A Two-Dimensional Partial Least Squares with Application to Biological Image Recognition

Shudong Hou, Quansen Sun and Deshen Xia
School of Computer Science and Technology
Nanjing University of Science and Technology
Nanjing 210094, China

***Abstract*—The Partial Least Squares(PLS) is a novel multivariate data analysis method developed from practical applications in real world. It is not influenced by the total scatter matrices of training samples being singular or not. So PLS can efficiently deal with the case of high-dimensional space with only small sample size such as biological feature recognition. The standard PLS firstly reshapes images into vectors. In order not to destroy the inherent structure information, a two-dimensional PLS is proposed which can extract features being more discriminative and dramatically reduces the computational complexity compared to the standard PLS. The proposed method is applied to face and palm biometrics and is examined using the FERET and PolyU palmprint database. Experimental results show that 2DPLS is a good choice for real-world biometrics recognition.**

***Index Terms*—Partial least squares, Pattern recognition, Face recognition, Feature extraction, Feature fusion**

I. Introduction

The Partial Least Squares(PLS) is a multivariate statistical data analysis and regression method which uses projection into latent variables to reduce high-dimensional and strongly correlated data to a much smaller data set that can then be easily interpreted [1]. It has been rapidly developed both in theory and applications over the past two decades. One of the most successful applications of PLS is in chemometrics and chemical industries for statistical data analysis. Recently, it is also applied to many other areas such as process monitoring, marketing analysis and image processing an so on.

In its general form, PLS creates orthogonal score vectors by maximizing the covariance between different sets of variables. PLS dealing with two blocks of variables is considered in this paper, although the PLS extensions to model relations among a higher number of sets exist [2].

PLS can be applied to classification problems by encoding the class membership in an appropriate indicator matrix and has a close connection with Fisher Discriminant Analysis(FDA). By the way, PLS is similar to Canonical Correlation Analysis(CCA) where latent vectors with maximal correlation are extracted. There are different PLS techniques to extract latent vectors, and each of them gives rise to a variant of PLS.

Biological image recognition is a important area in machine learning and pattern recognition, and has a wide application in people's lives. But limited by the difficulty of image acquisition and other reasons, there will not have enough training samples. So the small sample size(sss) problem always happens. In such a case, LDA and CCA can not directly used. There mainly have two methods to settle it . The first method is using Principal Component Analysis(PCA) to reduce the feature to a moderate dimension for solving the sss problem. The final recognition result badly depends on the dimension selected in the PCA. And how to determine a good choice dimension in PCA is still a hard question. The other method is employing the Perturbation strategy. It is need a cross-validate procedure to choose a good perturbation value and is time consuming. However, PLS can efficient extraction the discriminant feature and is not influenced by the sss problem [3] [4] .

The standard PLS is a linear learning algorithm. But in real world, there exist a lot of nonlinear phenomena that can not be well explained by linear method. So we need a nonlinear mapping, then the original feature can be mapped to a high or infinity dimension space where the nonlinear problem can solved by linear method. The difficult is how to choose the nonlinear mapping function. Luckily, we need not to know the concrete nonlinear mapping function. By the virtue of kernel-kick, the final projection can be calculated through the kernel matrix [5] [6]. Lately, the fuzzy inference system is incorporated in PLS. It is used to capture the nonlinearity and to increase the use of experts' knowledge [7].

The inherent structure information is important for classifying images and two-dimensional algorithms are well studied in subspace methods,such as 2DPCA,$(2D)^2PCA$, 2DCCA, 2DCOPLS and so on [8] [9] [10] [11]. In [11], 2DCOPLS is introduced as a tool used in one set of variables. It can not applied to two block sets of variables and is time consuming. Additionally, 2DCOPLS is based on column projection. Motivated from this, we proposed a new 2DPLS which can directly use the image as a whole feature without reshaping to a high-dimension vector. It extracts both the column projection and the row projection, so it will need less coefficients to store the extracted features . It not only largely reduces the time complexity, but also makes sure that the feature extracted is more representational, at the same time avoids the sss problem.

The organization of the paper is as follows. Section 2 introduces the basic theories of PLS and how to applied to feature extraction and recognition. Our method is detailed in section 3. Section 4 presents the experimental procedures and discusses the results. Finally, conclusions are given in section5.

