A New Face Recognition Framework: Symmetrical Bilateral 2DPLS plus LDA

Jiadong Song, Xiaojuan Li
Information Engineering College, Capital Normal University, Beijing, China
Email: rhadamanthys0407@126.com

Pengfei Xu, Mingquan Zhou
College of Information Science and Technology, Beijing Normal University, Beijing, China
Email: {pengfei, mqzhou}@bnu.edu.cn

Abstract—A novel face recognition framework is proposed in this paper to alleviate "Small Sample Size" (SSS) 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 Bilateral 2D Partial Least Square Analysis (2DPLS). The low-dimensional features are then used to train a LDA classifier which uses Frobenius-norm classification measure, and uses pseudo inverse to make sure between-class matrix $S_w$ be full rank. Experimental results on Yale Face Database B, ORL, and FERET Face Database demonstrate that our framework is highly efficient and gives the state-of-the-art recognition rate.

Index Terms—face recognition, pseudo inverse, Frobenius-norm, Linear Discriminant Analysis, Symmetrical Bilateral 2DPLS, two-dimension,

I. INTRODUCTION

Face recognition is a key to many applications ranging from video surveillance, person identification, human tracking and information security. 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. At first, the major challenges of face recognition lie in the effects resulted from illumination from 3D world to 2D image, and image noise etc. In the sequel, because of large computational cost brought by the high dimensionality of face image vector as a point in the high-dimensional vector space, researchers usually seek some statistical dimension reduction methods or frameworks, such as PCA, LDA, LPP, ICA and PLS, for extraction of low-dimensional features before classification. And though numerous successes based on vector space have been achieved, it still leads to the some problems.

1. The 2D image matrix must be transformed into a 1D long vector. Therefore, the dimensionality of face image space can range from several hundreds to thousands. That is so-called "curse of dimensionality".

2. Traditional method based on vector space will destroy the original structure of the data matrix, and lead to the face recognition rate dropped.

3. Traditional LDA is not robust, because the rank of $S_w$ matrix is not full rank.

Last but most importantly, Reference [1] shows that face recognition rate do not increase but decrease, when the illumination, brightness and other factors change. Therefore, some researchers use the image pro-processing before face recognition, whereas the effect is still not satisfactory.

In this paper, we propose a novel face recognition framework to alleviate the "Small Sample Size" (SSS) 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 the dimension reduction process via Symmetrical Bilateral 2D Partial Least Square Analysis (2DPLS). The low-dimensional features are then used to train a LDA classifier. Though we can extract feature via Symmetrical Bilateral 2DPLS which greatly reduces the possibility of "SSS" problem, it can never fundamentally solve the "SSS" problem. Therefore, we use pseudo inverse, which is embedded in the traditional LDA classification method, to solve this problem. In addition, we use Frobenius-norm measure instead of the traditional Euclidean distance, namely 2-norm, or Mahalanobis distance.

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, and classification in Section IV. What's more, discussion and experimental results are presented in Section V. Last but most importantly, we conclude in Section VI.

II. PREVIOUS WORKS

Many interesting global face recognition approaches have been proposed in some literatures. Generally, they are divided into two categories which are based on methods and frameworks.

Global face recognition approaches based on methods as follow:
Firstly, because traditional PCA [2] or LDA [3] in dealing with the 2D image was generally indicated as a matrix by a long row or column vector, 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. Yang et al. [4] first proposed 2DPCA method which improved PCA algorithm. Kong et al. [5] extended 2DPCA by generalizing 2DPCA. These two methods make the covariance matrix become quite small, so their feature extraction can be evaluated more accurately than those traditional vector space methods. However, the method like traditional PCA, 2DPCA, or G2DPCA is only good at image representation rather than discrimination. When there are large pose- and illumination-variations in face images, these methods don't model identity information but these external variations. However, Kong et al. [6] proposed 2DLDA method which improved LDA algorithm, and is proved that the "SSS" problem does not exist anymore because the between-class matrix $S_b$ is full rank, and 2DLDA can extract more discrimination information.

Secondly, Yang and Ding [7] proposed a novel approach SPCA which improved PCA via even and odd image transformation. The advantage of this method increases the number of sample, when the amounts of face samples are too little such as only 1 or 2 samples per class. And the face recognition rate which uses the event and odd image information slightly outperforms PCA, LPP, etc. However, there also is a fatal disadvantage, such that this method never embeds any other methods which own the discrimination information for classification.

