Constraint-based sparsity preserving projections and its application on face recognition

Jianjun Chen*

Yuanpei College of Shaoxing University, China

Received 1 March 2014, www.tsi.lv

Abstract

Aiming at the deficiency of supervise information in the process of sparse reconstruction in Sparsity Preserving Projections (SPP), a semi-supervised dimensionality reduction method named Constraint-based Sparsity Preserving Projections (CSPP) is proposed. CSPP attempts to make use of supervision information of must-link constraints and cannot-link constraints to adjust the sparse reconstructive matrix in the process of SPP. On one hand, CSPP obtains the high discriminative ability from supervised pairwise constraint information. On the other hand, CSPP has the strong robustness performance, which is inherited from the sparse representation of data. Experimental results on UMIST, YALE and AR face datasets show, in contrast to unsupervised SPP and existing semi-supervised dimensionality reduction method on sparse representation, our algorithm achieves increase in recognition accuracy based on the nearest neighbour classifier and promotes the performance of dimensionality reduction classification.

Keywords: semi-supervised dimensionality reduction, pairwise constraint, sparse representation, sparse reconstruction, sparsity preserving projections, face recognition

1 Introduction

Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2] fail to explore the essential structure of the data with non-linear distribution and how to select kernel and optimal kernel parameter in kernel version of them is still difficult. Representative Manifold learning algorithms [3-6] have been developed. Unfortunately, all of these algorithms are plagued by the out-of-sample problem. The solution for this problem is to apply a linearization procedure to construct explicit maps over new measurements. For example, Local Preserving Projections (LPP) [7] is a linearization version of LE; Neighbourhood Preserving Embedding (NPE) [8] is a linearization version of LLE; Isometric Projection (IsoProjection) [9] can be seen as a linearized Isomap; and Linear Local Tangent Spacen Alignment (LLTSA) [10] is a linearization of LTSA. But these algorithms fail to explore instinct geometry structure.

In recent years the study of sparse representation (SR) of signals has attracted many attentions. The purpose of the sparse representation is to optimize the most compact representation of a signal with linear combination of atoms in an over complete dictionary. SR has been successfully applied in many practical problems [11–15]. Researches [11] showed that classifier based on SR is exceptionally effective and achieves by far the best recognition rate on some face databases. Nowadays researches on dimensionality reduction based on SR have attached more and more attentions. Sparsity Preserving Projections (SPP) [16] is a representative algorithm. SPP firstly constructs an adjacent weight matrix of the data set based on SR and then evaluate the low-dimensional embedding of the data to best preserve such weight matrix. SPP is proved to outperform PCA, LPP and NPE, and avoids the difficulty of parameter selection as in LPP and NPE. Although SPP is effective, SPP is sensitive to large variations in whole-pattern based feature extractors. On the base of SPP, researchers combine other dimensionality reduction algorithm to overcome the defect of SPP under semi-supervised dimensionality reduction frameworks. [17] proposed a sparse representation-based classifier (SRC) [11] oriented unsupervised dimensionality reduction algorithm which combines SRC and PCA in its objective function. [18] proposed discriminant sparse neighbourhood preserving embedding (DSNPE) by adding the discriminant information into sparse neighbourhood preserving embedding. DSNPE not only preserves the sparse reconstructive relationship of SNPE, but also sufficiently utilizes the global discriminant structures. Discriminative Sparsity Preserving Projection (DSPP) [19] attempts to maintain the prior low-dimensional representation constructed by the data points and the known class labels and, meanwhile, considers the complexity of f in the ambient space and the smoothness of f in preserving the sparse representation of data. However, there is a common defect in above semi-supervised dimensionality reduction algorithms based on SR, namely, these algorithms ignore making use of supervised information to guide sparsity reconstruction of samples. In general, there are different forms of supervision information or

***Corresponding author** e-mail: chengjianjun_2012@163.com

prior knowledge, such as class label, pairwise constraint, and others. Class label may be strong information from the users and cost us many efforts. In contrast, it is more natural to specify which pairs of data points are similar or dissimilar [20]. As a kind of side information, pairwise constraint contain must-links where the pair of data points must be in a same class and cannot-links where the pair of data points must be in two different classes [21]. The utility of pairwise constraints has been demonstrated in many applications [20-30].

