Global Face Recognition Framework Based on Symmetrical 2DPLS by two Sides plus LDA

Jiadong Song, Xiaojuan Li
Capital Normal University
Information Engineering College
Beijing 100048, China
E-mail:rhadamanthys0407@126.com

Pengfei Xu,Mingquan Zhou
Beijing Normal University
College of Information Science and Technology
Beijing 100875, China
E-mail:pengfeixu@gmail.com

***Abstract*—A novel face recognition method is proposed in this paper to alleviate the "Small Sample Size" problem of the conventional Linear Discriminant Analysis (LDA). This method is based on the feature extraction of global odd and even face image representation, and a dimension reduction process via Symmetrical 2D Partial Least Square Analysis (2DPLS) by two sizes. The low-dimensional features are then used to train a LDA classifier. Experimental results on Yale Face Database B and Feret face Database demonstrate that our framework is highly efficient and gives the state-of-the-art recognition rate.**

Keywords- face recognition; LDA; two sides; Symmetrical PLS; dimension reduction

I. INTRODUCTION

Face recognition is a key to many applications ranging from video surveillance, person identification, information security and human tracking. Although significant progress has been made recently, there is still much room to boost before we can apply face recognition techniques to the real-world applications. On the one hand, the major challenges of face recognition lie in the effects resulted from illumination change, pose variation, information loss caused by projection from 3D world to 2D image, and image noise etc. On the other hand, because of large computational cost brought by the high dimensionality of face image vector, people usually seek some statistical dimension reduction methods, such as PCA, FLDA, CCA and PLS, for extraction of low-dimensional features before classification. Reference [1], however, shows that face recognition rate don't increase but decrease, as the illumination, brightness and other factors change. Therefore, some researchers use the image pre-processing before recognition, but effect is still not satisfactory.

In this paper, we propose a new face recognition framework to alleviate the "Small Sample Size" problem of the conventional Linear Discriminant Analysis (LDA). This framework is based on the feature extraction of global odd and even face image representation, and a dimension reduction process via symmetrical 2D partial least square analysis (2DPLS) by two sides. The low-dimensional features are then used to train a LDA classifier.

The remainder of this paper is organized as follows: after reviewing existing techniques in Section Ⅱ, we briefly describe our framework in Section Ⅲ, and then introduce selection of our framework for image pre-processing, feature extraction, classification in Section Ⅳ, Ⅴ and Ⅵ, respectively. What's more, discussion and experimental results are presented in Section Ⅶ. Last but not least, we conclude in Section Ⅷ.

II. PREVIOUS WORKS

Many interesting global face recognition approaches have been proposed in the literature. Generally, they are divided into two categories which are based on methods and frameworks. On the one hand, because traditional PCA [2] or FLDA [3] in dealing with the 2D image was generally indicated as a matrix by row or column vectors, which usually causes the problem of "curse of dimensionality", and makes it difficult to estimate covariance matrix. To deal with this problem, some 2D statistical dimension reduction methods are proposed recently. Therefore, Yang et al. [4] first proposed 2DPCA method which improved PCA algorithm. Kong et al. [5] extended 2DPCA by generalizing 2DPCA to a bilateral projection one. Kong et al. [6] proposed 2DFLDA method which improved FLDA algorithm. Yang and Ding [7] proposed a novel approach SPCA which improved PCA through even, odd image transformation. Yu and Yang [8] proposed that a direct LDA algorithm for high-dimensional data with application to face recognition. J and M [9] proposed partial least squares components for face recognition. Recently Yang et al. [10] proposed 2DPLS for face recognition, but this 2DPLS method was based on non-iterative partial least squares and not on classical iterative partial least squares; on the other hand, Zhou et al. [11] proposed a framework using BDPCA plus LDA, but this framework is still vulnerable to brightness, noise, illumination, etc. O et al. [12] proposed a framework using ICA plus SVM, but this framework produces such excellent results based on two categories not on multi-categories. Baback et al. [13] proposed a framework using Bayesian, but Bayesian needs lots of priori probability and we can only get these valid data through great computing and lots of experiments.

III. Global face recognition framework based on Symmetrical 2DPLS by two sides plus LDA

Our global face recognition framework consists of four components, specifically as follows:

- Image Pre-processing. This module will mainly

This work was supported by the BMDHR（20081D0501600187）and PHR（IHLB）Funding Project.

