Appl. Math. Inf. Sci. **7**, No. 3, 1065-1075 (2013)

1065

Semi-supervised Sparsity Pairwise Constraint Preserving Projections based on GA

Mingming Qi[1,2] *and Yang Xiang*[1]

[1]Department of Computer Science and Technology, Tongji University, Shanghai 201804, P. R. China
[2]School of Yuanpei, Shaoxing University, Shaoxing 312000, P. R. China

Abstract: The deficiency of the ability for preserving global geometric structure information of data is the main problem of existing semi-supervised dimensionality reduction with pairwise constraints. A dimensionality reduction algorithm called Semi-supervised Sparsity Pairwise Constraint Preserving Projections based on Genetic Algorithm (SSPCPPGA) is proposed. On the one hand, the algorithm fuses unsupervised sparse reconstruction feature information and supervised pairwise constraint feature information in the process of dimensionality reduction, preserving geometric structure in samples and constraint relation of samples simultaneously. On the other hand , the algorithm introduces the genetic algorithm to set automatically the weighted trade-off parameter for full fusion. Experiments operated on real world datasets show, in contrast to the existing typical semi-supervised dimensionality reduction algorithms with pairwise constraints and other semi-supervised dimensionality reduction algorithms on sparse representation, the proposed algorithm is more efficient.

Keywords: Semi-supervised Dimensionality Reduction, Pairwise Constraints, Sparsity Preserving Projections, Information Fusion, Genetic Algorithm, Trade-off Parameter

1. Introduction

The past few years have witnessed more and more research results on dimensionality reduction algorithms with pairwise constraints owing to great convenient in obtaining them and more supervised discriminant information in them [1,2]. Bar-Hillel et al. [3] proposed Constraints Fish Discriminant Analysis (CFDA), which is the pre-treatment step before relevant component analysis. CFDA only deal with the must-link constraints and ignore cannot-link constraints. Tang et al. [4] proposed Pairwise Constraints-guided Feature Projection (PCFP) for dimensionality reduction, which exploits both must-link constraints and cannot-link constraints but ignores unlabeled data. Zhang et al. [5] proposed Semi-Supervised Dimensionality Reduction (SSDR), which exploits both cannot-link and must-link constraints together with variance information in unlabeled data. However SSDR only deals with linear data. On the basic of LPP, Cevikalp et al. [6] proposed Constrained Locality Preserving Projections (CLPP), which finds neighborhood points to create a weighted neighborhood graph and make use of the constraints to modify the neighborhood relations and weight matrix to reflect this weak form of supervision. Although CLPP is fit for nonlinear data, the algorithm is still sensitive to noise and parameters. YU et al. [7] proposed Robust LPP (RLPP) with pairwise constraints based on robust path based similarity for overcoming these problems. Wei et al. [8] proposed Neighborhood Preserving based Semi-supervised Dimensionality Reduction (NPSSDR). The algorithm not only preserves the must-link and cannot-link constraints but also preserves the local structure of input data in the low dimensional embedding subspace by the regularization way, which makes NPSSDR easy to get into collapse in local structure. Chen et al. [9] proposed Semi-supervised Non-negative Matrix Factorization (SS-NMF) based on pairwise constraints on a few of documents. Peng et al. [10] proposed Semi-supervised Canonical Correlation Analysis Algorithm (Semi-CCA) which uses supervision information in the form of pairwise constraints in canonical correlation analysis. Davidson et al. [11] proposed Graph-driven Constrained Dimension

---

* Corresponding author e-mail: webqmm1974@sohu.com

Reduction via Linear Projection (GCDR-LP) that given a weighted graph attempts to find series of dimensions that are linear combinations of the old dimensions.

Recently, sparse representations attain more and more attention and are successfully applied in object detection and classification [12, 13, 14, 15,16]. Researches show, classification based on sparse representations has good robustness on face datasets with deformities, dressings and shelters. At present sparse learning has extended to dimensionality reduction, including Sparse Principal Component Analysis (SPCA) [17] Principal Component Analysis with Weighted Sparsity Constraint (PCAWSC) [18] and Sparsity Preserving Projections (SPP) [19]. SPP is a new unsupervised dimensionality reduction algorithm, which preserves sparse reconstruction relations of high-dimension data to low-dimension data. In contrast to other unsupervised dimensionality reduction algorithms, SPP preserves global geometry structure information contained in sparse reconstructions and is available of power discriminant analysis. However, SPP is sensitive to variations in whole pattern of data owing to deficiencies of supervised information. Gu et al.[20] proposed Discriminative Sparsity Preserving Projections(DSPP). DSPP provides an explicit feature mapping by fitting the prior low-dimensional representations which are generated randomly by using the labels of the labeled data points and, meanwhile, setting the smoothness regularization term to measure the loss of the mapping in preserving the sparse structure of data, so DSPP has highly discriminative ability.

Motivated by above analyses, a Semi-supervised Sparse Pairwise Constraint Preserving Projections based on Genetic Algorithm (SSPCPPGA) is proposed in the paper. The algorithm firstly exact respectively unsupervised information of sparse reconstruction and supervised information of pairwise constraint, then fuse two kinds of information by the linear weighted way and seek the optimized weighted trade-off parameter through the genetic algorithm. Finally projections are gotten to preserve fused information. The projected low-dimensional data not only preserve global geometric structure feature information contained in sparse reconstructions but also preserve pairwise constraint feature information. Experimental results on AR, Yale and UMIST show that our proposed algorithm improves the accuracy and stability of classification rate based on the shortest Euclidean distance, in contrast to other typical semi-supervised dimensionality reduction algorithm with pairwise constraints and other semi-supervised dimensionality reduction algorithms on sparse representation.

Several characteristics of our presented algorithm are listed as follows:

(1) PCFP and SPP have their own advantages and disadvantages. The fused algorithm inherits special character PCFP and SPP and overcomes their disadvantages.

(2)Sparse reconstruction information and pairwise constraint information vary greatly in different datasets, which is the reason that concrete weighted trade-off parameter is different in the process of linear fusion. The algorithm introduces the genetic algorithm to set automatically the weighted trade-off parameter value in order to get more performance.

The rest of the paper is organized as follows: Section 2 reviews PCFP and SPP. SSPCPGA is introduced in Section 3. In Section 4, we compare proposed SSPCPGA with PCFP, SPP, RLPP, SSDR, NPSSDR and DSPP. The experimental results are presented. Finally, we provide some concluding remarks and future work in Section 5.