Inspired by SPP and pairwise constraints, Constraints-based Sparsity Preserving Projections (CSPP) is proposed in the paper. Different from above semi-supervised dimensionality reduction algorithms based on SPP, CSPP makes use of supervised pairwise constraint information guide and adjust sparse reconstructive weights with penalty item. Experimental results on Yale, UMIST and AR show, in contrast to DSNPE and DSPP, our algorithm is more efficient.

The rest of the paper is organized as follows: Section 2 reviews sparse representation, sparse reconstruction and SPP. Our CSPP is introduced in Section 3. In Section 4, CSPP is compared with some related works. The experimental results are presented and made analyses. Finally, some concluding remarks and future work are provided in Section 5.

2 Related Background

2.1 SPARSE REPRESENTATION

Given a set of training samples $X=\{x_1,x_2,x_3,...,x_n\}\in R^{d\times n}$, sparse representation seeks a sparse reconstructive weight vector s_i for each x_i through the following minimization problem:

$$\min_{s_i} \| s_i \|_0 \quad s.t. x_i = Xs_i , \tag{1}$$

where S_{ij} denotes the contribution of each x_j to reconstructing x_i . $\| s_i \|_0$ is the pseudo- ℓ_0 norm which is equal to the number of non-zero components in S . However, Equation (1) is NP-hard. The solution of ℓ_0 minimization problem is equal to the solution of ℓ_1 minimization problem as follows:

$$\min_{s_i} \| s_i \|_1 \quad s.t. x_i = Xs_i . \tag{2}$$

2.2 SPARSE RECONSTRUCTION

Sparse reconstruction seeks a sparse reconstructive weight vector s_i for each x_i through the following modified ℓ_1 minimization problem:

$$\min_{s_i} \| s_i \|_1 \quad s.t. x_i = Xs_i , \quad 1 = 1^T s_i \tag{3}$$

where $\| s_i \|_1$ denotes the ℓ_1 normal of s_i, $s_i = [s_{i1},...,s_{ii-1},0,s_{ii+1},...,s_{in}]^T \in R^n$ 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.

$$x_i = s_{i1}x_1 + ... + s_{ii-1}x_{i-1} + s_{ii+1}x_{i+1} + ... + s_{in}x_n \quad . \tag{4}$$

The sparse reconstruction matrix $S=[s_1, s_2, ..., s_n]^T$ is attained through calculating S_i .

2.3 SPARSITY PRESERVING PROJECTIONS (SPP)

SPP aims to preserve sparse reconstruction relation of high-dimensional data space to low-dimensional data space. Given the projection matrix T, $T^T Xs_i$ is the projection point of x_i in high-dimensional data space. The objective function of SPP is as follows:

$$\min_T \sum_{i=1}^{n} \left\| T^T x_i - T^T Xs_i \right\|^2 . \quad s.t. \quad T^T XX^T T = I \tag{5}$$

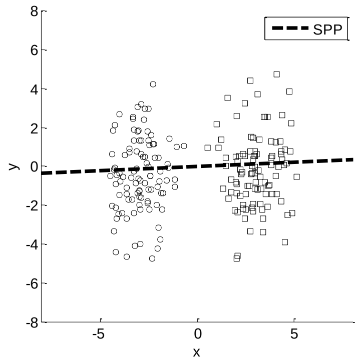
Equation (5) can be further transformed to

$$\max_T \left[T^T X \left(S + S^T - S^T S \right) X^T T \right] . \quad s.t. \quad T^T XX^T T = I \tag{6}$$

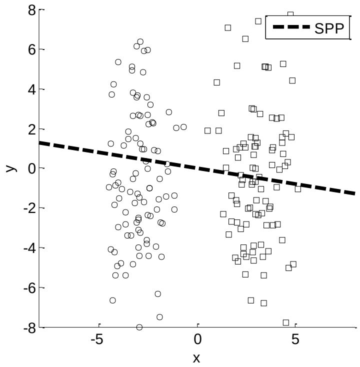
3 Constraints-based Sparsity Preserving Projections (CSPP)

3.1 BASIC IDEA

As described in section 1, in order to illustrate the problem existed in SPP, experiments of SPP on a two-dimensional dataset and the changed dataset are shown in Figure 1.



