Coupled Subspaces Analysis

Dong Xu[1,2], Shuicheng Yan[1], Lei Zhang[1] , ZhengKai Liu[2], Hongjiang Zhang[1]
[1]Microsoft Research Asia, Beijing, P. R. China
[2]Department of Electronic Engineering and Information Science
University of Science and Technology of China, Hefei, Anhui, P. R. China

Abstract

Subspace learning is a fundamental approach for face recognition and facial expression analysis. In this paper, we propose a novel subspace analysis scheme for the two applications. Unlike the traditional subspace algorithms, such as PCA and LDA, in which an image is treated as a vector; in our scheme, an image is directly treated as a 2D matrix, and a new criterion is proposed to infer two low dimensional coupled subspaces that optimally reconstruct the original matrices from row and column directions collaboratively. An efficient approach, namely Coupled Subspace Analysis (CSA), is applied to learn these two subspaces in an iterative manner. Then we reveal the essence of each step in CSA and propose an approach to select the dimension numbers for these two subspaces with the given rate of information lost. Moreover, we prove that PCA and the recently proposed 2DPCA are just simplified special cases of CSA and answer the unsolved theoretical problems in 2DPCA. The main contributions of this paper include: 1) for both face recognition and facial expression analysis, we propose a novel image matrix based scheme, and obtain a much lower dimensional face representation for subsequent discriminant analysis; 2) CSA effectively alleviates the curse of dimensionality dilemma and small sample size problem existed in face recognition problem; and 3) CSA clarifies the essence of 2DPCA and explains the superiority of 2DPCA compared with PCA. The extensive experiments on both face recognition and facial expression analysis demonstrate that CSA is superior to the classical algorithms.

1. Introduction

Face recognition and facial expression analysis have been two active research topics for decades, due to their potential applications in human machine interfaces, image/video analysis and et al. The face recognition problem can be classified into three types, *i.e.* verification, identification and watch list 20. Verification is to solve the problem “am who I say I am”, identification is for the question “who am I” and watch list is the task of “are you looking for me”. More specifically, for the verification task, a person claims his identity to a face recognition system, and the system then compares the presented biometric with the stored biometric of the claimed identity and decides to either accept or reject the claim. In the identification task, an image of an unknown person is provided to a system, and then the system compares the unknown image to each image of known people in the database to present a ranked listing of the top *n* “candidates”. Watch list task is more difficult than the above two tasks. Firstly face recognition system needs to determine whether an individual is in the watch list; and then if the individual is in the watch list, the system should identify the individual. For facial expression analysis, there are two different tasks: Facial Action Unit Recognition and Prototypic Emotional Expression Recognition 22. The former is to describe the subtle change of facial components, while the latter is to recognize a small set of prototypic emotional expressions, such as disgust, fear, joy, surprise, sadness and anger.

Many algorithms have been proposed for face recognition 12910151824 25262729 and facial expression analysis 35671213 142223283031. The related comprehensive surveys can be found in 411732. Among all these algorithms, the linear subspace learning algorithms, such as PCA, LDA and ICA, are the most popular ones for both applications. For face recognition problem, Turk *et al.* 24 applied Principal Component Analysis (PCA/ Eigenface), Belhumeur *et al.* 2 used Linear Discriminant Analysis (LDA), and Bartlett *et al.* implemented Independent Component Analysis (ICA) 1. For facial expression analysis problem, Lyons *et al.* 12 applied LDA, Buciu *et al.* 3 used ICA, X. Chen 5 applied variant LDA for Prototypic Emotional Expression Recognition and Donato *et al.* 6 applied PCA, LDA and ICA for Facial Action Unit Recognition. PCA 24 applies Karhunen-Loeve transformation to derive the most expressive subspace for face representation and decorrelates the input data using the second-order statistics, while ICA 1 minimizes both the second and higher-order dependencies of the data. Unlike the unsupervised algorithms PCA and ICA, LDA 2 is a supervised learning algorithm and aims at pursuing a set of features that can best distinguish different object classes.