2. Related works

In this section, we review related works, including PCFP in 2.1 and SPP in 2.2.

2.1. PCFP

Given training samples $X = \{x_1, x_2, x_3, \cdots, x_n\}$, must-link set $ML = \{(x_i, x_j) | x_i \text{ and } x_j \text{ are the same class}\}$ and cannot-link set $CL = \{(x_i, x_j) | x_i \text{ and } x_j \text{ are not the same class}\}$. PCFP aims to find an optimal projection matrix T that maximizes the following function[4]:

$$\max_{T} \left[\sum_{(x_i, x_j) \in CL} \| T^T x_i - T^T x_j \|^2 - \sum_{(x_i, x_j) \in ML} \| T^T x_i - T^T x_j \|^2 \right] \quad (1)$$

$$s.t. \quad T^T T = I$$

Where $\| \cdot \|^2$ denotes the l_2 norm.

Eq.(1) may be understood in such a sentence that two samples of must-link set in high-dimensional data space should be more near in low-dimensional data space and two samples of cannot-link set in high-dimensional data space should be more further in low-dimensional data space.

2.2. SPP

Given training samples $X = \{x_1, x_2, x_3, \cdots, x_n\} \in R^{d \times n}$, the goal of sparse representation is to reconstruct each sample with else samples. A sparse reconstructive weight vector s_i for x_i is described as follows:
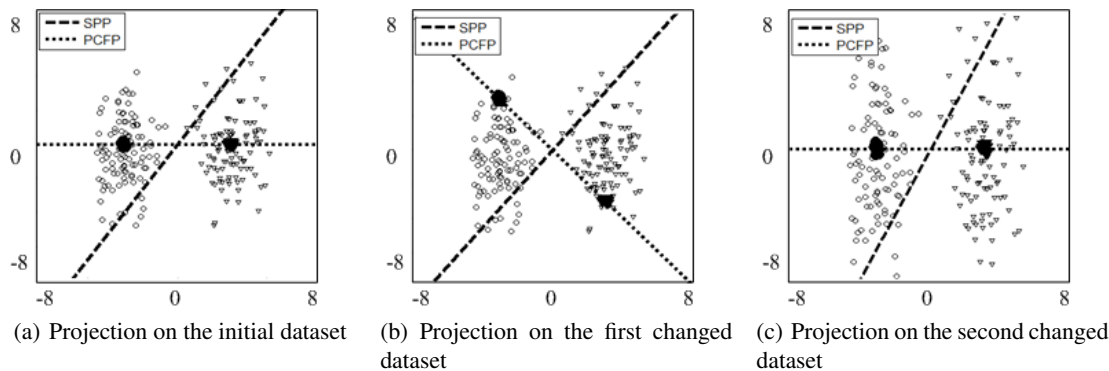
$$\min_{S_i} \| S_i \|_1$$

$$s.t. \quad x_i = X S_i \quad (2)$$

(a) Projection on the initial dataset

(b) Projection on the first changed dataset

(c) Projection on the second changed dataset

Figure 1 projected one-dimensional subspace

where S_{ij} denotes the contribution of each x_j to reconstructing x_i . $\| \cdot \|_1$ denotes the l_1 norm .

Sparse reconstruction is to seek a sparse reconstructive weight vector s_i for each x_i ,which is described as follows:

$$\min_{S_i} \| S_i \|_1$$
$$s.t. \quad x_i = XS_i \qquad (3)$$
$$1 = 1^T S_i$$

where $\|S_i\|_1$ denotes the l_1 normal of S_i , $S_i = [S_{i1}, \cdots, S_{ii-1}, 0, S_{ii+1}, \cdots, S_{in}]^T \in R^d$ is a vector in which S_{ij} denotes the contribution of each x_j to reconstructing x_i, and $1 \in R^n$ is a vector of all ones.

$$x_i = S_{i1}x_1 + \cdots + S_{ii-1}x_{i-1} + S_{ii+1}x_{i+1}$$
$$+ \cdots + S_{in}x_n \qquad (4)$$

The sparse reconstruction matrix $S = [S_1, S_2, \cdots, S_n]^T$ is attained through computing S_i. Sparse reconstructive wei- ghts have intrinsic geometric properties of the data.

The goal of SPP is to preserve sparse reconstruction relation of high-dimensional data space into low-dimensional data space. Given the projection matrix T, the objective function of SPP is gotten as follows [19]:

$$\min_T \sum_{i=1}^{n} \|T^T x_i - T^T XS_i\|^2$$
$$s.t. \quad T^T XX^T T = I \qquad (5)$$

Eq.(5) can be further transformed to

$$\max_T [T^T X(S + S^T - S^T S)X^T T]$$
$$s.t. \quad T^T XX^T T = I \qquad (6)$$

3. Semi-supervised Sparse Pairwise Constraint Preserving Projections based on GA (SSPCPPGA)

In this section, we first introduce the basic idea of our algorithm and then the objective function is gotten; finally we give steps of the algorithm.

3.1. Basic idea

In order to describe disadvantages of PCFP and SPP, we create a two-dimensional dataset that contains two classes represented by dot points and triangle points. The number of two kinds of samples is same.Two kinds of one-dimensional subspaces on the dataset and the changed dataset are gotten through SPP and PCFP. Concrete results are shown in Figure 1.

In Figure 1, solid dot points denote labeled data of the first class and solid triangle points denote labeled data of the second class. The number of two kinds of labeled samples is same. pairwise constraint sets are composed by these labeled samples. Figure 1(a) shows projection results of SPP and PCFP. When labeled samples happen to change in Figure 1(b),the projection result of SPP does not change while that of PCFP becomes poor. In contrast to data in Figure 1(a), data in Figure 1(c) are enlarged to twice as much in the vertical scalar to change global distribution patterns of the structure of data, which weak the projection result of SPP and don't affect that of PCFP. Figure 1 demonstrates that the performance of supervised PCFP is sensitive to pairwise constraint sets instead of global distribution patterns of the structure of training samples and SPP is contrary. Therefore fusing feature information of PCFP and SPP in the process of dimensionality reduction is a feasible way for overcoming disadvantages of them.

