In this paper, a novel sparse polynomial mapping approach for manifold learning is proposed. Our approach is developed from the graph embedding [14] and the simplified polynomial regression with L_1-norm regularization. This approach can obtain a sparse polynomial mapping from a high-dimensional input space to its low-dimensional representation space. Therefore, it could reasonably interpret the results of the dimensionality reduction. Moreover, the dimension of a new-come data sample can be reduced quickly, and the cost of subsequent computation and storage can be decreased considerably.

Through this approach, we can get new sparse nonlinear manifold learning algorithms. In practice, we develop a new algorithm named sparse locality preserving polynomial embedding using the approach. Experimental results on face recognition show effectiveness of the proposed approach.

Acknowledgments

This work was financially supported by the Natural Science Foundation of China (41201378), Natural Science Foundation Project of Chongqing CSTC (cstc2012jjA40014), and Doctoral Startup Foundation of Chongqing University of Posts and Telecommunications (A2012-34).

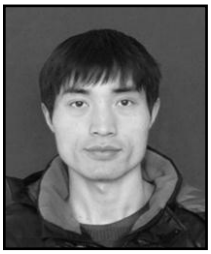
References

[1] R. E. Bellman, "Adaptive Control Processes: A Guided Tour", Princeton University Press, Princeton, **(1961)**.
[2] H. S. Seung and D. D. Lee, "The manifold ways of perception", Science, vol. 290, no. 5500, **(2000)**, pp. 2268–2269.
[3] J. B. Tenenbaum, V. de Silva and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", Science, vol. 290, no. 5500, **(2000)**, pp. 2319–2323.
[4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Science, vol. 290, no. 5500, **(2000)**, pp. 2323–2326.
[5] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation", Neural Computation, vol. 15, no. 6, **(2003)**, pp. 1373-1396.
[6] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming", International Journal of Computer Vision, vol. 70, no. 1, **(2006)**, pp. 77–90.
[7] R. R. Coifman and S. Lafon, "Diffusion maps", Applied and Computational Harmonic Analysis, vol. 21, no. 1, **(2006)**, pp. 5–30.
[8] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data", Proceedings of the National Academy of Sciences, vol. 100, no. 10, **(2003)**, pp. 5591–5596.
[9] X. He and P. Niyogi, "Locality preserving projections", in Advances in Neural Information Processing Systems, The MIT Press, Cambridge, vol. 16, **(2004)**, pp. 37-45.
[10] X. He, D. Cai, S. Yan and H. Zhang, "Neighborhood Preserving Embedding", Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing China, **(2005)**, pp. 1208-1213.
[11] H. Qiao, P. Zhang, D. Wang and B. Zhang, "An Explicit Nonlinear Mapping for Manifold Learning", IEEE Transactions on Cybernetics, vol. 43, no. 1, **(2013)**, pp. 51-63.

[12] H. Zhou, T. Hastie, and R. Tibshirani, "Sparse principle component analysis", Technical report, Statistics Department, Stanford University, **(2004)**.
[13] D. Cai, X. He and J. Han, "Spectral regression: A unified approach for sparse subspace learning", Proceedings of the 7th IEEE International Conference on Data Mining, **(2007)**, pp. 73-82.
[14] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, **(2007)**, pp. 40-51.
[15] R. Tibshirani, "Regression shrinkage and selection via the lasso", Journal of the Royal Statistical Society, Series B (Methodological), **(1996)**, pp. 267-288.
[16] J. Friedman, T. Hastie and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent", Journal of statistical software, vol.33, no. 1, **(2010)**, pp. 1-22.
[17] A. Martinez and R. Benavente, "The AR face database", CVC Tech. Report #24, **(1998)**.
[18] T. Sim, S. Baker and M. Bsat, "The CMU Pose, Illumination, and Expression Database," IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 25, no. 12, **(2003)**, pp. 1615-1618.
[19] P. N. Belhumeur, J. P. Hepanha and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, **(1997)**, pp. 711–720.

Authors

Ying Xia, she received the Ph.D. degree in computer science and technology from the Southwest Jiaotong University, China, in 2012. Currently, she is a professor at Chongqing University of Posts and Telecommunications, China. Her research interests include spatial database, GIS, and Cross-media retrieval.



Qiang Lu, he received the bachelor degree in computer science and technology from Zhengzhou University in 2012. Currently, he is studying for the master degree at Chongqing University of Posts and Telecommunications. His current research interests include machine learning and Cross-media retrieval.

Hae-Young Bae, he is tenured full professor of Inha University of Korea, and he is honorary professor of the Chongqing University of Posts and Telecommunications of China. His research area mainly includes database and spatial information processing.