In all aforementioned subspace learning algorithms, a face image matrix is typically transformed to a vector by concatenating all row vectors, which usually results in some serious problems in practical applications. Firstly, the intrinsic spatial structure information is lost. Secondly, the feature dimension is extremely high even in moderate cases, which will result in the *curse of the dimensionality* dilemma. Finally, in many cases, the available number of training samples is relatively very small compared to the feature dimension, which will make the algorithms suffer from the *small sample size* problem.

Recently, Yang *et al.* 29 proposed an algorithm called 2DPCA for face recognition, in which the image covariance (scatter) matrix as in PCA/Eigenface is directly computed from the image matrix representations. However, as the authors stated in 29 , there are still three fundamental questions not solved for 2DPCA: one is whether the eigenvalues have the same characteristics as in Eigenface; another one is why 2DPCA can outperform Eigenface; and the last one is that it is still unclear how to directly reduce the dimension of 2DPCA.

In this paper, we directly treat an image as a two dimensional matrix and the image spatial structure information is explicitly utilized for face recognition and facial expression analysis. Firstly, we propose a novel image reconstruction criterion to reconstruct the original image matrices with two low dimensional coupled subspaces in the sense of least square error. These two subspaces encode the row and column information of the image matrices, respectively, which is different from traditional algorithms that encode all the information in one subspace. To obtain the optimal solution, it needs to solve a biquadratic programming problem with biquadratic constraint, yet there is no closed-form solution. In this work, an iterative approach, called Coupled Subspace Analysis (CSA), is proposed to pursue the local optimum of the new criterion. In each sub-step of CSA, the optimization criterion is changed to an eigenvector decomposition problem as in Eigenface. As analyzed later in the paper, each sub-step of CSA is intrinsically a specialized PCA in which the row/column vectors of the image matrices are considered as the new objects to be analyzed and the Principle Component Analysis is conducted on these new vectors 2627.

Furthermore, we investigate PCA and 2DPCA from a novel perspective and reveal the relationship among PCA, 2DPCA and CSA. We prove that PCA and 2DPCA are just the simplified special cases of CSA. The proposed CSA has the advantages of PCA and 2DPCA algorithms and meanwhile throws away the disadvantages of them. More specifically, CSA pursues the low dimensional representation aiming at reconstructing the original image set as PCA does, and it can remove the intrinsic redundancies in row and column vectors of the image. A much lower dimensional image representation is acquired after CSA, thus LDA can be used directly to further improve the performance. In this way, CSA overcomes the drawback of 2DPCA. As the object to be analyzed is the row/column vector, and the object set in each step is significantly enlarged, CSA avoids the curse of dimensionality dilemma and the small sample size problem. Consequently, we answer the three questions mentioned in 2DPCA 29. For the first problem, as 2DPCA is a special case of CSA, its eigenvalues have the same meaning as in PCA, *i.e.* the larger the eigenvalue is, the more important its corresponding eigenvector is in reconstructing the original image. For the second problem, as 2DPCA can also avoid the curse of dimensionality dilemma and the small sample size problem as CSA, it is superior to PCA. For the third question, our proposed CSA is the method that directly reduces the dimensions based on the image matrix representations.

The remainder of this paper is organized as follows. In Section 2, we introduce the coupled subspaces based criterion for image reconstruction and present the iterative algorithm for the optimization of the new criterion. In Section 3, we study the relationship between CSA and PCA, 2DPCA. In Section 4, the exhaustive experiments on face recognition and facial expression analysis are presented to demonstrate the effectiveness of the proposed CSA algorithm. Finally, we conclude this paper in Section 5.