3.2. Objective function

According to the information level, information fusions are divided into the data level, the feature level and the decision-making level. Information fusion based on the feature level may make sure relations among different feature information [21]. Fusion of supervised feature information and unsupervised feature information based on the linear weighted way has proved to be an efficient fusion way of semi-supervised dimensionality reduction [22, 23].
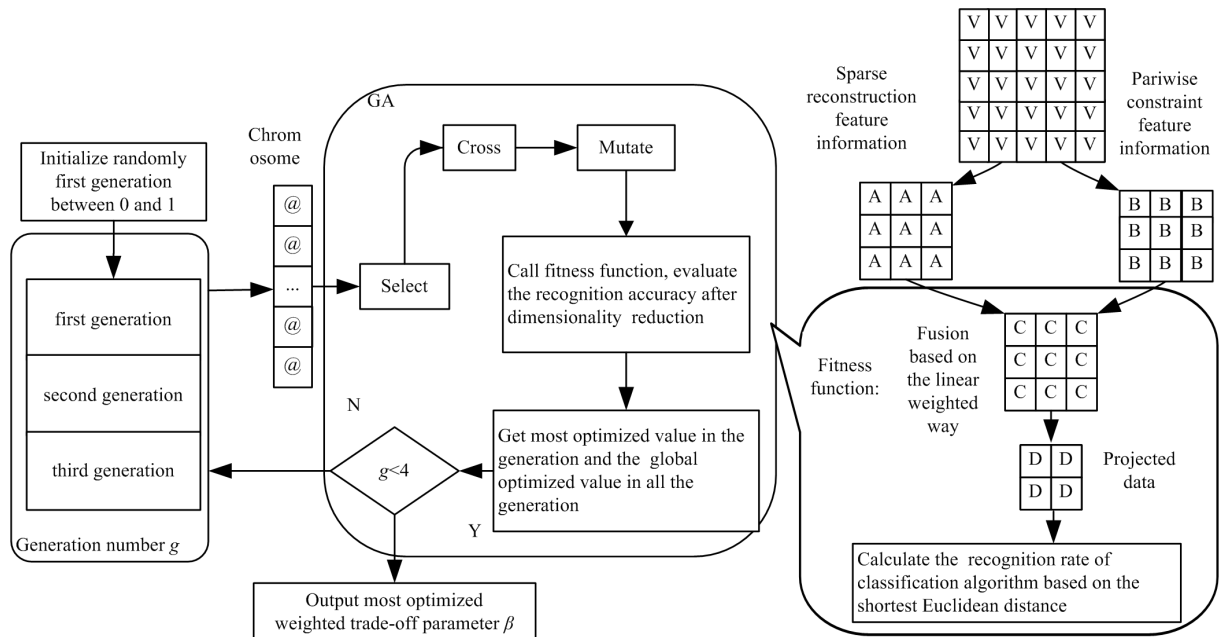
Figure 2 The framework of setting the linear weighted parameter β based on GA

SPCPPGA aims to find a projection T to preserve simultaneously pairwise constraint feature information and sparse reconstruction feature information by making use of the linear weighted way to fuse two kinds of information.

Eq.(1) and Eq.(6) can respectively further be transformed to an equivalent maximization problem as follows:

$$\max_{T} \left[\frac{\sum\limits_{(x_i,x_j)\in CL} \|(T^T x_i - T^T x_j)\|^2}{T^T T} - \frac{\sum\limits_{(x_i,x_j)\in ML} \|(T^T x_i - T^T x_j)\|^2}{T^T T} \right] \quad (7)$$

$$\max_{T} \frac{[T^T X(S + S^T - S^T S)X^T T]}{T^T X X^T T} \quad (8)$$

We infuse Eq.(7) and Eq.(8) to get the objective function of SPCPPGA through the linear weighted way as follows:

$$\max_{T} \frac{T^T[\beta X S_\alpha X^T + (1-\beta)P_\alpha]T}{T^T[\beta X X^T + (1-\beta)I]T} \quad (9)$$

where

$$S_\alpha = S + S^T - S^T S \quad (10)$$

$$P_\alpha = \sum_{(x_i,x_j)\in CL} (x_i - x_j)(x_i - x_j)^T \\ - \sum_{(x_i,x_j)\in ML} (x_i - x_j)(x_i - x_j)^T \quad (11)$$

The weighted trade-off parameter β denotes the contribution of sparse reconstruction feature information to SPCPPGA and $1 - \beta$ denotes the contribution of pairwise constraint feature information to SPCPPGA.

3.3. Optimize the weighted trade-off parameter based on Genetic Algorithm

According to above analyses, the weighted trade-off parameter β plays an important role in SPCPPGA. The necessities of the optimization for the weighted trade-off parameter β are embodied as follows:

(1) Sparse reconstruction feature information is the model of information decomposition and information reconstruction that are based on over-complete dictionary while pairwise constraint feature information is side information. Their natures are great different owing to their basic math ideas and representations.

(2) Two kinds of feature information from different training samples and different dimensions of projected subspaces are very different , determining different weighted trade-off parameters in the process of fusing them.

Genetic Algorithm (GA) is the optimal solution by simulating the natural evolutionary process search [24]. GA has power of inherent implicated parallel and better ability of global optimization, obtaining automatically and guiding to optimize the searching space through adjusting adaptively searching directions [25, 26]. Therefore, GA is introduced to guide to set the linear weighted trade-off parameter . Figure2 gives us the

framework of setting the linear weighted trade-off parameter based on GA.

The main processes of optimizing the weighted trade-off parameter β in Figure 2 are as follows:

(1) Create randomly a certain number of populations represented by the binary chromosome and initialize their values between 0 and 1.

(2) Achieve basic operation on these populations, including selecting, crossing and mutating. Selecting operation is achieved by eliminating the first half of individuals in ascending order about the fitness value and creating randomly some individuals for supplement.

(3) Get most optimized value in the generation and the global optimized value in the total generation.

(4) If the number of GA generation $g < 4$, then jump to (1), otherwise output the global optimized value in all the generation.

3.4. Algorithm steps

Input: training samples $X = \{x_1, x_2, x_3, \cdots, x_n\}$ and $x_i \in R^{w \times h}$,
$ML = \{(x_i, x_j) | x_i \text{ and } x_j \text{ are in the same class}\}$,
$CL = \{(x_i, x_j) | x_i \text{ and } x_j \text{ are not in the same class}\}$, The size of the populations p, the length of the chromosome c, the crossover probability pc and mutation probability pm.
Output: the project matrix $T^{d \times r}(r < d)$.
Steps:

(1) construct S using to Eq.(3).