(a) SPP on a two-dimensional dataset



(b) SPP on the changed dataset

FIGURE 1 Experiments of SPP on a two-dimensional dataset and the changed dataset

Figure 1 depicts 2-dimensional 2-class examples. The circles and triangles denote the samples in positive and negative classes. The solid and dashed lines denote the 1-dimensional embedding spaces of SPP on the dataset. In contrast to the dataset in (a), the vertical scaling of the data is doubled in (b), which lead to change in the whole structure of the dataset. This change of scales affects SPP solutions, which illustrates a possible weakness of SPP arising from its unsupervised nature.

3.2 OBJECTIVE FUNCTION

According above analyses, it is important to introduce supervised information to guide sparse reconstruction in order to overcome the weakness of SPP. Given training samples $X=\{x_1,x_2,x_3,...,x_n\}\in R^{d\times n}$, containing a must-link (ML) set and a cannot-link (CL) set. According to Equation (3) and Equation (4), the sparse reconstructive weight matrix $S=\left(s_{ij}\right)_{n\times n}$ of samples is adjusted [13] proposed to make use of pairwise constraint supervised information to refine adjacency relations of samples with the weighted parameter way. Inspired by [13], the paper adjust the sparse reconstructive weight vector s_i of $x_i\in X$ on the base of s_i. The adjustment of the adjusted sparse reconstructive coefficient $\tilde{s}_{ij}$ of x_j to x_i is described as follows:

$$\tilde{s}_{ij}=\begin{cases} s_{ij}+\dfrac{\alpha\times n_M}{(n_M+n_C)} & if\,(x_i,x_j)\in ML \\ s_{ij}+\dfrac{\beta\times n_C}{(n_M+n_C)} & if\,(x_i,x_j)\in CL\ , \\ s_{ij} & other \end{cases} \tag{7}$$

where n_M denotes this size of *ML* and n_C denotes this size of *CL*. $\tilde{s}_{ij}$ denotes the adjusted sparse reconstructive coefficient of x_j to x_i. α and β denote adjustment parameters.

Equation (7) may be understood in such two sentences: if two samples are in the same class, the greater sparse reconstructive coefficient strengthens their relation as much as possible. If two samples are in two different classes, the less sparse reconstructive coefficient alienates their relation as much as possible.

According to Equation (6), with $S=\left(\tilde{s}_{ij}\right)_{n\times n}$ replacing *S*, the objective function is gotten as follows:

$$\max_T \frac{T^{\mathrm{T}}X(S+S^{\mathrm{T}}-S^{\mathrm{T}}S)X^{\mathrm{T}}T}{T^{\mathrm{T}}XX^{\mathrm{T}}T}. \tag{8}$$

3.3 ALGORITHM STEPS

Input: face training sample $X=\{x_i\,|\,x_i\in R^d\}_{i=1}^n$.

Output: projection matrix $T\in R^{d\times l}(l\le d)$.

Steps:

1) construct the sparse reconstructive matrix s using of Equation (3).

2) get the adjusted sparse reconstructive matrix s with pairwise constraint supervised information using Equation (7).

3) transform Equation (8) into the generalized matrix problem $X(S+S^{\mathrm{T}}-S^{\mathrm{T}}S)X^{\mathrm{T}}t_i=\lambda_i XX^{\mathrm{T}}t_i, 1\le i\le l$ and get the projection matrix $T=[t_1,t_2,...t_l]$.

3.4 COMPUTATIONAL COMPLEXITY ANALYSES

Given samples $X=\{x_1,x_2,x_3,...,x_n\}\in R^{d\times n}$. CSPP contains main steps for solving s and the eigen-decomposition using Equation (8). According to Equation (7), solving S is the key part of solving the sparse reconstructive weight matrix s .The computational complexity of sparse learning is nearly that of solution of` l_1 norm minimization problems which is $O\left(d^3\right)$ [31]. Therefore the computational complexity of solving S is $O\left(d^3\right)$. The eigen problem on a symmetric matrix can be efficiently computed by the singular value decomposition

(SVD), which is $O(d^3)$. Hence, the computational complexity of CSPP is $O(d^3)$.