2. Coupled Subspaces Analysis

Before describing the Coupled Subspace Analysis algorithm, we give some terminologies on matrix operations. The inner product of two matrices $A \in \mathbb{R}^m \times \mathbb{R}^n$ and $B \in \mathbb{R}^m \times \mathbb{R}^n$ is defined as $\langle A, B \rangle \triangleq \sum_{i=1}^m \sum_{j=1}^n A_{ij} B_{ij}$ and the Frobenius norm of the matrix A is defined as $\| A \|_F = \sqrt{\langle A, A \rangle}$. Let $X_i \in \mathbb{R}^m \times \mathbb{R}^n$, $i = 1, \ldots N$ be the training samples, where N is the total number of training images. The samples are assumed to be zero centered, i.e. $\sum_{i=1}^N X_i = 0_{m \times n}$.

2.1. Optimal Matrix Reconstruction Criterion

Denote matrix $Y_i \in \mathbb{R}^{m'} \times \mathbb{R}^{n'}$ be the lower-dimensional matrix representation of sample X_i derived from two projection matrices $U \in \mathbb{R}^m \times \mathbb{R}^{m'}$ and $V \in \mathbb{R}^n \times \mathbb{R}^{n'}$, *i.e.* $Y_i = U' X_i V$. Then, the optimal matrices U and V that best reconstruct the original matrices in the sense of least square error should satisfy the following objective function, i.e. *Optimal Matrix Reconstruction Criterion*:

$$(U^*, V^*) = Arg \underset{U,\,V}{Min} \sum_i \| U Y_i V' - X_i \|_F^2 \qquad (1)$$

2.2 Coupled Subspaces Analysis

The objective function in Eq. (1) is biquadratic and has no closed-form solution. Therefore, we design an iterative procedure to obtain the local optimal solution.

For given $U \in \mathbb{R}^m \times \mathbb{R}^{m'}$, the objective function of Eq. (1) can be rewritten as

$$V^* = Arg\,\underset{V}{Min} \sum_i \| X_i^U VV' - X_i \|_F^2 \qquad (2)$$

where $X_i^U = UU'X_i$. As proved in the following *theorem*-1, the solution of Eq (2) is the eigenvectors of the eigen-decomposition problem $FF'x = \lambda x$ with

$$F = [X_1^U(1,*)', ..., X_1^U(m,*)', X_2^U(1,*)', ..., ...X_N^U(m,*)'] \qquad (3)$$

where $X_i^U(r,*)$ is the *r-th* row of the image matrix X_i^U and F is the concatenated matrix of all dimension-reduced samples X_i^U.

Similarly, for given $V \in \mathbb{R}^n \times \mathbb{R}^{n'}$, the optimization problem in Eq. (1) is changed to

$$U^* = Arg\,\underset{U}{Min} \sum_i \| UU'X_i^V - X_i \|_F^2 \qquad (4)$$

where $X_i^V = X_i VV'$. And as proved in the following *theorem*-1, the solution of Eq (4) is the eigenvectors of the eigen-decomposition problem $GG'x = \lambda x$ with

$$G = [X_1^V(*,1), ..., X_1^V(*,n), X_2^V(*,1), ..., ...X_N^V(*,n)] \qquad (5)$$

where $X_i^V(*,c)$ is the *c-th* column of image matrix X_i^V.

By iteratively optimizing the objective function with respect to U and V, respectively, we can obtain a local optimum of the solution. The whole procedure, namely Coupled Subspace Analysis, is listed as in Figure 1.

Coupled Subspaces Analysis: Given the sample set $\{X_i \in \mathbb{R}^m \times \mathbb{R}^n,\ i=1,...,N\}$ and the final reduced dimensions (m', n').