(2) construct S_α using Eq.(10).

(3) construct P_α using Eq.(11).

(4) The simplest Nearest Neighbor Classifier is adopted in the fitness function. According to the model of Figure2, calculate the optimized weighted parameters with different datasets and different dimensions.

(5) Transform Eq.(9) to

$$(\beta X S_\alpha X^T + (1-\beta)P_\alpha)\varphi = \lambda(\beta X X^T + (1-\beta)I)\varphi$$

and calculate the projections matrix $T^{d \times r}(r < d)$.

4. Experiments and analyses

In this section, we firstly introduce experimental datasets, which is followed by experimental settings. Finally, we give experimental results and detail analyses on the classification performance of SPCPPGA and the adaptivity of the weighted trade-off parameter β.

4.1. Experimental datasets

(1) Yale contains 165 face images of 15 individuals. There are 11 images per subject, and these 11 images are

respectively, under the following different facial expression or configuration: center-light, wearing glasses, happy, left-light, wearing no glasses, normal, right-light, sad, sleepy, surprised and wink. In our experiment, we resize theses face images of Yale to 30×30 pixels.

(2) AR consists of over 4000 face images of 126 individuals. For each individual, 26 pictures were taken in two sessions that separated by two weeks and each section contains 13 images, which include front view of faces with different expressions, illuminations and occlusions. In our experiment, we resize theses face images of AR to 30×30 pixels.

(3) UMIST is composed of 564 face images of 20 individuals. UMIST face images cover the front face of different side. In our experiment, we resize theses face images of UMIST to 34×28 pixels.

A group of face samples on Yale, AR and UMIST are shown in Figure3.

(a) A group of face samples on Yale

(b) A group of face samples on AR

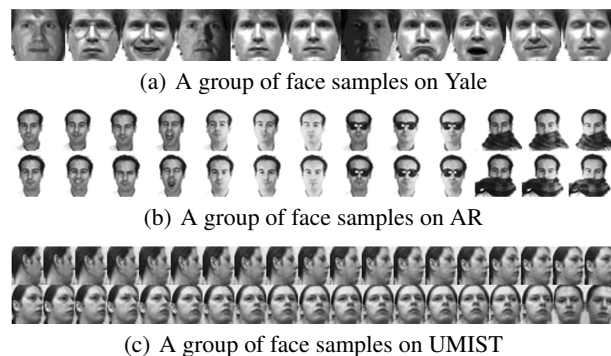(c) A group of face samples on UMIST

**Figure 3** A group of face samples on various datasets

4.2. Experimental settings

SSDR[5], RLPP[7], NPSSDR[8] and DSPP [20] are three typical semi-supervised dimensionality reduction algorithms with pairwise constraints and are compared with our proposed algorithm for testing classification performances. In addition, SPP and PCFP are also added compared algorithms in order to verify the integration performance. Table 1 shows specific parameter settings in various algorithms.

In Table 1, p denotes the population size, c denotes the individual chromosome length, pc denotes the crossover probability and pm denotes the mutation probability in SSP CPPGA. Usually the crossover probability parameter pc and the mutation probability parameter pm of GA are set respectively to 0.6 and 0.1. For expression convenience, SSPCPPGA with the different population size and the different individual chromosome length is denoted by SSPC PPGA (p, c).
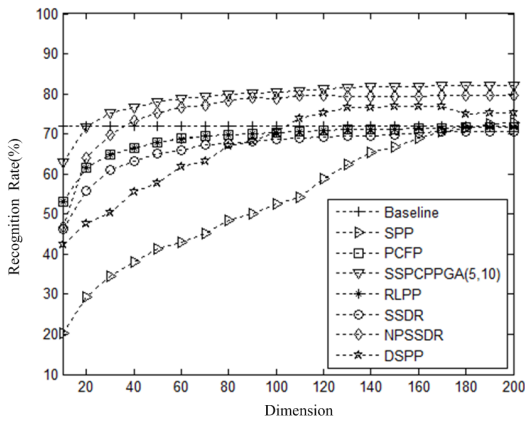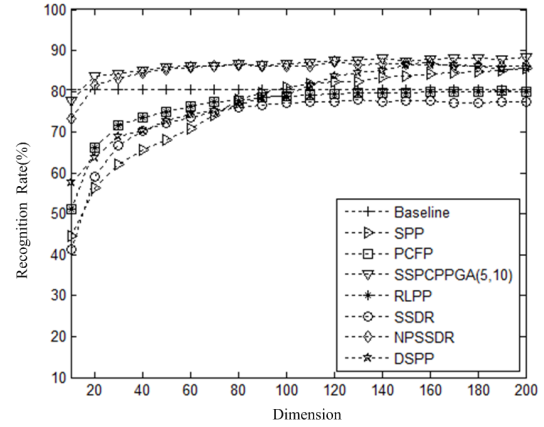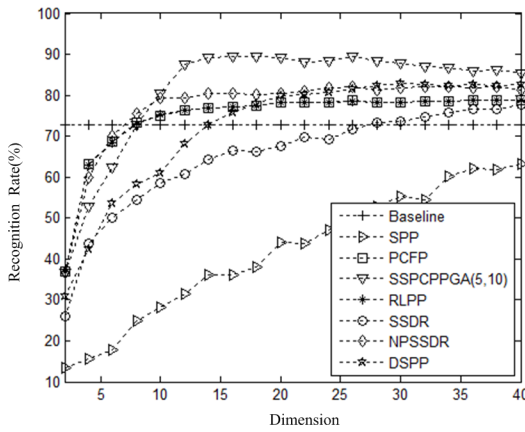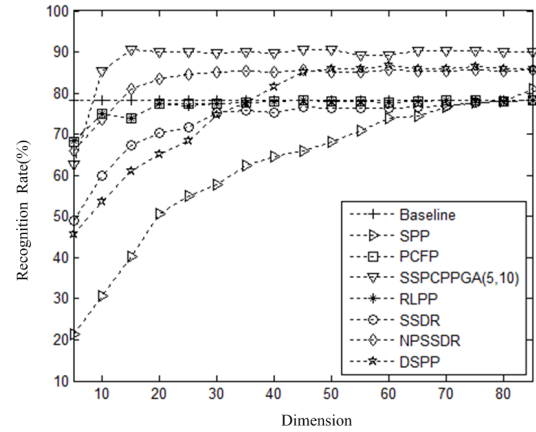
(a) *L*=5 and *D*=10   (b) *L*=10 and *D*=10