II. Theory of Partial Least Squares (PLS)

Basically, the PLS method is a multivariable linear regression algorithm that can handle correlated inputs and limited data. The objective criterion for constructing components in PLS is to sequentially maximize the covariance between the class variable and a linear combination of the features. Let $X \in R^{n\times p},\ Y \in R^{n\times q}$ respectively be predictor matrix and response matrix. Usually, the response matrix is defined as follows. Suppose there are C classes $(person : 1, 2, \cdots, C)$ to be recognized. We define C-dimensional random vector $y_s = (y_1^s, y_2^s, \cdots, y_C^s)'$, where $y_i^s = 1$ and $y_j^s = 0$ for all $j \neq i$ when the image s belongs to the class i, $i = 1, 2, \cdots, C$. Then we can get the class vector observations $y_1, y_2, \cdots, y_n$ from the training sample of images to construct the $n \times C$ class matrix $Y = [y_1, y_2, \cdots, y_n]'$. PLS is to find the weight vector b_k and c_k such that

$$b_k = arg \max_{b'b=1,\ c'c=1} Cov^2(Xb, Yc) \quad for \quad k = 1, 2, \cdots, K \tag{1}$$

subject to the orthogonality constraint

$$b_k' b_j = 0 \quad for \quad all \quad 1 \leq j < k \tag{2}$$

where b, c are unit vectors. The procedure is called the multivariate PLS. The *k* th PLS component is the linear combination of the original features Xb_k.

In [3], A Non-iterative PLS Modeling Method is proposed. The number of effective weight vectors, which satisfy constraint (2), are composed of vectors which are selected from the eigenvecotrs corresponding to the first d maximum eigenvalues of eigenequations (3) and (4)

$$S_{xy}S_{yx}b = \lambda^2 b \tag{3}$$

$$S_{yx}S_{xy}c = \lambda^2 c \tag{4}$$

where $S_{xy} = X'Y$ and $S_{yx} = Y'X$ are the covariance matrix.

In the classification step, there will be have several fusion method to design a classifier for the acquired two feature vectors from the same pattern. Serial and parellel feature fusion strategies are always used. In [3] a new fusion form is adopted. First, we get two projection matrices $B = [b_1, b_2, \cdots, b_k]\ and\ C = [c_1, c_2, \cdots, c_k]$ through equations (3) and (4). We can acquire the feature vector $z_1 = t_1 B, z_2 = t_2 C$ by projecting it to the projection matrices when facing a sample $t = \{t_1, t_2\}$. Then each a pair of PLS components can constitute the correlation feature matrix $M = [z_1; z_2]$. The distance between any two correlation feature matrices $M_i = [z_1^{(i)}; z_2^{(i)}]\ and\ M_j = [z_1^{(j)}; z_2^{(j)}]$ is defined as:

$$d(M_i, M_j) = \sum_{k=1}^{2} \| z_k^{(i)} - z_k^{(i)} \|_2 \tag{5}$$

where $\| \cdot \|$ represents the vector's Euclidean distance.

Let $\omega_1, \omega_2, \cdots, \omega_c$ be the c known pattern classes, and assume that $\xi_1, \xi_2, \cdots, \xi_n$ are the all training samples and their corresponding correlation feature matrices are $M_1, M_2, \cdots, M_n$. For any testing sample ξ, its correlation feature matrix $M = [z_1; z_2]$. If $d(M, M_l) = \min_j\ d(M, M_j)\ and\ \xi_j \in \omega_k\ ,\ then\ \xi \in \omega_k$

In this paper, we change the indicator matrix as another feature matrix in order to fuse more features so that it can extract more discriminative information. Although PLS in which the two blocks are both features is belonged to a unsupervised learning algorithm while the original is a supervised method, but the recognition accuracy will be higher almost.

III. The Two-dimensional Partial Least Squares (2DPLS)

Now we consider two sets of image data, $\{X_t \in R^{m_x \times n_x}, t = 1, \cdots, N\}$ and $\{Y_t \in R^{m_y \times n_y}, t = 1, \cdots, N\}$ that are realizations of random variable matrix X and Y,respectively. In the conventional PLS, each image data is reshaped into a long vector, and then ,the method described in section 2 is applied. Here, we present a 2DPLS where we directly use image data to determine the maximum relation components between them.