1. Initialization: $U_0 = I_m$.
2. For t=1,2,…, T_{max}, Do
   a) For given U_{t-1}, $X_i^U \Leftarrow U_{t-1}U_{t-1}'X_i$
   $F = [X_1^U(1,*)', ..., X_1^U(m,*)', X_2^U(1,*)', ..., ...X_N^U(m,*)']$
   Compute the optimal projection matrix V_t as the first m' leading eigenvectors of symmetry matrix FF'.
   b) For given V_t, $X_i^V \Leftarrow X_i V_t V_t'$
   $G = [X_1^V(*,1), ..., X_1^V(*,n), X_2^V(*,1), ..., ...X_N^V(*,n)]$
   Compute the optimal projection matrix U_t as the first n' leading eigenvectors of symmetry matrix GG'.
   c) If $t > 2$ and
   $$\left\langle \begin{array}{l} \| U^k - U^{k-1} \|_F < m * \varepsilon \\ \| V^k - V^{k-1} \|_F < n * \varepsilon \end{array} \right.$$
   Go to step 2; else, continue.
3. Output the final coupled projection matrices $U = U_t \in \mathbb{R}^{m \times m'}$ and $V = V_t \in \mathbb{R}^{n \times n'}$.

Figure 1. Coupled Subspaces Analysis Procedure

2.2. Algorithmic Analysis and Justification

In this subsection, we discuss and reveal the essence of each sub-step of our proposed Coupled Subspace Analysis algorithm. First, we solve the problem in each sub-step of CSA with the *theorem*-1.

Theorem 1. For given $U \in \mathbb{R}^m \times \mathbb{R}^{m'}$, the solution of objective function (2) is the leading eigenvectors of the symmetry matrix FF'; and for given $V \in \mathbb{R}^n \times \mathbb{R}^{n'}$, the solution of the objective function (4) is the leading eigenvectors of the symmetry matrix GG'.

Proof. We take the case with given $V \in \mathbb{R}^m \times \mathbb{R}^{m'}$ as example to prove the theory and another case can be proved in the same way. Denote $f(U) = \sum_i \| UU'X_i^V - X_i \|_F^2$, then we have

$$\begin{aligned} f(U) &= \sum_i \| UU'X_i VV' - X_i VV' + X_i VV' - X_i \|_F^2 \\ &= \sum_i \| UU'X_i VV' - X_i VV' \|_F^2 + \| X_i VV' - X_i \|_F^2 \\ &= \sum_i \| UU'X_i VV' - X_i VV' \|_F^2 + c_1 \\ &= \sum_i \| UU'X_i^V - X_i^V \|_F^2 + c_1 \\ &= \sum_i Tr[(X_i^V{}'UU' - X_i^V{}')(UU'X_i^V - X_i^V)] + c_1 \\ &= \sum_i Tr[-X_i^V{}'UU'X_i^V + X_i^V{}'X_i^V] + c_1 \\ &= -Tr[U'(\sum_i X_i^V X_i^V{}')U] + c_2 \\ &= -Tr(U'GG'U) + c_2 \end{aligned} \qquad (6)$$

The second equality is obtained owing to the orthogonality of the columns of the projection matrix V; *i.e.* $S_1 = UU'X_i VV' - X_i VV'$ lies in the sub-space spanned by V; while $S_2 = X_i VV' - X_i$ is the reconstruction error, which is orthogonal to S_1. The third equality is derived because the projection matrix V is assumed known in this sub-step. The fifth equality is obtained because $\| A \|_F^2 = Tr(AA') = Tr(A'A)$, where $Tr(\cdot)$ is the trace of a matrix. The sixth equality stands because $U'U = I$. And the last equality is obtained as

$$\begin{aligned} \sum_i X_i^V X_i^V{}' &= \sum_i \sum_c X_i^V(*,c) X_i^V(*,c)' \\ &= \sum_{i,c} X_i^V(*,c) X_i^V(*,c)' = GG' \end{aligned} \qquad (7)$$

In the derivation,

$$c_2 = c_1 + Tr(\sum_i X_i^V X_i^V{}') \\ = \| X_i VV' - X_i \|_F^2 + Tr(\sum_i X_i^V X_i^V{}') \tag{8}$$

From Eq. (6), the optimal U of function $f(U)$ is the first m' eigenvectors of the symmetry matrix GG' and can be obtained by solving the eigenvector decomposition problem as

$$GG'u = \lambda u \quad s.t. \quad \|u\| = 1 \tag{9}$$

Thus, the solution of the objective function (4) is the leading eigenvectors of the symmetry matrix GG'. ■

From *theorem*-1, we can easily obtain the following corollary which reveals the essence of each sub-step of CSA.