978-0-7695-3987-4/10 $26.00 © 2010 IEEE
DOI 10.1109/ETCS.2010.60

make the image clearer, and reduce larger amount of different noise.

- Feature Extractions. This module will extract the global features using Symmetrical 2DPLS by two sides. Two sides method looks very similar to a bilateral projection one [5], but it's so different. It reduces dimension of matrix by horizontal direction, then reduces dimension of Eigen matrix by vertical direction rather than reduce dimension at the same space. Therefore, a bilateral projection one needs much larger space and is less accurate than ours. Next, we will introduce odd and even images. Because even image energy is larger than odd image, even eigenvector which is stable will give priority to be selected. Experimental results also confirm that Symmetrical 2DPLS by two sides is better than most other methods, e.g. 2DPCA, 2DFLDA, etc.
- Classification. This module will use LDA method, because LDA method can better balance the relationship between class and class in low-dimensional space, while SVM only produces such excellent result based on two categories. Although the SVM can also handle the multiclass of recognition by One-against-One or One-against-Rest method, it requires a large amount of cost and increases the complexity. Therefore, this is not much value. Experimental results show that not only LDA Classification can better improve the recognition rate than SVM, but also requires only a small amount of cost.
- Statistics recognition rate. Usually, leave one out [14] method is used, when the set is too small, especially.

The global face recognition chain is shown as in fig.1

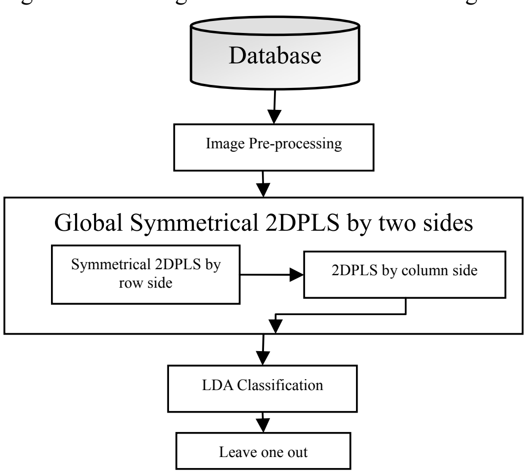


Figure 1. the Global Face Recognition Chain

This global face recognition framework requires neither any probability statistics such as Bayesian, nor the complicated hierarchical structure such as artificial neural networks and decision trees, only requires a simple chain structure. Experimental results show that our framework is very simple and effective, too.

IV. Image Pre-processing

Image Pre-processing plays a critical role in the recognition rate. Because the images of people's face are too vulnerable to the effects of light direction and strength, we need to use the two methods about equalizations obviously. One is based on the global histogram equalization; the other is based on the local illumination compensation normalization. Reference [1] says that recognition rate based on histogram equalization of compensation is lower than local illumination compensation. Therefore, this paper will use the method based on local illumination compensation method.

$$f_p(x,y)=\frac{f(x,y)-E(f(x,y))}{D(x,y)+0.01} \tag{1}$$

Where E(f(x,y)) and D(f(x,y)) denote the mean and variance of neighborhood of pixel (x,y). $f_p(x,y)$ is the new value of pixel by (1).And the 0.01 is used to prevent denomination which is zero.

Through the above processing, there will be some noise, still. Hence, we will use the method which is corrosion as well as expansion the image by morphology. At the same time, this method obtains excellent results.

V. Global Feature Extraction

In this section we will describe some global feature extraction. We first will present feature extraction methods which have often been used. Next, we will introduce 2DPLS and Symmetrical 2DPLS by two sides.

A. 2DPCA Feature Extraction

2DPCA [2] feature extraction has improved the PCA algorithm. Traditional PCA in dealing with the 2D image was generally indicated as a matrix by row or column vectors, which usually causes the problem of "curse of dimensionality", and makes it difficult to estimate covariance matrix. To overcome this, 2DPCA is described as follows:

Let A_{ij} be the image of the i^{th} category and the j^{th} sample, μ_i be the mean of the i^{th} category, μ be the mean of all training samples, M be the number of each category in all training samples, and N be the number of all categories. The true μ, μ_i are given as (2) and (3), respectively.

$$\mu = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} A_{ij} \tag{2}$$

$$\mu_i = \frac{1}{M}\sum_{j=1}^{M} A_{ij} \tag{3}$$

Covariance matrix, Cor, is given as (4)

$$Cor = \frac{1}{N}\sum_{i=1}^{N}(\mu_i-\mu)(\mu_i-\mu)^T \tag{4}$$

We first solve the eigenvalues and eigenvectors of (4). Next, we choose the first k largest eigenvalues and corresponding eigenvectors. In the sequel, the eigenvectors matrix is, here defined as V= [v1, v2,…, vk].