Thirdly, Yu and Yang [8] proposed that a direct LDA algorithm which incorporates the concept of null space for high-dimensional data with application to face recognition. It first removed the null space of the $S_b$ matrix, and then seeks a projection to minimize the trace of within-class covariance in the range space of $S_b$. The rank of $S_b$ is smaller than that of $S_a$, so removing the null space of $S_b$ maybe lose part of or the entire null space of $S_a$, which is very likely to be full rank after the removing operation.

Fourthly, Cai and He [9, 10, 11] proposed the novel method named LPP. This method is a novel local feature extraction based on Graph embedding, and is also proved that global method such as PCA equal to LPP, when embedding adjacent graphs are some special graphs. Meanwhile, LPP is more approximate than LDA, while LPP is an unsupervised method, but LDA is supervised method. However, there also are two disadvantages. Firstly, LPP needs lots of cost to calculate the adjacent graph; next, we don't know which adjacent graph is best such as based on K-NN, similarity, Fuzzy set, etc. Hu, et al. [18] proposed 2DLPP method which improved the traditional LPP method. This 2DLPP method has a fatal error that $S_b$ should be a matrix not a value. Meanwhile, this method is applied to palmprint recognition not to face recognition.

Last but most importantly, J Baek and M kim [12] proposed partial least square components for face recognition. Recently, Yang et al. [13] proposed 2DPLS for face recognition, and yet this 2DPLS method was based on non-iterative partial least square and not on iterative partial least square. Because 2DPLS inherit the main advantages of 2DPCA and 2DCCA, we will create the framework based on 2DPLS and LDA embedded for classification information.

Global face recognition approaches based on frameworks as follow:

Zhou et al. [14] first proposed a framework using brightness, noise, illumination, etc.

In the sequel, O. Deniz et al. [15] proposed a framework using ICA plus SVM, but this framework produces such excellent results based on two classes not on multi-class. More importantly, the calculation cost of SVM is too large.

Finally, Baback et al. [16] proposed a framework using Bayesian, but Bayesian needs lots of priori probability information and we can only get these valid data through great computing and lots of experiments from a mounts of samples. More importantly, when the amounts of samples are very little, the probability usually is distortion.

### III. Global Face Recognition Framework

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

- **Image Pre-processing.** This module will mainly make the image clearer, and reduce larger amount of different noise. Usually, the local illumination compensation normalization method slightly outperforms the global illumination compensation normalization [1].

- **Feature Extraction.** This module will extract the global features using Symmetrical Bilateral 2DPLS. Bilateral feature extraction method looks very similar to Bilateral Projection One [5], but it is so different. Bilateral feature extraction reduces dimension of Eigen-matrix through horizontal direction, and then reduces dimension of Eigen-matrix through vertical direction rather than reduces dimension at the same original 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 usually larger than odd image, even eigenvector which is stable will give priority to be selected. Experimental results also confirm that Symmetrical Bilateral 2DPLS is better than most other methods, e.g. 2DPCA, 2DLDa, 2DLPP, etc.

- **Classification.** In this module, we will use LDA method, because LDA method can better balance the relationship between class and class in dimensional space, while SVM only produces such excellent result based on two classes. Although the SVM can also handle the
multiclass of recognition by One-against-One or One-against-Rest strategy, it requires a large amount of calculation cost and increases the complexity. Therefore, this is not much value. Experimental results show that not only LDA Classification can better and simpler improve the recognition than SVM, but also requires only a small amount of cost.

- Statistics recognition rate. Usually, Leave-One-Out [17] 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](image)

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

IV. SYMMETRICAL BILATERAL 2DPLS PLUS LDA

In this section, we will introduce a novel framework for face recognition. Image pre-processing, feature extraction, and classification will be described by theoretical analysis and justification, respectively.

A. Image Pre-processing

Image Pre-processing plays a critical role in the face recognition rate. Because the images of people's face are too vulnerable to 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, we will use the method based on local illumination compensation method.

\[
f_p(x,y) = \frac{f(x,y) - E(f(x,y))}{D(x,y) + 0.01}
\]

(1)

Where

- \(E(f(x,y))\) is the mean of neighborhood of pixel \((x,y)\);
- \(D(f(x,y))\) is the variance of neighborhood of pixel \((x,y)\);
- \(f_p(x,y)\) is the new pixel value;
- \(0.01\) is used to prevent denominator equal to 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. And this method obtains lots of excellent results.