Corollary 1. *For each sub-step in Coupled Subspace Analysis, the optimal projection matrix is obtained from the Singular Value Decomposition by taking each row/column vector of each image matrix as a new object to be analyzed.*

Proof. Similarly to the proof of *theorem*-1, we take the case with given $V \in \mathbb{R}^n \times \mathbb{R}^{n'}$ as example to prove the corollary. From *theorem*-1, the optimal projection matrix in each sub-step is the leading eigenvector of GG'. If we consider G in a different perspective, each column of G is a column vector of the reconstructed image X_i^V and G can be considered as a new sample matrix with each column as an object. Denote the Singular Value Decomposition of G as

$$G = W_1 S W_2 \tag{10}$$

where U and W are orthonormal matrices and S is a diagonal matrix. Then, we have

$$GG'W_1 = W_1 S^2 \tag{11}$$

which means that the leading column vectors in Singular Value Decomposition matrix W_1 are the solution for optimal projection matrix U. ■

In the following, we introduce how to determine the proper dimension numbers with given rate of information loss as in *theorem-2*.

Theorem 2. (*Dimensions Selection*) For any given $\varepsilon > 0$ (assumed less than 0.5), if the retained row and column numbers in the first step are determined by making the energy/information loss rate not larger than ε_1 which satisfies (ε_1 +2) $\varepsilon_1 \le \varepsilon$, then the total information/energy loss rate for final results will not be larger than ε .

Proof. From the analysis in *theorem*-1, $f(U,V) = \sum_i \| UU' X_i VV' - X_i VV' \|_F^2 + \| X_i VV' - X_i \|_F^2$. It is obvious that the information loss is not increased in each iteration; therefore, $f(U,V) \le \sum_i \| U_1 U_1' X_i V_1 V_1' - X_i V_1 V_1' \|_F^2 + \| X_i V_1 V_1' - X_i \|_F^2$. As the retained row and column numbers are determined by satisfying the energy/information loss rate which is less than ε_1 in the first step, then $f(U,V) \le \varepsilon_1 \sum_i \| X_i V_1 V_1' \|_F^2 + \varepsilon_1 \sum_i \| X_i \|_F^2$ and $\sum_i \| X_i V_1 V_1' - X_i \|_F^2 \le \varepsilon_1 \sum_i \| X_i \|_F^2$, then, $\sum_i \| X_i V_1 V_1' \| \le (\varepsilon_1 + 1) \sum_i \| X_i \|_F^2$. Then, $f(U,V) \le (\varepsilon_1 + 2)\varepsilon_1 \sum_i \| X_i \|_F^2 \le \varepsilon \sum_i \| X_i \|_F^2$. Therefore, the rate of final lost information/energy is not larger than ε . And the dimension can be determined according to $\sum_{i=n'+1}^{n} \lambda_i^v \le \varepsilon_1 \sum_{i=1}^{n} \lambda_i^v$ and $\sum_{i=m'+1}^{m} \lambda_i^u \le \varepsilon_1 \sum_{i=1}^{m} \lambda_i^u$, where λ_i^u and λ_i^v are the eigenvalues of FF' and GG' in the first step. ■

3. Connections with PCA and 2DPCA

In this section, we discuss the relationship between the proposed CSA and the PCA, 2DPCA. Figure 2 illustrates the flowchart of the CSA algorithm. Firstly, as shown in Figure 2(a) and (b), the row vectors of the image matrices are considered as the objects to be analyzed, and Singular Value Decomposition is performed to learn the subspace in the row direction V with optimal reconstruction capability. Secondly, With the matrix V, the original image matrices in Figure 2(a) are projected to the low-dimensional matrices in Figure 2 (c) and then reconstructed, as in Figure 2 (d); and then, as in Figure 2(d) and (e), the column vectors of the image matrices are considered as the objects to be analyzed, and similarly, Singular Value Decomposition is applied to derive the optimal subspace in the column direction. Finally, the Figure 2(f) shows the low-dimensional matrices after the projection with U and V. then we reconstruct the original images with U and V again, and the algorithm continues to run until the procedure goes on until converged. The derived coupled subspaces collaboratively reconstruct the original images in the sense of least square error.