Figure 4 Experimental results of the classification performance with *L* and *D* on AR
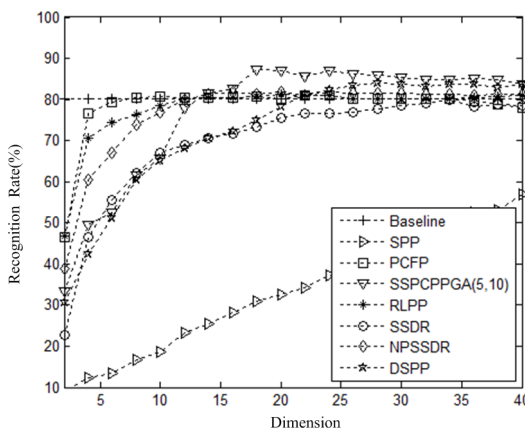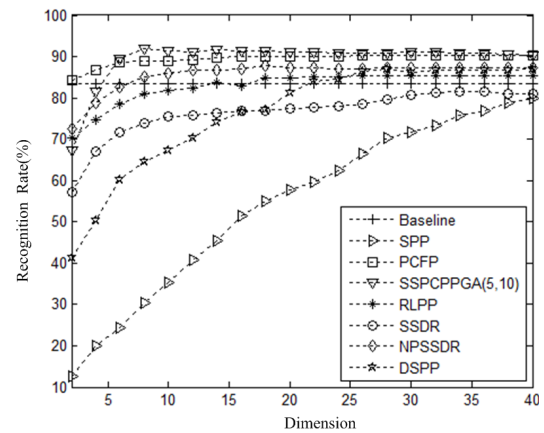

(a) *L*=3 and *D*=2   (b) *L*=6 and *D*=5

Figure 5 Experimental results of the classification performance with *L* and *D* on Yale

(a) *L*=3 and *D*=2   (b) *L*=6 and *D*=5

Figure 6 Experimental results of the classification performance with *L* and *D* on UMIST

Table 2 The most recognition rates of various algorithms on different face datasets under different L

Algorithm name	Face datasets					
	AR		Yale		UMIST	
	5	10	3	6	3	6
SPP	72.55%	85.53%	63.33%	81.00%	57.03%	79.81%
PCFP	71.71%	80.04%	78.75%	78.33%	80.79%	88.08%
SSPCPPGA(5,10)	**82.15%**	**88.39%**	**89.38%**	**91.04%**	**87.34%**	**91.83%**
RLPP	71.73%	80.02%	78.75%	78.33%	81.20%	85.38%
SSDR	70.71%	77.97%	77.71%	78.33%	79.77%	81.62%
NPSSDR	79.73%	87.34%	81.96%	85.67%	81.77%	87.92%
DSPP	77.01%	85.53%	82.96%	86.67%	83.94%	86.90%

Table 1 Specific settings of parameters on algorithms

Algorithm name	Parameters settings
SPP	no
PCFP	no
SSPCPPGA(p,c)	$pc = 0.6, pm = 0.1$
RLPP	$\kappa = 7, t = 1$
SSDR	$\alpha = 1, \beta = 20$
NPSSDR	$\alpha = 0.1, \kappa = 5$
DSPP	no

The simplest Nearest Neighbor classification algorithm is adopted. We select randomly L images from each group face for training samples and remains for testing. Retained feature dimensions are increased with the increment D and corresponding classification accuracies are calculated. All experiments are repeated 20 times and average recognition rates are gotten.

Besides, the matrix XX^T probably is singular since the number of training samples is much smaller than the feature dimension. To deal with the problem, training data firstly projected into the PCA subspace $\widetilde{X} = T_{PCA}^T X$. The ratio of PCA is set to 1.

4.3. Experimental results and analyses

4.3.1. Results and analyses on the classification performance of SSPCPPGA

In the experiment, we select SSPCPPGA(5,10) where the parameter p is set to 5 and the parameter c is set to 10. Fig.4-Fig.6 show experimental results of the classification performance with L training samples of each group and D increment of retained dimensions.

Moreover, in order to verify more accurately the classification performance of SSPCPPGA, the most recognition rates of various algorithms on different face datasets under different L are given and shown in Table2. Bold items represent maximum recognition rates.

From above Fig.7-Fig.9 and Table 2, we draw following conclusions:

(1) SSPCPPGA is obviously superior to SPP. Besides, SSPCPPGA outperforms PCFP and also can get top recognition rate under lower dimensions. This illuminates that the linear weighted fusion way in Eq(10) is efficient and SSPCPPGA inherits advantages of SPP and PCFP.

(2) In contrast to SSDR, the advantages of SSPCPPGA is obvious. The reason is that SSDR preserves structure information based on linear scatter matrices while SSPCPPGA adopts to preserve structure information with sparse representation that has more power for describing geometric structures in data.

(3) Although RLPP and SSPCPPGA share nonlinear advantages, SSPCPPGA outperforms RLPP owing much to different description ways for manifold structure in data. RLPP only capture local manifold structures with the nonlinear approximation, which ignore natural intrinsic geometric structure information in data while Sparse reconstruction of SSPCPPGA is available of intrinsic geometric properties.

(4) NPSSDR preserves local structure information by adding a regularization item instead of constructing the adjacency matrix of data. However, NPSSDR does not change the way of preserving structure information with the nonlinear approximation. Therefore the performance of NPSSDR is not so good as that of SSPCPPGA.

(5) DSPP has a high discriminative ability which is inherited from the sparse representation of data that is shared by DSPP and SSPCPPGA. But DSPP attempts to maintain the prior low-dimensional representation constructed by the data points and the known class labels, which is the main reason for that SSPCPPGA is obviously superior to DSPP.

4.3.2. Analyses on the effect of parameters of GA on the algorithm

To evaluate the effect of the parameter p and the parameter c on the algorithm, we set different p and c in experiments. Experimental results on the effect of parameters are shown in Fig.7-Fig.9.