We define mean matrices of X_t and Y_t as

$$M_x = \frac{1}{N}\sum_{t=1}^{N} X_t, \quad M_y = \frac{1}{N}\sum_{t=1}^{N} Y_t \tag{6}$$

Then centered image data are denoted by

$$\tilde{X}_t = X_t - M_x, \quad \tilde{Y}_t = Y_t - M_y$$

The 2DPLS seeks left transforms l_x and l_y and right transforms r_x and r_y such that covariance between $l_x' X r_x$ and $l_y' Y r_y$ are maximized. Then 2DPLS is formulated as

$$\begin{aligned} arg \max_{l_x, r_x, l_y, r_y} cov(l_x^T X r_x, l_y^T Y r_y) \\ s.t. \ \ var(l_x) = 1,\ var(r_x) = 1 \\ var(l_y) = 1,\ var(r_y) = 1 \end{aligned} \tag{7}$$

We define

$$\Sigma_{xy}^r =< \tilde{X} r_x r_y' \tilde{Y}_t' >= \frac{1}{N}\sum_{t=1}^{N} \tilde{X}_t r_x r_y' \tilde{Y}' \tag{8}$$

Note that we can write

$$cov(l_x' X r_x, l_y' Y r_y) =< l_x' \tilde{X} r_x r_y' \tilde{Y}_t' l_y >= l_x' \Sigma_{xy}^r l_y \tag{9}$$

With these definition, the formulation of 2DPLS given in (7) is rewritten as

$$arg\ max\ l_x' \Sigma_{xy}^r l_y,\ s.t.\ l_x' l_x = 1,\ l_y' l_y = 1 \tag{10}$$

Note that (7) can also written as

$$arg\ max\ r_x' \Sigma_{xy}^l r_y,\ s.t.\ r_x' r_x = 1,\ r_y' r_y = 1 \tag{11}$$

where

$$\Sigma_{xy}^l =< \tilde{X}' l_x l_y' \tilde{Y} >= \frac{1}{N}\sum_{t=1}^{N} \tilde{X}_t' l_x l_y' \tilde{Y}_t \tag{12}$$

So the 2DPLS determines l_x and l_y by solving (10) with r_x and r_y fixed. The right transforms r_x and r_y are found by solving (11) with l_x and l_y are fixed. While given r_x and r_y, the optimization in (10) involves the following two eigenvalue decomposition problem:

$$\Sigma^r_{xy}\Sigma^r_{yx}l_x = \lambda^2 l_x \tag{13}$$
$$\Sigma^r_{yx}\Sigma^r_{xy}l_y = \lambda^2 l_y \tag{14}$$

where $\Sigma^r_{yx} = \Sigma^{r\ '}_{yx}$.

In a similar manner, given l_x and l_y, the optimization in (11) involves the following two eigenvalue decomposition problem:

$$\Sigma^l_{xy}\Sigma^l_{yx}r_x = \lambda^2 r_x \tag{15}$$
$$\Sigma^l_{yx}\Sigma^l_{xy}r_y = \lambda^2 r_y \tag{16}$$

where $\Sigma^l_{yx} = \Sigma^{l\ '}_{yx}$.

Left transforms(l_x and l_y) and right transforms(r_x and r_y) are determined by iteratively solving the aforementioned eigenvalue decomposition problems until convergence. And the four eigenvalue decompositions in itself can solved by two singular value decomposition problems for much smaller size matrices compared to the standard PLS.

Let $m = m_x * n_x$ and $n = m_y * n_y$ denote the dimensions of two feature matrices respectively, then the time complexity of PLS is $O(m^2n + mn^2 + n^3)$ if $m > n$. The time complexity of our methods is $O(t(m_x^2 n_x + m_x n_x^2 + n_x^3 + m_y^2 n_y + m_y n_y^2 + n_y^3))$ if $m_x > n_x$ and $m_y > n_y$ where t is the iterative time. In our numerical experiments, it takes almost five or six iterations for convergence. So t is usually small and the complexity will dramatically reduced although an iteration algorithm is introduced because $m_x, n_x << m$ and $m_y, n_y << n$. For example, when FERET is used in experiment, the computational complexity is $O(10^7)$ while PLS is $O(10^{12})$ if the face images are with size of $80 * 80$.

If we need reduce the image to a feature matrix with size of $d_1 \times d_2$, we will get the left transforms matrices $L_x \in R^{m_x \times d_1}$ and $L_y \in R^{m_y \times d_1}$ and the right transform matrices $R_x \in R^{n_x \times d_2}$ and $R_y \in R^{n_y \times d_2}$. The PLS components of a sample T with two needed two-dimensional images T_1, T_2 are calculated as: $Q_1 = L_x^{'} T_1 R_x$ and$Q_2 = L_y^{'} T_2 R_y$ and let the feature matrix be $Q = [Q_1; Q_2]$. The distance between any two feature matrices $Q_i = [Q_1^i; Q_2^i]$ and $Q_i = [Q_1^i; Q_2^i]$ is defined as:

$$D(Q_i, Q_j) = \sum_{k=1}^{d1} \|Q_1^i(k,:) - Q_1^j(k,:)\| + \|Q_2^i(k,:) - Q_2^j(k,:)\| \tag{17}$$

Then we can also use the classifier depicted in Section 2.