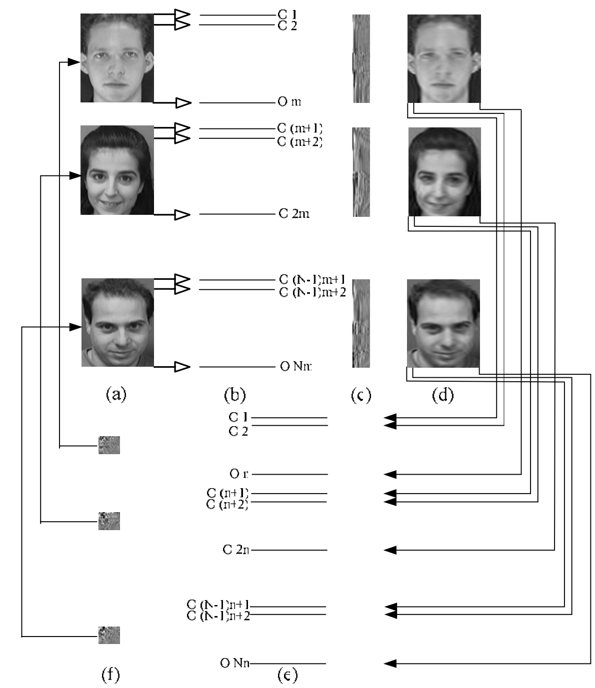

Figure 2. The flowchart of Coupled Subspaces Analysis (a) the original image X_i (b) all the rows of the images are treated as the objects, i.e. F' (c) the low-dimensional representation after the projection with V (d) The reconstructed images X_i^V with V (e) all the columns of the images are treated s the objects, i.e. G' (f) the low dimensional representation Y_i after the projection with U and V.

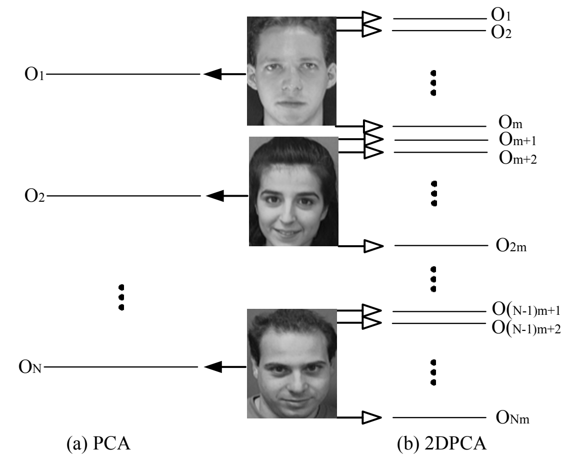

Figure 3. Object representations of Eigenface and 2DPCA

3.1. Connection with PCA

In PCA, all the pixels in one image are concatenated as a vector as shown in Figure 3(a). The principal components are the leading eigenvectors of the covariance matrix. It can be proved that PCA is a special case of CSA as follows.

Claim1. Principal Component Analysis is a special case of Coupled Subspace Analysis algorithm with n=1.

Proof: As n=1, the optimal V is 1, and the matrix X_i can be directly represented as a vector $x_i = X_i$. Here, we assume the data is not centered and $\overline{X} = \overline{x}$ are the mean vector; then the objective of CSA is:

$$\begin{aligned}&\sum_i \| UY_iV' - (X_i - \overline{X}) \|_F^2 \\ &= \sum_i \| UU'(x_i - \overline{x}) - (x_i - \overline{x}) \|^2 \\ &= -Tr[U' \sum_i (x_i - \overline{x})(x_i - \overline{x})' U] + c_3 \\ &= -Tr(U'CU) + c_3\end{aligned} \tag{23}$$

where c_3 is a constant. Therefore, CSA with n=1 is equal to the traditional PCA algorithm.