4 Results and Analysis

4.1 EXPERIMENTAL DATASETS

Some following face datasets are selected in the experiment:

1) UMIST: This set contains 564 images of 20 individuals. Each face image is resized to 112 × 92 pixels with 256 gray levels. The images are covering a range of poses from profile to frontal views. A group of faces in UMIST are shown in Figure 2.

FIGURE 2 A group of faces in UMIST

2) Yale: this database 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. A group of faces in Yale are shown in Figure 3.

FIGURE 3 A group of faces in Yale

3) AR: this database consists of over 4000 face images of 126 individuals. For each individual, 26 pictures were taken in two sessions (separated by two weeks) and each section contains 13 images. These images include front view of faces with different expressions, illuminations and occlusions. A group of faces in AR are shown in Figure 4.

FIGURE 4 A group of faces in AR

4.2. EXPERIMENTAL SETTINGS

L images are selected randomly from a group face and remains for test samples. Besides, pairwise constraints set with the size PC are created randomly from training samples. In order to eliminate the singular problem, training samples are projected into the PCA [1] subspace. The performance of the proposed algorithm is evaluated and compared with that of several methods using the Nearest Neighbour Classifier (NNC). As a baseline, the classification results of NNC directly used the raw data without dimensionality reduction is given. SPP, DSNPE and DSPP are also introduced for comparing with our algorithm. Parameters of various algorithms are set in Table 1.

TABLE 1 Parameter settings of various algorithms

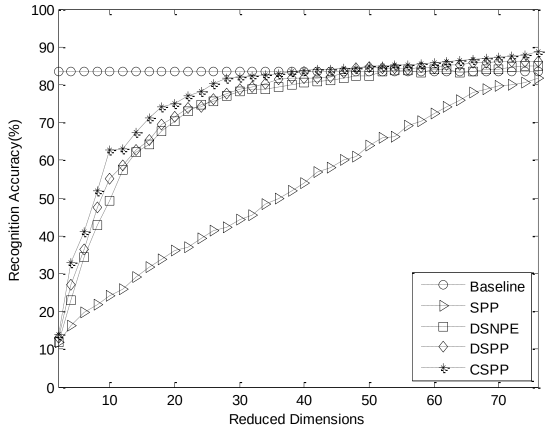
Algorithms name	Parameter settings
Baseline	no
SPP	no
DSNPE	$\gamma = 0.5$
DSPP	$\gamma_A = 0.001$, $\gamma_I = 1$
CSPP	$\alpha = 10$, $\beta = 30$

4.3 EXPERIMENTAL RESULTS

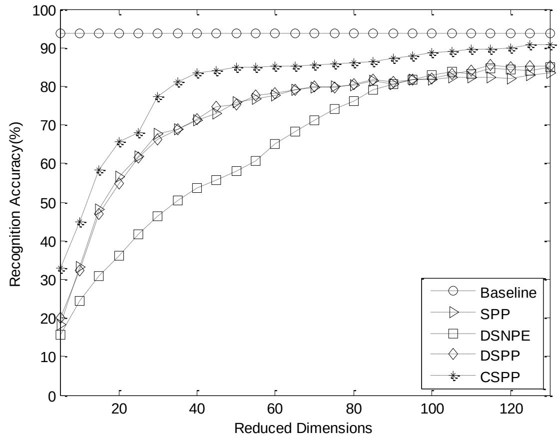
In order to verify efficiently the performance of our proposed algorithm, experiments and analyses are made under different reduced dimensions and pairwise constraints sets with the different size. All experiments are repeated twenty times and average recognition accuracy rates are gotten.

4.3.1 Effect of reduced dimension on the performance

Reduced dimensions are selected with the certain increment and corresponding average recognition accuracy are calculated. Concrete experimental results are shown in Figures 5–7.