B. Symmetrical Bilateral 2DPLS Feature Extraction

The main idea about Symmetrical Bilateral 2DPLS 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 bilateral, namely the matrix of horizontal and vertical directions achieve dimension 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 \(d_h \times d_v\) size (\(d_h<100\) and \(d_v<100\)).

In the first place, because any function can be decomposed into an odd function and even function, we think the image as a complex function and Reference [7] gives three functions as (2),(3), and (4)

\[
I_{ordi} = I_{even} + I_{odd}
\]

(2)

\[
I_{odd} = \frac{(I_{ordi} - I_{sym})}{2}
\]

(3)

\[
I_{even} = \frac{(I_{ordi} + I_{sym})}{2}
\]

(4)

Where

- \(I_{ordi}\) represents an original image;
- \(I_{sym}\) represents a symmetrical image of the original image;
- \(I_{odd}\) represents an odd image of the original image;
- \(I_{even}\) represents an event image of the original image.

We should notice three points.

Firstly, angle, uneven illumination and other factors have been created non-symmetrical face images based on \(I_{even}\).

Secondly, though \(I_{ordi}\) has a lot of features which have lots of noise, we don't give up all but only choose the most important parts of feature.

Finally, we should measure to do everything possible to exclude noise. We combine the \(I_{ordi}\) and \(I_{even}\) together; furthermore, we give different weights to them and the weight of \(I_{even}\) owing to the stability of identification information should be larger than that of \(I_{ordi}\).

In the second place, we will introduce 2DPLS in detail.

PLS is a new multivariable analysis method which can reduce the impacts from noise and makes the face
recognition become more robust, especially when multi-correlation exists and the number of variables is too large, yet the number of observation data is too little. 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 (5)

\[ 2DPLS \approx 2DPCA + 2DCCA \] (5)

2DPLS algorithm is specifically given as follows: The covariance matrix is given as (6)

\[
Cor = \frac{1}{NC} \sum_{i=1}^{C} \sum_{j=1}^{Ni} (x_{ij} - \mu)(x_{ij} - \mu)^T
\]

\[
= \frac{1}{C} \sum_{i=1}^{C} (\mu_i - \mu)(\mu_i - \mu)^T
\]

Where

- \( \mu \) represents the total mean matrix of all training samples;
- \( \mu_i \) represents the i-th class mean matrix;
- \( C \) represents the number of all classes;
- \( Ni \) represents the number of each class;
- \( x_{ij} \) represents the i-th class and j-th image sample.

According to (6), we first solve its Eigen-values and Eigen-vectors. Moreover, the iterative formula is given as (9)

\[
Cor_{(j)} = \left( I - Cor_{u_i u_j} Cor_{u_j}^T \right) Cor_{(j-1)} \quad j=1, \ldots, k
\]

Where \( u_i \) is the Eigen direction, \( k \) is the dimension, and we will not stop iterative formula until \( j=k \). Then, the Eigen-vectors matrix is, here defined as \( U = [u_1, u_2, \ldots, u_k] \) . Last but not least, \( Y_{ij} \), as (8), is the new training sample of \( A_{ij} \) after dimension reduction of 2DPLS.

\[
Y_{ij} = U^T A_{ij}
\] (8)

According to (8), we can first achieve two group Eigen-value and corresponding Eigen-vector matrices which is named \( A_{even}, A_{odd}, V_{even} \) 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 Eigen-value 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 Eigen-vector matrix, as \( V_{even} \) corresponding to the first \( k \) largest values. \( V_{even} \), which mixes the Eigen-vectors \( v_{even} \) and \( v_{odd} \) may be expressed as \( [v_{1 even}, v_{2 even}, v_{3 even}, \ldots, v_{1 odd}, \ldots, v_{k odd}] \).

The Eigen-vector matrix of Symmetrical Bilateral 2DPLS is given as (11)

\[
F_1 : Y_{ij} = V_1^T A_{ij}
\]

\[
F_2 : S_{ij} = Y_{ij} V_2
\] (9)

Where

- \( F_1 \) represents the Eigen-space of the original space;
- \( F_2 \) represents the Eigen-space of the \( F_1 \) space;
- \( V_1 \) represents the Eigen-vector matrix of the \( F_1 \) space;
- \( V_2 \) represents the Eigen-vector matrix of the \( F_2 \) space;
- \( A_{ij} \) represents the i-th class and the j-th sample image.

Therefore, we can find that Eq. (9) is not same as a bilateral projection one. On the one hand, a bilateral projection one is based on one space, which is the original image space, so memory requirement and time cost are larger than our bilateral method. On the other hand, our method is based on \( F_1 \) then \( F_2 \) Eigen-space, so it will be little for influence by noise. Therefore, feature extraction is more effective than bilateral projection one.