3.2. Connection with 2DPCA

2DPCA was recently proposed by Yang et. al 29. It treats the input image as a matrix and replaces the vector with matrix to compute the covariance matrix.

$$C_{2D} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{M})^T (X_i - \overline{M}) \tag{33}$$

However, there are three questions unsolved in 2DPCA 29: 1) what is the meaning of the largest eigenvectors of the computed covariance matrix; 2) why 2DPCA outperforms PCA; and 3) how to continue reduce the dimensionality. Here, we systematically answer these questions in the following claim 2.

Claim2. 2DPCA is a special case of Coupled Subspace Analysis algorithm with fixed U= I.

Proof. Similar to the proof of claim 1, we assume the data is not centered and $\overline{X}$ are the total mean matrix of all the samples, i.e. $\overline{X} = 1/N \sum_{i=1}^{N} X_i$. With given fixed U = I, the objective function of CSA can be rewritten as

$$\begin{aligned}&\sum_i \| UY_iV' - (X_i - \overline{X}) \|_F^2 \\ &= \sum_i \| (X_i - \overline{X})VV' - (X_i - \overline{X}) \|^2 \\ &= -Tr[V' \sum_i (X_i - \overline{X})'(X_i - \overline{X})V] + c_4 \\ &= -Tr(U' C_{2D} U) + c_4\end{aligned} \tag{44}$$

where c_4 is a constant. That is, CSA with U=I is the 2DPCA algorithm. ■

As the 2DPCA is a special CSA algorithm, the leading eigenvectors are the optimal components to reconstruct the original image matrix in the sense of Frobenius norm, which is similar to the eigenvectors of the PCA algorithm. On the other hand, 2DPCA shares the some characteristics of CSA, and thus it can avoid the curse of dimensionality and the small sample problem. Therefore, 2DPCA has the potential of being superior to PCA. Furthermore, the third question is solved by using the proposed CSA to reduce the dimensions from both row and column directions.

3.3. Discussions

As described in the above two subsections, PCA, 2DPCA are special cases of the proposed CSA algorithm. More strictly speaking, they are simplified versions of CSA, thus they still have some limitations.

In PCA, all the pixels in one image are concatenated as a vector, which usually results in the well known *curse of dimensionality* dilemma and the small sample size problem. 2DPCA overcomes the above two problems to some extent. However, as stated in 29, the fundamental theory for why to do so and why it has the superiority was not presented.

The proposed CSA is motivated from the optimal matrix reconstruction criterion and its purpose is to pursue two coupled subspaces to reconstruct the original sample image matrices in the sense of Frobenius norm. As a general method, it reveals the essence of the 2DPCA algorithm and clearly explains why 2DPCA is able to outperform traditional PCA. Moreover, compared with 2DPCA, CSA has the following advantages: 1) CSA removes the redundant information and noise in both row and column vector directions; while 2DPCA only considers the row vector direction; 2) CSA can derive a much lower dimensional representation than 2DPCA, which makes the following LDA convenient for further supervised learning.

4. Experiments

In this section, we compare the proposed Coupled Subspaces Analysis algorithm with other classical subspace algorithms for face recognition and facial expression analysis. For face recognition, two databases CMU PIE 21 and ORL 16 are used, and for facial expression analysis, JAFFE database 11 is used.

4.1. Experiments for Face Recognition

In this sub-section, the CSA is evaluated in different scenarios with pose, illumination and expression variations as well as the small number of samples problem. ORL database contains 400 images of 40 individuals. Some images were captured at different times and have different variations including expression (open or closed eyes, smiling or non-smiling) and facial details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20 degrees. All images are grayscale and normalized to a resolution of 56*46 pixels. Five sample images of one individual in the ORL database are displayed in Figure 4. Data set is randomly partitioned into gallery and probe sets with given sample numbers. We conducted three sets of experiments with the training samples for each individual varying from 4 to 2.