IV. Numerical Experiments

In this section, the performance of 2DPLS is evaluated on the PolyU Palmprint database [12] and FERET face image database [13], compared with the performances of standard PLS, PCA and $(2D)^2PCA$.

The PloyU Palmprint database contains 600 gray-scale images of 100 different palms with six samples for each palm

TABLE I
The Maximal Recognition Rates of PCA, PLS, $(2D)^2PCA$ and 2DPLS on a Subset of the PolyU Palmprint Database and Corresponding Dimensions

Method	PCA	PLS	$(2D)^2PCA$	2DPLS
Recognition Rate	86%	88.7%	95%	95%
Dimension	94	138	27	22

(http://www.comp.polyu.edu.hk/ biometrics/). Six samples from each of these palms were collected in two sessions, where the first three were captured in the first session and the other three in the second session. The average interval between the first and the second sessions is two months. Fig.1 shows some sample images of two palms.

In our experiments, the central part of each original image was automatically cropped using the algorithm mentioned in [14]. The cropped images were resized to 64×64 pixels. Performing the wavelet transforming on each original image using Daubechies orthonormal wavelet, the low-frequency image (LL1) with 32×32 resolution are used as the second feature matrices. The first three samples are used for training while the other three are used for testing. By combining two sets of feature matrices, we use the proposed 2DPLS algorithm and the nearest neighbor(NN) classifier depicted in Section 3. As a matter of convenience we set $d = d_1 = d_2$, so d is not larger than 32. In the recognition step of our experiments, the second set of extracted features is not considered because the second set of features is the wavelet transform of the first set of features and cannot carry more information than the first one. The maximal recognition rate of each method and the corresponding dimension are listed in table I.

From Table I, we can see that 2DPLS and $(2D)^2PCA$ outperform PLS and PCA and PLS is better than PCA. Although $(2D)^2PCA$ and 2DPLS have the same accuracy, 2DPLS can use much less feature coefficients. 2DPLS needs to store much more coefficients compared to 1D methods because we set the left transforms and right transforms have same dimensions. If the dimensions of left and right transforms are both carefully selected by cross validation, the coefficients in 2DPLS can much less.

The FERET face image database has become a standard database for testing and evaluating state-of-the-art face recognition algorithms. The proposed method is tested on a subset of the FERET database. This subset includes 1400 images of 200 individuals (each one has seven images). It is composed of the images whose names are marked with two-character strings:"ba","bj","bk","be","bf","bd",and"bg".This subset involves variations in facial expression,illumination, and pose. Some sample images of one person are shown in Fig.2

In our experiment, the facial portion of each original image is automatically cropped based on the location of eyes and mouth, and the cropped image was resized to 80×80 pixels. Performing the quadratic wavelet transforming on each original images using Daubechies orthonormal wavelet, the low-frequency image (LL2) with solution 20×20 are used as the

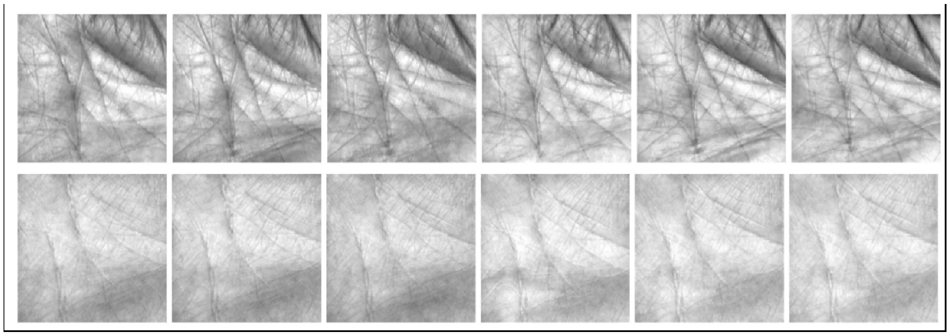

Fig. 1. Samples of the cropped images in the PolyU Palmprint database

Fig. 2. Seven images for one subject in FERET face database

second feature matrices. We use the first k samples for training and the remainder for testing and adopt the NN classifier in Section 3. The same dimension of left transforms and right transforms is also adopted. The final recognition results is listed in table II.

From Table II, we can see that 2DPLS overall outperforms PLS and $(2D)^2PCA$, irrespective of the variation in training sample size. And because preserving the structure information, $(2D)^2PCA$ is better than PLS. So, it is validated that two-dimensional methods have prominent superiority.