In a word, Symmetrical Bilateral 2DPLS has the following advantages.

- To reduce the impact of brightness, noise and illumination.
- To reduce more dimension than most methods because of two directions.
- To reduce more memory requirements than a bilateral projection one.
- Global feature extraction using Symmetrical Bilateral 2DPLS outperforms some more popular feature extraction method, e.g. PLS, 2DPCA, 2DLDA, 2DPP, et al.
- To composite factors based on the image of stability and instability and choice of the Eigen-values, which is different from the previous methods e.g. PCA, etc, using the product of the weight value and Eigen-value.

C. LDA Classification

After a transformation by Symmetrical Bilateral 2DPLS, a feature matrix is obtained for each face image. Then, we will first use LDA extract discrimination information, and then use the simplest and most effective method for classification, which is named by "Nearest Neighbor classifier ".

The main idea of LDA classification makes larger between-class distance and smaller within-class variance so that separate the different types of data as much as possible.

Specifically as follow:

At first, we define between-class scatter matrix and within-class scatter matrix as \( S_b \) and \( S_w \) which are given as (10) and (11)

\[
S_b = \sum_{i=1}^{C} P_i (\mu_i - \mu)(\mu_i - \mu)^T
\]

\[
S_w = \sum_{i=1}^{C} P_i E_i (x - \mu_i)(x - \mu_i)^T
\] (10) (11)

Where

- \( \mu \) is the total mean vector of all training samples;
- \( \mu_i \) is the mean vector of the i-class training samples;
Let \( C_i \) be the \( i \)-th class, if \( d(A, \text{Mean}_i) = \min_j (d(A, \text{Mean}_j)) \) and \( A_j \in C_i \), then the resulting decision is \( A \in C_i \).

**Theorem 1:** Based on matrix and \( \| \cdot \| \) based on vector are compatible.

The proof is given in Appendix A.

V. EXPERIMENTS AND ANALYSIS

In this section, we present some experimental results for our framework. We will first explain the overall framework of our experiments; in the sequel, we will introduce our experiments on Yale Face database B, ORL, and FERET Face database; then show the results of different training set and testing set, and compare with other methods or frameworks; finally, we will evaluate and analysis the experimental results.

A. The Overall Framework of Experiment

Our overall framework consists of the following four components, namely (a) Image Pre-processing, (b) Feature Extraction based on Symmetrical Bilateral 2DPLS, (c) Classification for training based on LDA in low-dimensional space, (d) Statistics recognition rate such as Leave-One-Out method.

B. Experimental Data on the Yale Face Database B, ORL, and FERET Face Database

Yale Face Database B contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. Hence, the total number of images is in fact 5760 + 90 = 5850.

There are the 65 (64 illuminations + 1 ambient) images of a subject in a particular pose. Fig. 2 shows the thumbnails based on some images of 11 persons, which ten persons own to Yale Face Database B and another one person owns to Extend Yale Database B.

In our experiment, we choose 11 class sets of front face, and each class contains 64 face images by different illumination (see Fig.2).

Figure 2. Thumbnails based on Some Images of Yale Face Database B
ORL Face Database contains ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). A preview image of the Database of Faces is available.

In our experiment, we choose all 40 class sets, which contain different facial expression and facial details. And each class contains 10 faces images. The thumbnails face images is given as Fig.3.

![Figure 3. Thumbnails based on Some Images of ORL Face Database](image1)

The FERET image corpus was assembled to support government monitored testing and evaluation of face recognition algorithms using standardized tests and procedures. The final corpus, presented here, consists of 14051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles.

In our experiment, we choose 130 classes of FERET Face Database; each class contains the 2 samples which are "fa" and "fb" images.

![Figure 4. Thumbnails based on Some Images of FERET Face Database](image2)

Importantly, for the sake of simple calculation, all images are normalized (in scale and orientation) such that the two eyes are aligned at the same position. Then, the facial areas are cropped into final images for matching. The size of each cropped image in all the experiments is $100 \times 100$ pixels. The image is given as Fig.5.

C. Experiment results and analysis

We first test recognition rate using Yale Face Database B. Let the number of training set in each class be $i$ ($i=5, 10, 15, \ldots, 60$), the number of test set per class be $65-i$, and the number of classes be 10. According to the test, recognition rate curves of some different popular methods or frameworks are given as Fig.6.