Although either PCA or 2DPCA has many advantages such as reducing the dimension of eigenspace, they have two fatal disadvantages. One is susceptible to noise, brightness, etc. The other one is disadvantaged when the number of training samples per category is too large.

B. 2DFLDA Feature Extraction

Similarly, 2DFLDA [3] feature extraction has improved the FLDA algorithm, so we directly introduce 2DFLDA feature extraction. The main idea of 2DFLDA feature extraction makes larger between-class distance and smaller within-class variance so that separate the different types of data as much as possible. And the most importance is that the "small sample size problem" does not exist anymore because the between-class and within-class scatter matrices constructed in 2DFLDA are both of full rank [3]. Specifically as follows:

$A_{i,j}$, μ_i,μ,M and N first are the same as parameters of 2DPCA feature extraction section. Moreover, we define between-class scatter matrix and within-class scatter matrix as S_b and S_w which are given as (5) and (6)

$$S_b = \frac{1}{N}\sum_{i=1}^{N}(\mu_i - \mu)(\mu_i - \mu)T \tag{5}$$

$$Sw = \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}(A_{ij} - \mu_i)(A_{ij} - \mu_i)T \tag{6}$$

According to Fisher criterion, $J_F(w)$ is given as (7)

$$J_F(w) = \frac{w^T S_b w}{w^T S_w w} \tag{7}$$

The eigenvalues and eigenvectors are first solved by (7). Furthermore, we choose the first k largest eigenvalues and corresponding eigenvectors. In the sequel, the eigenvectors matrix is, here defined as W= [w1, w2,…, wk].

C. 2DPLS Feature Extraction

PLS is a new multivariable analysis method which can reduce the impacts from noise and makes the face recognition more robust, especially when the number of variables is too large and multi-correlation exists, while the number of observational data is too small. Before PLS had conducted derived feature space of the least squares regression, it chose the Eigen by the covariance. Meanwhile, as already pointed out, because 2DPLS inherited the main advantages of 2DPCA and 2DCCA, 2DPLS is described as (8).

$$2DPLS \approx 2DPCA + 2DCCA \tag{8}$$

2DPLS algorithm is specifically given as follows:

According to (4), we first solve its eigenvalues and eigenvectors. Moreover, the iterative formula is given as (9)

$$Cor_{(i+1)} = \left(\frac{I - \left(Cor_j u_j u_j^T Cor_j^T\right)}{\left(u_j^T Cor_j^T Cor_j u_j\right)} \right) Cor_j, \quad j = 1\ldots,k \tag{9}$$

Where u_j is the eigen direction, k is the dimension, and we will not stop Iterative formula until j=k. Then, the eigenvectors matrix is, here defined as U=[u1,u2,…,uk]. Last but not least, $Y_{i,j}$,as (10),is the new training sample of A_{ij} after dimension reduction of 2DPLS.

$$Y_{i,j}=U^T \times A_{ij} \tag{10}$$

D. Symmetrical 2DPLS by two sides

The main idea about Symmetrical 2DPLS by two sides has two aspects. On the one hand, we will introduce symmetrical image, odd image and even image representation [7]; on the other hand, our 2DPLS is based on two sides, namely matrix of horizontal and vertical directions achieve dimensionality reduction in turn. If the original matrix is a 100×100 size, then we can reduce dimension of matrix, and make the Eigen matrix become a d1×d2 size (d1<100 and d2<100).

Because any function can be decomposed into an odd function and an even function, we think the image as a complex function and reference [7] gives three formulas as (11), (12) and (13)

$$I=I_{even}+I_{odd} \tag{11}$$

$$I_{odd}=(I-I_{sym})/2 \tag{12}$$

$$I_{even}=(I+I_{sym})/2 \tag{13}$$

Where I is an original image, I_{sym} is a symmetrical image of the original image, I_{even} is an even image of the original image, and I_{odd} is an odd image of the original image.

Firstly, angle, uneven illumination and other factors have been created face images of non-symmetry based on I_{odd}, while relatively stable features of face have been created based on I_{even}.

Secondly, there is some identification information except noise, so though I_{odd} have a lot of features which are noise, we don't give up all but choose the most important parts of features.