From experimental results of Fig.7-Fig.9, we draw the conclusion that the performance of SSPCPPGA is not sensitive to the parameter p and the parameter c when they exceed the threshold value.
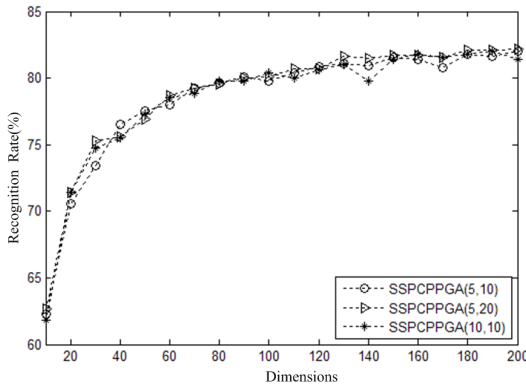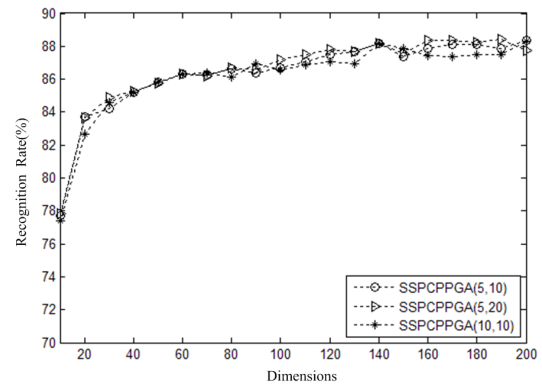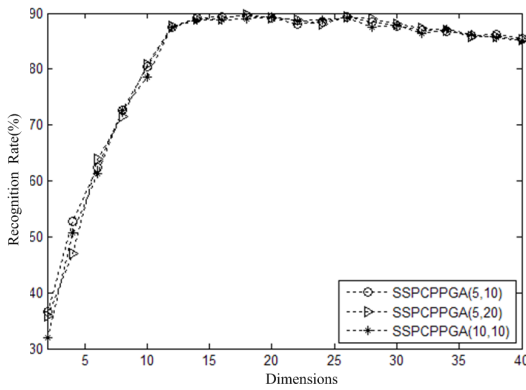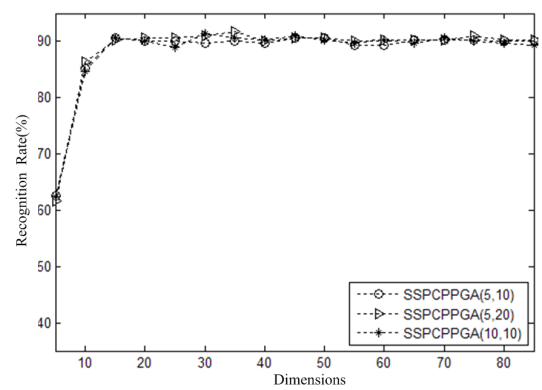
(a) $L=5$ and $D=10$ (b) $L=10$ and $D=10$

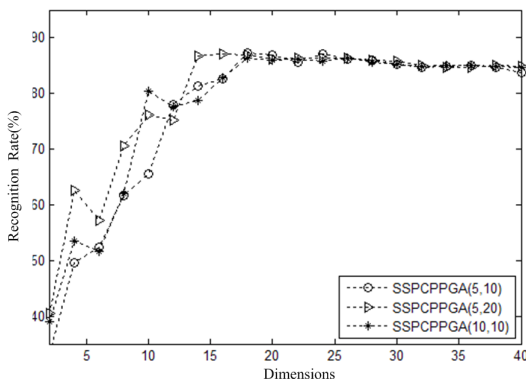**Figure 7** Experimental results of the effect of parameters with L and D on AR
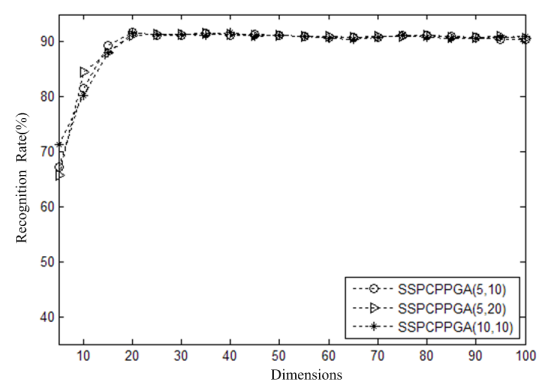