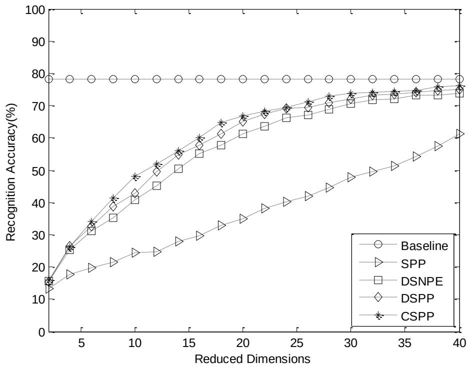


(a)L=4 and PC=800

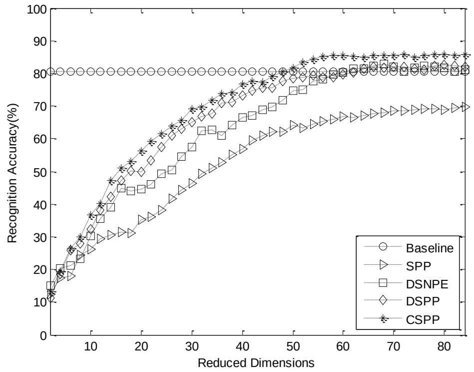


(b) L=8 and PC=1600

FIGURE 5 Recognition accuracy (%) VS. Reduced dimension on UMIST with L and PC

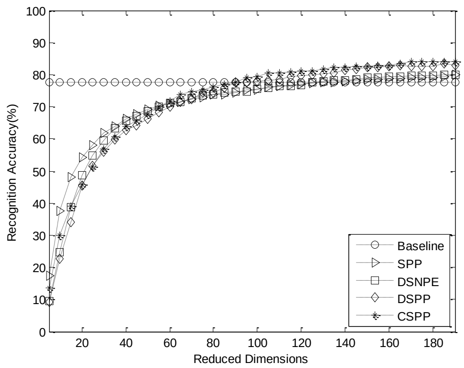


(a) L=3 and PC=200

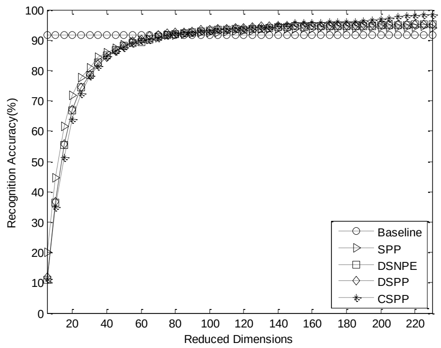


(b) L=6 and PC=400

FIGURE 6 Recognition accuracy (%) VS. Reduced dimension on Yale with L and PC



(a) L=5 and PC=2000



(b) L=10 and PC=4000

FIGURE 7 Recognition accuracy (%) VS. Reduced dimension on AR with L and PC

From Figures 5–7, the following conclusions are drawn:

1) With increase on reduced dimensions, the recognition accuracy of SPP and CSPP promote. CSPP is superior to SPP in face datasets with different character, which illustrates that the adjustment way on the sparse reconstructive weight matrix in Equation (7) is efficient.

2) Although DSNPE combines sparsity criterion and maximum margin criterion (MMC) together to project the input high-dimensional image into a low-dimensional feature vector, integrating both the robustness advantage of sparse representation and distinctiveness advantage of MMC, the recognition accuracy of CSPP still is higher than DSNPE. This show that, in contrast to the way of infusing sparse reconstruction information of SR and discriminative information of MMC, the way of making use of supervised pairwise constraints information to guide the adjustment on sparse reconstructive weight matrix is more efficient.

3) Owing to providing s an explicit feature mapping by fitting the prior low-dimensional representations which are generated randomly by using the labels of the labelled data points, DSPP has a high discriminative ability which is inherited from the sparse representation of data. However, the CSPP outperform DSPP, which is caused by the reason that DSPP pay attention to set the smoothness regularization term to measure the loss of the mapping in preserving the sparse structure of data and ignore the defect of sparse reconstruction in SR.

4.3.2 Effect of the size of pairwise constraints sets on the performance

Under different L on different face datasets, pairwise constraints sets are created with the different size PC and calculated the corresponding maximum recognition accuracy. Concrete experimental results are shown in Tables 2–4.