Finally, we should take measures to do everything possible to exclude noise. We combine the I_{odd} and I_{even} together, furthermore, we give different weights to them and the weight of I_{even} owning the stability of identifying information should be larger than that of I_{odd}.

According to (10), we can first achieve two eigenvalues and corresponding eigenvectors matrices named E_{even}, $E_{odd,}$ V_{even} and V_{odd} via I_{even} and I_{odd}, respectively. Moreover, because the even image representation is stable, the w_{even}, which is the product of eigenvalue and weight, is usually larger than that of w_{odd}. Then, we merge two matrices W_{even} and W_{odd} into W. Last but not least, all w are sorted by descending, and we choose eigenvectors matrix, as V_1, corresponding to the first k largest values. V_1 which mixes the eigenvectors v_{even} and v_{odd} may be expressed as [v_{even}1, v_{even}2, v_{odd}3, … , v_{even}i,…v_{odd}k].

The eigenvector matrix of symmetrical 2DPLS by two sides is given as (11)

$$\begin{aligned} F_1 &: Y_{ij} = V_1^T \times A_{ij} \\ F_2 &: S_{ij} = Y_{ij} \times S_2 \end{aligned} \tag{11}$$

Let F_1 be Eigen space of A, F_2 be Eigen space of F_1, V_1 be the eigenvector matrix of the row direction of space A and S_2 be the eigenvector matrix of the column direction of Eigen space F_1, namely $V_1^T \times A$ which is nothing other than feature extraction matrix.

We can find that (11) is not same as a bilateral projection one. On the one hand, because a bilateral projection one is based on one space, memory requirement is larger than our method by two sides. On the other hand, because our method by two sides is a serial manner, feature extraction of Eigen space F2 will be little for influence by noise, namely feature extraction is more effective than bilateral projection one.

In a word, Symmetrical 2DPLS by two sides has the following advantages.

- To reduce the impact of brightness, noise and illumination.
- To reduce more dimension than most methods because of two sides.
- To reduce the more memory requirements than a bilateral projection one.
- Global feature extraction using Symmetrical 2DPLS is more effective than PLS, 2DPCA, 2DFLDA, et al.
- To composite factors based on the image of stability and instability and choice of the eigenvalues, which is different from the previous methods e.g. PCA, et al, using the product of the weight value and eigenvalue.

VI. CLASSIFICATION METHODS

The main idea of LDA Classification Method is similar to 2DFLDA feature extraction, so we will not introduce it too much. According to (10), we can achieve the feature mean of each category as μ_j, and then the metric dispersion is given as (12)

$$D_i = \arg\min \sum_{j=1}^{N} d^2(\mu_j, \text{testing}) \tag{12}$$

If i=j then count=count+1 otherwise e=e+1, where e and count represent the number of false recognition and true recognition, respectively. D_i represents that the shortest distance between the testing sample and μ_{lj} of each category is the i^{th} category.

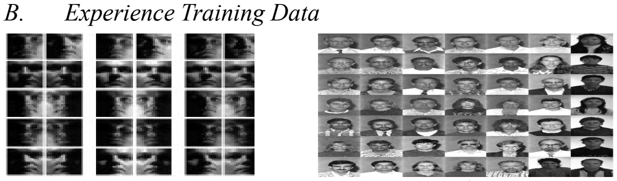
VII Experience

In this section we present experimental results for our framework. We first explain the overall framework of our experience, then show the results of different training set and testing set, and compare with other methods or frameworks.

A. The overall framework of experience

Our overall framework consists of the following four components, namely a) Image pre-processing, b) Feature Extraction based on Symmetrical 2DPLS by two sides, c) Classification based on LDA in low-dimensional space, d) Statistics recognition rate.

B. Experience Training Data



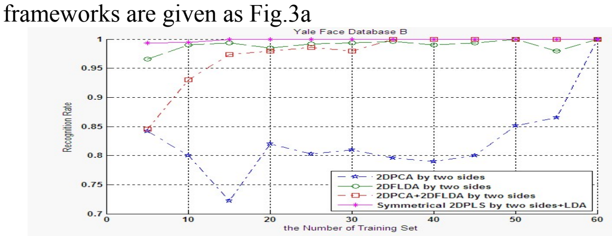
a b

Figure 2. Thumbnails based on Some images of Yale Face Database B and Feret Face Database