There is lots of noise, brightness, and illumination in the images of Yale Face Database B, so the effect of contrast will be more obvious than others. According to Fig.6, we can find that the face 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 axis is fixed. Because LDA classification method makes the distance between the different classes increase, recognition rate is so stable; because 2DPLS feature extraction inherits the advantages of 2DPCA and 2DCCA, the extracted feature are superior to 2DPCA or 2DCCA; because we use the symmetrical 2DPLS in the context of the 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 B2DPCA is low than others. This is because that B2DPCA is vulnerable to brightness, noise and illumination factors. We still find that methods of B2DLDA, left 2DLDA plus right 2DPCA and of left 2DPCA plus right 2DLDA are also stable. This is because that 2DLDA makes the within-class closer but the between-class farther, importantly the $S_w$ matrix of 2DPCA is full rank, so the "Small Size Sample" will never happen anymore [6]. Therefore, LDA 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 space. And then we still find that B2DLPP slightly outperforms B2DPCA, and is inferior to B2DLDA. Therefore, it’s true that Local feature extraction is superior to global feature extraction [9]. However, the calculation cost of B2LPP is too large.
In the sequel, we test recognition rate using FERET Face Database. The amount of this database is so large. Through the test, we give curves based on the new framework using different dimension as Fig.7 using Leave-One-Out method. Meanwhile, when dimension reduces to $40 \times 40$, this new framework has the general phenomenon of under-fitting, but this is quite normal. For instance, when there are 120 classes and dimension is $40 \times 40$, the face recognition rate is 99.17% and the face error rate is 0.83% (see Fig.8 or Fig.9). 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.8). To this end, this new framework can be also used by the large face database through verification of FERET Face Database (see Fig.7).

Next, the recognition rate of some different methods and frameworks is tested by Yale Face Database B, when the dimension is reduced as in Table 1. In this Experiment, the number of train samples per class is 5 on the ORL Face Database. The "B" of some methods represents bilateral dimension reduction, and SB2DPLS represents Symmetrical Bilateral 2DPLS. We find that the recognition rate of our framework is significantly higher than others through dimension reduction. When the dimension reduces $20 \times 20$ size, the face recognition rate of Symmetrical Bilateral 2DPLS is still 100%.

Then, we test the robust using pseudo inverse, when "SSS" is happen. In this Experiment, the number of train samples per class is 2 on the FERET Face Database and use Leave-One-Out method for statistics. According to Fig.10, we can find LDA embedding pseudo inverse can prevent the "SSS" problem, because the between-class matrix $S_w$ is full rank. Meanwhile, we also find that the method using pseudo inverse to well prevent the "SSS" problem is very effective and importantly simple.

<table>
<thead>
<tr>
<th>Framework</th>
<th>Size of Dimension reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2DPC A</td>
<td>$80 \times 80$</td>
</tr>
<tr>
<td></td>
<td>$60 \times 60$</td>
</tr>
<tr>
<td></td>
<td>$40 \times 40$</td>
</tr>
<tr>
<td></td>
<td>$20 \times 20$</td>
</tr>
<tr>
<td>B2DPL A</td>
<td>82.05%</td>
</tr>
<tr>
<td></td>
<td>70.86%</td>
</tr>
<tr>
<td></td>
<td>90.41%</td>
</tr>
<tr>
<td></td>
<td>95.86%</td>
</tr>
<tr>
<td>BDL DA</td>
<td>85.56%</td>
</tr>
<tr>
<td></td>
<td>91.33%</td>
</tr>
<tr>
<td></td>
<td>97.78%</td>
</tr>
<tr>
<td></td>
<td>98.76%</td>
</tr>
<tr>
<td>(2DPC A+2DL DA)+L DA</td>
<td>89.20%</td>
</tr>
<tr>
<td></td>
<td>89.60%</td>
</tr>
<tr>
<td></td>
<td>95.60%</td>
</tr>
<tr>
<td></td>
<td>100.00%</td>
</tr>
<tr>
<td>SB2DPL A+LD A</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>100.00%</td>
</tr>
</tbody>
</table>
Finally, we test the face recognition rate using different distance measures. According to Fig.11, we can find the Frobenius-norm based on matrix measure and 2-norm based on vector (traditional 2-norm) measure slightly outperforms any others, and the highest recognition rate is 98%. Meanwhile, the recognition rate using Frobenius-norm is similar to using traditional 2-norm. Therefore, the proof of Theorem 1 is true, namely $||A||_F$ based on matrix measure and $|| \cdot ||$, based on vector measure are compatible.