V. Conclusions and future work

In this paper, we have presented a two-dimensional extension of PLS that directly takes the image data as inputs without reshaping them into vectors, in order to correlate relationships between them. The main advantage of 2DPLS is two fold: 1) low computational complexity; 2) preserving spatial structure of image data in the calculation of canonical variable matrices.

The conjugate orthogonal property is very important in feature extraction. But there is still no clear similar concept in two-dimensional methods. How to develop and construct a perfect 2D-PLS framework for pattern recognition is our future work, such as two-dimensional conjugate orthogonal PLS.

Acknowledgment

This work was supported by the National Science Foundation of China, the National Science Foundation of Jiangsu Province and Doctoral Fund of Ministry of Education of China(RFDP) under grant no. 60773172, BK2008411 and 200802880017.

References

[1] Roman Rosopal and Nicole Krämer. Overview and Recent Advances in Partial Least Squares. *Subspace Latent Structure and Feature Selection. In: Statistical and Optimization Perspectives Workshop, SLSFS 2005. Revised Selected Papers (Lecture Notes in Computer Science Vol.3940)*,2006 pp.34-51.

[2] Wencong Lu, Nianyi Chen, Guozheng Li and Jie Yang. Multitask Learning Using Partial Least Squares Method. *Seventh International Conference on Information Fusion 2004* Stockholm, Sweden, 2004, Vol.1, pp.79-84.

[3] Quan-Sen Sun, Zhong Jin, Pheng-Ann Heng, and De-Shen Xia. A Novel Feature Fusion Method Based on Partial Least Squares Regression. *The third International Conference on Advances in Pattern Recognition(Bath,UK), lecture Notes in Comoputer Science*,Springer-Verlag, Heidelberg,Berlin,2005,3686:268-277.

[4] Jangsun Baek, Minsoo Kim. Face Recognition using Paritial least squares components. *Pattern Recognition*, 37(6):1303-1306,2004.

[5] Charanpal Dhanjal, Steve R. Gunn, and John Shawe-Taylor. Efficient Sparse Kernel Feature Extraction Based on Partial Least Squares *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 31(8): 1347-1361, 2009.

[6] Xiujuan Li, Jie Ma and Shutao Li. A Novel Face Recognition Method Based on Principal Component Analysis and Kernel Partial Least Squares. *IEEE International Conference on Robotics and Biomimetics 2007* Sanya, China, 2007, pp.1773-1777.

[7] Yoon Ho Bang, Chang Kyoo Yoo and In-Beum Lee. Nonlinear PLS modeling with fuzzy inference system. *Chemometrics and Intelligent laboratory systems.* 64(16):137-155, 2003.

[8] Jian Yang, David Zhang. Two-Dimensional PCA: A New Appoach to Appearance-Based Face Representation and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 26(1):131-137, 2004.

[9] Daoqiang Zhang, Zhihua Zhou. $(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. *NeuroComputing.* 69(13):224-231,2005.

[10] Sun Ho Lee, Seungjin Choi. Two-Dimensional Canonical Correlation Analysis. *IEEE Signal Processing Letters.* 14(10):735-738, 2007.

[11] Maolong Yang, Quansen Sen, Deshen Xia. Conjugate orthonormalized partial least squares regression and its application in image recognition. *Computer Engineering and Applications.* 44(29):36-39,2008.

[12] David Zhang, Wai-Kin Kong, Jane You and Michael Wong. On-line palmprint identification, *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 25(9):1041-1050, 2003.

TABLE II

THE MAXIMAL RECOGNITION RATES OF PLS, $(2D)^2PCA$ AND 2DPLS ON A SUBSET OF THE FERET DATABASE AND CORRESPONDING DIMENSIONS (SHOWN IN PARENTHESES)

Accuracy	k=1	k=2	k=3	k=4
PLS	43.3%(68)	55.9%(66)	52.4%(68)	70.2%(42)
$(2D)^2PCA$	45.6%(9)	58.8%(11)	55% (10)	72.5%(11)
2DPLS	48%(12)	60%(13)	56.8% (10)	77.2%(10)

[13] Phillips P.J., Moon H.J., Rizvi S.A., Rauss P.J.. The FERET evaluation methodology for face recognition algorithms. *IEEE Transacation on Pattern Analysis and Mahine Intelligence.* 20(10):1090-1104, 2000.

[14] D.Zhang.*Palmprint Authentication* . Kluwer Academic, 2004.