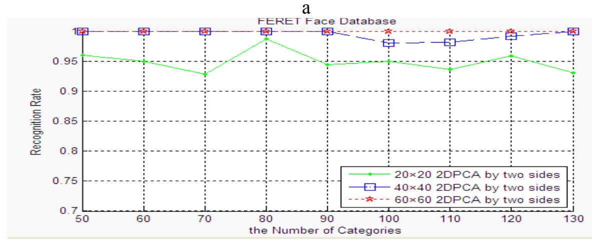
All experiments are conducted using each category i training samples and remaining for test. All experiments use Yale Face Database B and Feret Face Database (see Fig.2a, Fig.2b).Yale Face Database B is chosen because of illumination, noise, etc and Feret Face Database is chosen because of its large quantities. Between them, we use Yale Face Database B of 10 categories, each category contains the 65 samples, and Feret Face Database of 130 categories, each category contains the 2 samples which are "fa" and "fb" images. Meanwhile, for the sake of simple calculation, all images of two Face Databases are scaled to a 100×100 size.

C. Experience results and analysis

We first test recognition rate using Yale Face Database B. Let the number of training set in each category be i (i=5, 10, 15,…, 60) , the number of testing set in each category be 65-i, and the number of categories be 10. According to the test, recognition rate curves of some different methods and frameworks are given as Fig.3a



a



b

Figure 3. Two different Testing Results

There is lots of noise, brightness, and illumination in images of Yale Face Database B, so the effect of contrast will be more obvious than others. According to Fig.3a, firstly, we can find that the recognition rate of saturation using this new framework is significantly earlier than other methods or frameworks, but more importantly is that the recognition rate is significantly higher than other methods, when the horizontal coordinate is fixed. Secondly, this new framework has a high recognition rate and high stability. Because LDA classification method makes the distance between the different categories increase, recognition rate is so stable; because 2DPLS dimension reduction inherits the advantages of 2DPCA, 2DFLDA and 2DCCA, extracted features are superior to 2DPCA, 2DFLDA, etc; because we use the symmetrical 2DPLS in the context of odd and even image representation which can decrease interference of noise, brightness and illumination, it leads to this new framework which is highly robust and effective, eventually. At the same time, we find that recognition rate of 2DPCA by two sides is lower than others. This is because that 2DPCA is vulnerable to brightness, noise and illumination factor. We still find that methods of 2DFLDA by two sides and of 2DPCA plus 2DFLDA are also stable. This is because that 2DFLDA makes the same category closer and closer but the different categories farther and farther. Therefore, it is more concerned about the dimension reduction data to distinguish optimization of the different data types, rather than focusing on optimization based on the original high-dimensional data fidelity.

We then test recognition rate using Feret Face Database. The amount of this database is so large. Through the test, we give curves based on new framework using the different dimension as Fig.3b using Leave one out method. Meanwhile, when dimension reduces to 40×40, this new framework has the general phenomenon of under-fitting, but this is quite normal. For instance, when there are 120 categories and dimension is 40×40, the recognition rate is 99.17% and the error rate is 0.83% (see Fig.3b or Fig.4). We can see that two people face expression and the appearances of their faces are very similar except for the impact of their hair (see Fig.5). To this end, this new framework can be also used by the large database through verification of Feret Face Database (see Fig.3.b).

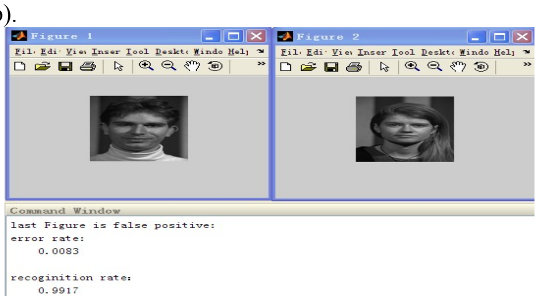


Figure 4. the Display of Experiment Result based on 130 Categories and 40×40 Dimension

Figure 5. Comparison after Figure4 is processed

Finally, the recognition rate of some different methods and frameworks is tested by Yale Face Database B, after the dimension is reduced as in Table1.The superscript represents two sides using dimension reduction, and S2DPLS represents Symmetrical 2DPLS. We find that recognition rate of our framework is significantly higher than others through dimension reduction.