VI. CONCLUSION AND FUTURE WORK

This paper proves an in-depth experimental study on face recognition. The experimental results show that the Symmetrical Bilateral 2DPLS plus LDA framework has some advantages.

At first, Symmetrical Bilateral 2DPLS plus LDA which extracts the feature of global odd and even face image representation can better avoid the interference of light, noise and other factors via extracting discriminate information.

Furthermore, dimension reduction by two directions needs less cost of space and time than that by one direction, so Eigen-vector space is far smaller than the original space. To do this, we will better avoid the "curse of dimensionality" than before. In the sequel, Symmetrical Bilateral 2DPLS plus LDA framework can avoid a "Small Sample Size" through reducing dimension by Symmetrical Bilateral 2DPLS and using pseudo inverse, which $S_{2D}$ matrix is full rank.

Finally, when the number of Face database is very small, this framework can increase about double number of samples. So it decreases the extent of the lack of samples.

Through Symmetrical Bilateral 2DLPP plus LDA framework has lots of advantages, there are still some aspects of Symmetrical Bilateral 2DLPP that deserve further study. Firstly, usually, although 2DPLS is significantly better than traditional PLS, we can't explain why PLS outperforms 2DPLS when the sample number is very little. Secondly, 2DPLS needs more coefficients for image representation than traditional 2DPLS. Thirdly, we can never still explain how large the parameter should be chosen when using pseudo inverse. Fourthly, we can't still explain how the dimension of Symmetrical Bilateral 2DPLS could be reduced directly not via experiments result. Last but most importantly, because reference [10] shows that the local feature outperforms the global feature, we will combine global and local features of face, and then design the new mixture framework. But it's our future work.

APPENDIX A THE PROOF OF THEOREM 1

Proof: Firstly, $||A||_F$ is clearly a non-negative and homogeneity. Let $a_j (j=1,2,\ldots,n)$ be the $j$-th column of $A$ matrix, and $b (j=1,2,\ldots,n)$ be the column of $B \in \mathbb{C}^{m \times d}$. Thus, we can conclude that

\[ A+BF = \sum_{k=1}^{n} (a_k b_k)_{m \times d} \in \mathbb{C}^{m \times d} \]

Secondly, we assume that $B$ is equal to $(b_j)_{m \times 1} \in \mathbb{C}^{m \times 1}$, and then we can conclude that

\[ AB = \sum_{k=1}^{n} a_k b_k \]

Thus,

\[ \|AB\|_F^2 = \sum_{j=1}^{n} \sum_{k=1}^{n} a_k b_k^2 \leq \sum_{j=1}^{n} a_j \sum_{k=1}^{n} b_k^2 \]

\[ = \sum_{j=1}^{n} \sum_{k=1}^{n} a_k b_k^2 = \sum_{j=1}^{n} \sum_{k=1}^{n} a_k \sum_{k=1}^{n} b_k^2 \]

which is the matrix norm of $A$. According to (18), we can conclude the Theorem is established.

\[ \|Ax\|_2^2 \leq \|A\|_F \|B\|_F \|x\|_2 \]

(19)

Therefore, $||A||_F$ and the vector $|| \cdot ||_2$ are compatible.

ACKNOWLEDGMENT

This work was supported by the BMDHR (2008I0 501600187) and PHR(HLB) Funding Project. We also thank Doctor Zhou and Li, Teacher Xu for their kindly help.

REFERENCES


Jiadong Song received the BS degree in computer science from Dalian University, China, in 2008. He is currently working toward the MS degree in the Information Engineering College, the Capital Normal University, China. His research interests are machine learning, pattern analysis, and computer vision.

Xiaojuan Li received the Ph. D degree from China Agricultural University, China. Dr. Li is now a professor in Capital Normal University, China. And she is the Master Supervisor in Capital Normal University. She has been working on research in image analysis, multimedia sensor networks.

Pengfei Xu received the Ph. D degree from Nanyang Technological University, Singapore in 2007. Dr. Xu is now a lecturer in Beijing Normal University, China. His research interests include computer vision, signal and image processing, stochastic optimization and variation techniques.

Mingquan Zhou received the MS degree from Northwest University, China. Dr. Zhou is now a professor in Beijing Normal University, China. And he is the Ph. D Supervisor in Beijing Normal University. He has been working on research in virtual reality, computer graphic, and Chinese information processing.

© 2011 ACADEMY PUBLISHER