(a) $L=3$ and $D=2$ (b) $L=6$ and $D=5$

Figure 8 Experimental results of the effect of parameters with L and D on Yale

(a) $L=3$ and $D=2$ (b) $L=6$ and $D=5$

Figure 9 Experimental results of the effect of parameters with L and D on UMIST

4.3.3. Analyses on the optimization performance of the weighted trade-off parameter β

In order to evaluate the adaptivity of the weighted trade-off parameter β , we firstly select some retain dimension and increase gradually β from 0 to 1 with the increment of 0.05 and calculate corresponding recognition rate. Moreover, the optimized weighted
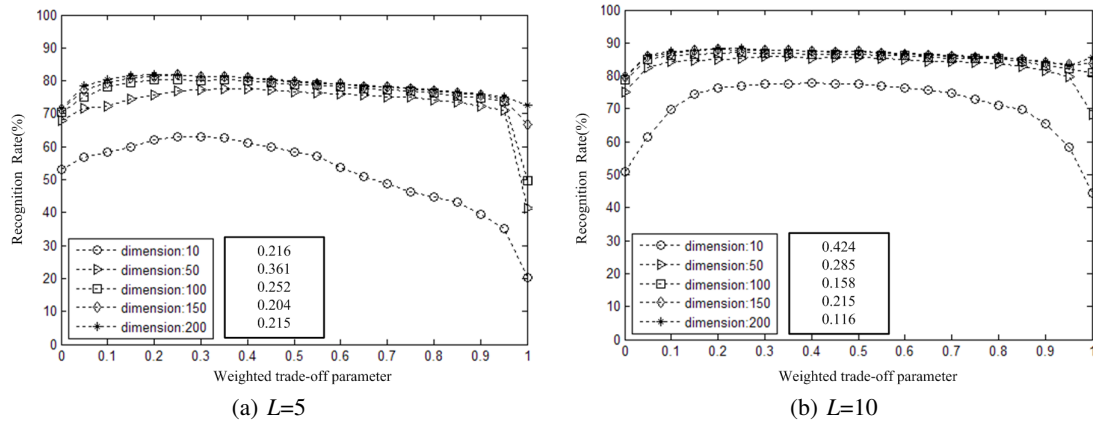
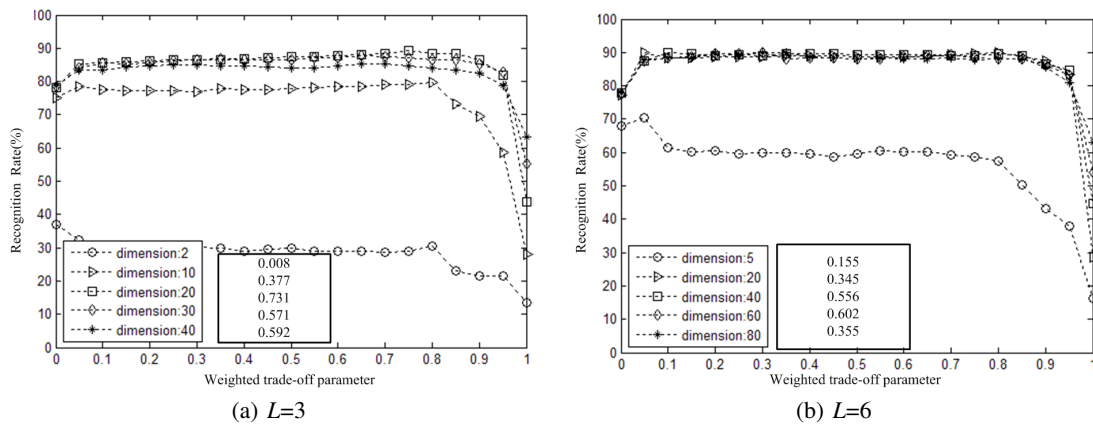**Figure 10** Experimental results of the optimization of the weighted trade-off parameter β with L on AR

Figure 11 Experimental results of the optimization of the weighted trade-off parameter β with L on Yale
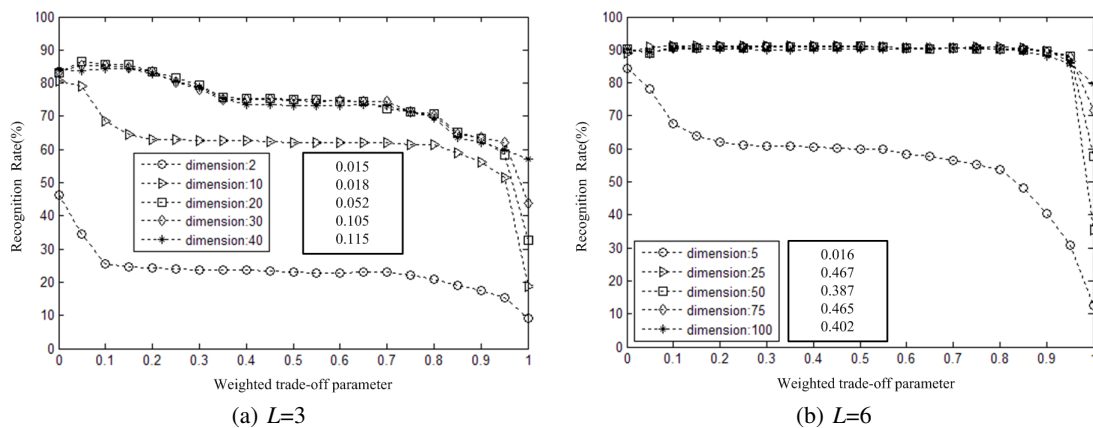

Figure 12 Experimental results of the optimization of the weighted trade-off parameter β with L on UMIST

trade-off parameter β of different retained dimension are listed, which are obtained by SSPCPPGA(5,10).

Fig.10-Fig.12 show experimental results on AR,Yale and UMIST.

From experimental results of Fig.10-Fig.12, we draw following conclusions:

(1) When retained dimension number exceeds the threshold of dimension, the role of the weighted trade-off parameter β on the classification performance is nearly identical, which illuminates the stability of the adaptivity of the weighted trade-off parameter β in SSPCPPGA(5,10).

(2) The way of interval of 0.05 increments need to calculate 20 individual value to obtain the optimized value while SSPCPPGA(5,10) only to calculate 15 individual value, which demonstrates less time cost of SPCPPGA in searching the optimized value.

(3) The optimized trade-off parameter β calculated by SSPCPPGA(5,10) is almost near to the optimal value in experimental results, which confirm its effectiveness of the adaptivity of the weighted trade-off parameter β in SSPCPPGA(5,10).

(4) The optimized trade-off parameter β is different on different datasets, which demonstrates the necessity of the adaptivity of the weighted trade-off parameter β in SSPCPPGA(5,10).

4.4. Computational complexity analyses

For samples $X = \{x_1, x_2, x_3, \cdots, x_n\} \in R^{d \times n}$,SSPCPPGA contains main steps for solving S_a, P_a ,the trade-off β based on GA and the eigen-decomposition using Eq.(10). The computational complexity of sparse learning is nearly that of solution of l_1 norm minimization problems which is $O(d^3)$ [27]. Therefore the computational complexity of solving S_a is $O(d^3)$. The computational complexity of solving P_a is $O(d^2)$.The eigen-problem on a symmetric matrix can be efficiently computed by the singular value decomposition (SVD) which is $O(d^3)$. Main computational time in GA algorithm is caused by calculating the adaptability that contains steps for solving S_a, P_a, and the eigen-decomposition. So the computational complexity of solving the trade-off β based on GA is $O(d^3)$. To sum up, the computational complexity of SSPCPPGA is $O(d^3)$. Computational complexity analyses on various algorithms are shown in Table 3.

Table 3 Computational complexity analyses

Algorithm name	Computational complexity
SPP	$O(d^3)$
PCFP	$O(d^3)$
SSPCPPGA	$O(d^3)$
DSPP	$O(d^3)$
RLPP	$O(d^3)$
SSDR	$O(d^3)$
NPSSDR	$O(d^3 + n^2)$

Conclusion

In the paper, a dimensionality reduction algorithm called Semi-supervised Sparsity Pairwise Constraint Preserving Projections based on Genetic Algorithm (SSPCPPGA) is proposed to solve the problem of deficiency of the ability for preserving global geometric structure of data in existing semi-supervised dimensionality reduction with pairwise constraints. On one hand, the algorithm fuses sparse reconstruction feature information and pairwise constraint feature information through the linear weighted way. Projected data preserve geometrical structure in samples and constraints relation of samples. On the other hand, the algorithm adopts GA to get the optimized weighted trade-off parameter. Experiments operated on AR, Yale and UMIST demonstrate the effectiveness of our proposed algorithm.