TABLE I. RECOGNITION RATE OF DIFFERENT DIMENSION REDUCTIONS

	Size of Dimension Reduction			
	80×80	***60×60***	***40×40***	***20×20***
$2DPCA^2$	52. 05%	53. 86%	53. 41%	79. 86%
$2DFLDA^2$	85. 56%	91. 33%	97. 78%	98. 76%
(2DPCA+2DFLDA) +LDA	89. 2%	89. 6%	95. 6%	100%
$S2DPLS^2$+LDA	100%	100%	100%	100%

VIII SUMMARY and CONCLUSIONS

This paper proves an in-depth experimental study on face recognition. Multiple global features extraction combinations have been examined with respect to their experience curves and efficiency on a large database set, such as Yale Face Database B and Feret Face Database, with ground truth.

Global features, here represented by Symmetrical 2DPLS plus Symmetrical 2DPLS feature extraction coefficients, are found to be superior to global features represented by 2DPLS plus 2DPLS, 2DPCA, etc. At first, this feature extraction can better avoid the interference of light, noise and other factors. Furthermore, dimension reduction by two sides needs less cost of space-time than that by one side, while Eigenvector space is far smaller than the original space. Last but not least, with regarding to classification method, there is an advantage about LDA which is similar to cohesion and heterogeneous coupling, so our approach, on the one hand, has a higher stability, on the other hand, has a higher recognition rate. In conclusion, the new framework about face recognition outperforms the other state-of-the-art methods. In the next step, we will combine global and local features of face, and then design the new mixture framework. But it's our future work.

ACKNOWLEDGMENT

This work was supported by the BMDHR (20081D050 1600187) and PHR (IHLB) Funding Project. We also thank Doctor Zhou, Teacher Kong and Xu for their kindly help.

REFERENCES

[1] Xudong Xie, Kin-Man Lam, "An Efficient Illumination Normalization Method for Face Recognition, " Pattern Recognition Letters, vol.27(6), pp.609-617, 2006

[2] I. T. Jolliffe, "Principal Component Analysis," Series:Springer Series in Statistics,2nd ed., Springer, NY, 2002

[3] Belhumeur P N,Hespanha J P,Kriegman D J, "Eigenfaces Verus Fisherfaces: Recognition Using Class Specific Linear Projection, " IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.19(7), pp.711-720, 1997

[4] J Yang,D Zhang,A F Frangi,J Y Yang, "Two dimensional PCA: A new approach to appearance based face representation and recognition," IEEE Trans.Pattern Anal. Machine Intell, vol.26(6), pp131-137,2004

[5] Kong H,Wang L,Teoh EK Li X ,et al., "Generalized 2D principal component analysis for face image representation and recognition," Neural Networks, vol.18(5-6), pp. 585-594, Jul-Aug 2005

[6] Hui Kong,Lei Wang, Eam Khwang Teoh,et al., "A Framework of 2D Fisher Discriminant Analysis: Application to Face," CVPR 2005.

[7] YANG Qiong and DING Xiao-Qing, "Symmetrical PCA and Its Application to Face Recognition," Chinese Journal of Computers, vol.26(9), pp.1146-1151, Sept 2003, (in Chinese)

[8] Hua Yu and Jie Yang, "A Dirct LDA Algorithm for High-Dimentional Data with Application to Face Recognition," Pattern Recgntion, vol.34(10), pp.2067-2070, Oct 2001

[9] J Baek and M kim, "Face Recognition using partial least squares component," Pattern Recognition, vol.37(6), pp.1303-1306, June 2004

[10] Mao-Long Yang, Quan-Sen Sun and De-Shen Xia, "Two-Dimensional Partial Least Squares and Its Application In Image Recognition," Proceedings of 4th International Conference on Intelligent Computing (ICIC'08) Shanghai vol.15 pp.208-215,Sep 2008

[11] Wangmeng Zhou, David Zhang, Yang Jian,et al., "BDPCA Plus LDA: A Novel Fast Feature Extraction Technique for Face Recognition," IEEE Trans on systems,man,and cybernetics, vol.36(4), pp.946-953, Aug 2006

[12] O.Deniz,M.Castrillon and M.Herandez, "Face recognition using independent component analysis and support vector machines," Pattern Recognition Letters,vol.24(13), pp.2153-2157,Sep 2003

[13] Baback Moghaddam, Tony Jebara and Alex Pentland, "Bayesian face recognition, " Pattern Recognition, vol.33(11), pp.1771-1782, Nov 2000

[14] Ray Kumar S and Ghoshal J, "Approximate reasoning approach to pattern recognition," Fuzzy Sets and Systems, vol.77(2), pp.125-150, Jan 1996