TABLE 2 Experimental results on UMIST

L	*PC*	Recognition accuracy (%)
4	400	81.50
	800	85.43
	1200	86.25
	1600	86.50
8	500	85.07
	1000	87.35
	1500	89.23
	2000	90.73

TABLE 3 Experimental results on Yale

L	*PC*	Recognition accuracy (%)
3	100	73.87
	200	74.87
	300	75.87
	400	76.16
6	200	81.06
	400	82.13
	600	83.73
	800	84.80

TABLE 4 Experimental results on AR

L	*PC*	Recognition accuracy(%)
5	1000	82.87
	2000	83.19
	3000	84.15
	4000	85.65
10	2000	97.08
	4000	97.19
	8000	98.27
	10000	98.85

From Tables 2–4, some conclusions are drawn as follows:

1) With increase in the size of pairwise constraints sets, the recognition accuracy become more higher, which illustrate that more supervised pairwise constraints information is effective for adjustments of sparse reconstructive weight matrix in Equation 6.

2) Although there are less increment of the pairwise constraints set on UMIST and Yale than AR, the performance of CSPP is more sensitive to the increment on UMIST and Yale. It is reason for the problem that the performance of CSPP is influenced by the ratio of the increment of the pairwise constraints set to the size of the training sample instead of the absolute increment of the pairwise constraints set.

5 Conclusion

Constraints-based Sparsity Preserving Projections (CSPP) is proposed for dimensionality reduction in the paper. On the base of SPP, CSPP adjust the sparse reconstructive weight matrix through the penalty way with supervise pairwise constraints information. Experimental results on UMIST、YALE and AR demonstrate the effectiveness of our proposed algorithm. However, for CSPP, the certain size of the pairwise constraints set is needed. Although attaining pairwise constraints information is simpler than class label information, the work cost us much effort. Therefore, how to introduce new supervised information guide sparse reconstruction is our next work.

References

[1] Fukunaga K 1990 Introduction to Statistical Pattern Recognition 2nd ed *Academic Press* Boston USA

[2] Duda R O, Hart P E Stork, D G 2000 Pattern Classification second ed *John Wiley & Sons* New York

[3] Sam T R, Lawrence K S 2000 Nonlinear dimensionality reduction by locally linear embedding *Science* **290**(5500) 2323–6

[4] Joshua B T, Vin de S, John C L 2000 A global geometric framework for nonlinear dimensionality reduction *Science* **290**(5500) 2319–23

[5] Mikhail B, Partha N 2003 Laplacian eigenmaps for dimensionality reduction and data representation *Neural Computation* **15**(6) 1373–96

[6] Zhenyue Z, Hongyuan Z 2003 Nonlinear dimension reduction via local tangent space alignment *In Proceedings of Intelligent Data Engineering and Automated Learning (IDEAL'2003)* Springer 21-3

[7] Xiaofei H, Partha N 2003 Locality preserving projections *Advances in Neural Information Processing Systems (NIPS'2003)*

[8] Xiaofei H, Deng C, Shuicheng Y, Hong-Jiang Z 2005 Neighborhood preserving embedding *In Proceedings Of International Conference on Computer Vision (ICCV2005)* 1208-13 IEEE Computer Society October 2005

[9] Deng C, Xiaofei H, Jiawei H 2007 Isometric projection *In Proceedings Of AAAI Conference on Artificial Intelligence* Wikipedia

[10] Tianhao Z, Jie Y, Deli Zhao, Xinliang Ge 2007 Linear local tangent space alignment and application to face recognition *Neural Computation* **15**(6) 1547-53 *(in Chinese)*

[11] Yang A Y, Ganesh A, Sastry S S, Yi M 2009 Robust face recognition via sparse representation *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(2) 210–27

[12] Patrik O H 2004 Non-negative matrix factorization with sparseness constraints *Journal of Machine Learning Research* **5**(11) 1457–69