However, to optimize weighted trade-off parameter based on GA still costs us some time, the fusion way without selecting parameters is more convenience, which is the future work.

Acknowledgments

References

[1] R. Yan, J. Zhang, J. Yang and A. Hauptmann, A discriminative learning framework with pairwise constraints for video object classification, Proceedings of the IEEE computer society conference on computer vision and pattern recognition, **2**,284-291(2004).

[2] E. Xing, A. Ng, M. Jordan and S. Russell, Distance metric learning with application to clustering with side information, Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, **15**, 505-512 (2003).

[3] A. Bar-Hillel, T. Hertz, N. Shental and D. Weinshall, Learning a mahalanobis metric from equivalence constraints, Journal of Machine Learning Research, **6**(6), 937-965 (2006).

[4] W. Tang and S. Zhong, Pairwise constraints-guided dimensionality reduction, SDM Workshop on Feature Selection for Data Mining, 2006.

[5] D. Zhang, Z. Zhou and S. Chen, Semi-Supervised dimensionality reduction, Proceedings of the SDM, 629-634 (2007).

[6] H. Cevikalp, J. Verbeek, F. Jurie and A. Klaser, Semi-Supervised dimensionality reduction using pairwise equivalence constraints, Proceedings of the VISAPP, 489-496 (2008).

[7] G. Yu, H. Peng, J. Wei and Q. Ma, Robust Locality Preserving Projections with Pairwise Constraints, Journal of Computational Information Systems, **6**(5), 1631-1636 (2010).

Appl. Math. Inf. Sci. **7**, No. 3, 1065-1075 (2013) / www.naturalspublishing.com/Journals.asp

1075

[8] J. Wei and H. Peng,Neighbourhood preserving based semi-supervised dimensionality reduction,Electronics Letters, **44**(20), 1190-1191 (2008).

[9] Y. Chen, M. Rege, M. Dong and J. Hua, Incorporating user provided constraints into document clustering, Proceedings of the ICDM, 103-112 (2007).

[10] Y. Peng and D. Q. Zhang, Semi-Supervised canonical correlation analysis algorithm, Journal of Software, **19**(11), 2822-2832 (2008).

[11] I. Davidson, Knowledge driven dimension reduction for clustering, Proceedings of the IJCAI, 1034-1039 (2009).

[12] M. Davenport, M. Duarte, M. Wakin, D. Takhar, K. Kelly and R. Baraniuk, The smashed filter for compressive classification and target recognition, Proceedings of IS&T/SPIE Symposiumon Electronic Imaging: Computational Imaging, 2007.

[13] J. Wright, A. Yang, S. Sastry and Y. Ma, Robust face recognition via sparse representation, IEEE Transaction on Pattern Analysis and Machine Intelligence, **31**(2), 210-227 (2009).

[14] T. Guha, and R.K. Ward, Learning Sparse Representations for Human Action Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, **34**(8), 1576-1588 (2012).

[15] Y. Liu and W. Wei, Sparse Kernel Learning-based Nonlinear Predictive Controllers, Applied Mathematics & Information Sciences, **5**(2), 195-201 (2011).

[16] S. Wang, L.J. Cui ,D.C. Liu, R. Huck, P. Verma, J.J . Sluss, Jr. And S. Cheng, Vehicle Identification Via Sparse Representation, IEEE Transactions on Intelligent Transportation Systems, **13**(2), 955-962 (2012).

[17] H. Zou, T. Hastie and R. Tibshirani, Sparse principal component analysis,Technical report, 2004.

[18] T. D. X. Duong, and V. N. Duong, Principal Component Analysis with Weighted Sparsity Constraint, Applied Mathematics & Information Sciences, **4**(1), 79-91 (2010).

[19] L. Qiao, S. Chen and X. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recognition, **43**(1),331-341(2010).

[20] N. N. Gu, M. Y. Fan, H. Qiao and BoZhang, Discriminative Sparsity Preserving Projections for Semi-Supervised Dimensionality Reduction, IEEE Signal Processing Letters, **19**(7), 391-394 (2012).

[21] J. Esteban, A. Starr and R. Willets, A Review of Data Fusion models and Architectures: towards engineering guidelines, Neural Computing & Applications, **14**(4), 273-287 (2005).

[22] Y. Q. Song, F. P. Nie , C. S. Zhang and S. Xiang , A unified framework for semi-supervised dimensionality reduction,Pattern Recognition, **41**(9), 2789-2799 (2008).

[23] M. Sugiyama, Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis, Journal of Machine Learning Research, **8**(5), 1027-1061 (2007).

[24] K. J. Leng and X.J. Chen, A Genetic Algorithm Approach for TOC-based Supply Chain Coordination, Applied Mathematics & Information Sciences, **6**(3), 767-774 (2012).

[25] A. Isazadeh, H.Izadkhah and A.H. Mokarram, A Learning based Evolutionary Approach for Minimization of Matrix Bandwidth Problem, Applied Mathematics & Information Sciences, **6**(1), 51-57 (2012)

[26] J. C. Hung, C. C. Lin, Hybridization of Genetic Algorithm and IBMUSIC applied to DOA estimation under coherent signals, Applied Mathematics & Information Sciences, **6**(2), 489-496 (2012).

[27] D. L. Donoho and Y. Tsaig, Fast Solution of l_1-norm Minimization Problems When the Solution May be Sparse,IEEE Transactions on Information Theory,**54**(11), 4789-4812 (2008).

MingMing Qi was born in Jiangxi, China, in 1974. He received the MS degree in computer application from Guizhou University in 2003. Currently he is a Ph.D.student at the Department of Computer Science and Engineering in Tongji university and a lecturer in Shaoxing university. His research interests are in the areas of machine learning and image processing.

Yang Xiang was born in Shandong, China. He received Ph.D. from Dalian University of Technology. Currently he is a prefessor of the Department of Computer Science and Engineering in Tongji university. His research interests are in the areas of machine learning ,web semantic and image processing.