[13] Xiaorong P, Zhang Y, Ziming Z, Wei Z, Mao Y 2005 Face recognition using Fisher non-negative matrix factorization with sparseness constraints *In Proceedings of the Second international conference on Advances in neural networks (ISNN'2005)* Springer-Verlagpp 112–7

[14] Ran H, Bao-Gang H, Wei-Shi Z, Xiang-Wei K 2011 Robust principal component analysis based on maximum correntropy criterion I*EEE Transactions on Image Processing* **20**(6) 1485–94

[15] Ran H, Wei-Shi Z, Bao-Gang H 2011 Maximum correntropy criterion for robust face recognition *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(8) 1561–76

[16] Lishan Q, Songcan C, Xiaoyang T 2010 Parsity preserving projections with applications to face recognition *Pattern Recognition* **43**(1) 331–41

[17] Lei Z, Meng Y, Zhizhao F, Zhang D 2010 On the dimensionality reduction for sparse representation based face recognition *In Proceedings of International Conference on Pattern Recognition (ICPR'2010)* IEEE Computer Society 1237–40

[18] Jie G, Zhenan S, Wei J, Rongxiang H, Yingke L, Shuiwang J 2012 Discriminant sparse neighborhood preserving embedding for face recognition *Pattern Recognition* **45**(8) 2884–93

[19] Nannan G, Mingyu F, Hong Q, Bo Z 2012 Discriminative Sparsity Preserving Projections for Semi-Supervised Dimensionality Reduction *IEEE Signal Processing Letters* **19**(7) 391–4

[20] Rong Y, Jian Z, Jie Y, Hauptmann A G 2006 A discriminative learning framework with pairwise constraints for video object classification *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(4) 578–93

[21] Xing E, Ng A, Jordan M, Russell S 2003 Distance metric learning with application to clustering with side-information *In Proceedings of Advances in Neural Information Processing Systems* MIT Press 505-12

[22] Aharon B H, Tomer H, Noam S, Daphna W S 2006 Learning a mahalanobis metric from equivalence constraints *Journal of Machine Learning Research* **6**(6) 937–65

[23] Zhang D, Zhou Z, Chen S 2007 Semi-Supervised dimensionality reduction *In Proceedings of SIAM International Conference on Data Mining (SDM' 2007)* ACM 629–34

[24] Chen Y, Rege M, Dong M, Hua J 2007 Incorporating user provided constraints into document clustering *In Proceedings of International Conference on Data Mining(ICDM' 2007)* Nebraska USA 103–12

[25] Wei J, Peng H 2008 Neighborhood preserving based semi-supervised dimensionality reduction *Electronics Letters* **44**(20) 1190-1

[26] Hakan C, Jakob V, Frédéric J, Alexander K 2008 Semi-Supervised dimensionality reduction using pairwise equivalence constraints *In: Proc. Of International Conference on Computer Vision Theory and Applications (VISAPP'2008)* ACM 205–14

[27] Guoxian Y, Hong P, Jia W, Qianli M 2010 Robust Locality Preserving Projections with Pairwise Constraints *Journal of Computational Information Systems* **6**(5) 1631–36

[28] Shiguo C, Daoqiang Z Semisupervised 2011 Dimensionality Reduction With Pairwise Constraints for Hyperspectral Image Classification *IEEE Geoscience and Remote Sensing Letters* **8**(2) 369–73

[29] Mariam K, Philippe B, Ludovic M, Denis H 2011 Constraint scores for semi-supervised feature selection: A comparative study *Pattern Recognition Letters* **32**(5) 656–65

[30] Maggini M, Melacci S, Sarti L 2012 Learning from pairwise constraints by Similarity *Neural Networks* **26**(10) 141–58

[31] Donoho D L, Tsaig Y, 2008 Fast Solution of` l1-norm Minimization Problems When the Solution May be Sparse *IEEE Transactions on Information Theory* 54**(11)** 4789–812

Authors

Jianjun Chen, born in May, 1966, Zhejiang, China

Current position, grades: associate professor of Shaoxing University, Shaoxing, China.
University studies: Bachelor's degree in physics science from Zhejiang University in 1987, the Master's degree from Taiyuan University of Technology in 2004
Scientific interest: machine learning and image processing
Publications: 11