Figure 4. Five images in the ORL database

The CMU PIE (Pose, Illumination and Expression) database contains more than 40,000 facial images of 68 individuals. The images were acquired across different poses, under variable illumination conditions and with different facial expressions. In our experiments, we use two sub-databases. In the first sub-database PIE-I, we use five near frontal poses (C27, C05, C29, C09 and C07) and illumination 08 and 11. The flash 08 and 11 are placed near the center and the illumination can be considered as the nearly frontal illumination. Due to the data collection problem, we can only use 63 persons. The preprocessing steps include: 1) all ten images are aligned by fixing the locations of two eyes, and the images are resized to 64*64 pixels; 2) Histogram equilibrium is applied to remove the illumination affection. Figure 5 shows five examples of one individual. Data set is randomly partitioned into gallery and probe sets with given sample numbers. We conducted three sets of experiments with the training samples for each person varying from 4 to 2. In the second sub-database PIE-II, we choose the same five poses and the illumination 08, 10, 11 and 13. The flash 10 is placed at the one-quarter left profile side and flash 13 is placed at the one-quarter right profile side, so this database is more difficult. Thus there are 63 individuals, and each individual has 20 images. The same preprocessing strategies are applied. We also conducted three sets of experiments with the training samples for each individual varying from 6 to 4.

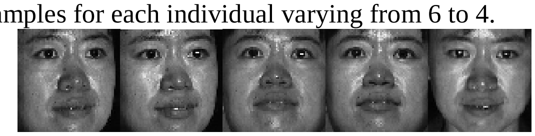

Figure 5. Five images in the PIE database

4.1.1. CSA vs. Eigenface and 2DPCA

In this part, we compare the performance of CSA with Eigenface and 2DPCA on CMU PIE and ORL databases. The top-one recognition results are shown in the Table 1-3. From these results, we find that the proposed CSA consistently outperforms Eigenface and 2DPCA.

Table 1. The top-one recognition rate of CSA, Eigenface and 2DPCA on ORL database

	G4/P6	G3/P7	G2/8
Eigenface	87.9%	84.6%	76.9%
2DPCA	91.7%	89.3%	81.9%
CSA	92.1%	90.7%	84.7%

Table 2. The top-one recognition rate of CSA, Eigenface and 2DPCA on PIE_I database

	G4/P6	G3/P7	G2/8
Eigenface	70.1%	70.1%	56.9%
2DPCA	78.0%	77.3%	62.5%
CSA	79.6%	79.1%	64.5%

Table 3. The top-one recognition rate of CSA, Eigenface and 2DPCA on PIE-II database

	G4/P16	G5/P15	G6/14
Eigenface	43.4%	49.2%	49.2%
2DPCA	46.9%	52.1%	50.6%
CSA	60.0%	65.1%	63.2%

We also plot the top-one recognition rate of G2/8 in ORL database with the number of selected eigenvectors of Eigenface in Figure 6, 2DPCA in Figure 7 and CSA in Figure 8.

4.1.2.CSA+LDA vs. Fisherface

We also compare the performance of the proposed CSA+LDA with traditional Fisherface. The top-one recognition rates on ORL and CMU PIE database are shown in Table 4-6. In Fisherface, the dimension in PCA step is set as $N-c$ and in LDA step it is set as $c-1$. For fair comparison, we also report the result in PCA+LDA by exploring all PCA dimension no larger than $N-c$ and LDA dimension no larger than $c-1$.

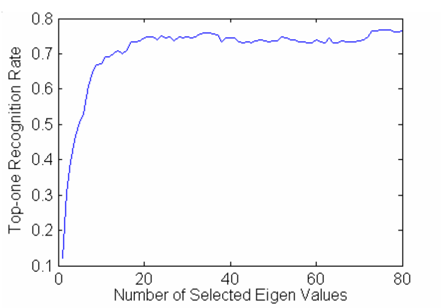

Figure 6. The top-one recognition rate of Eigenface with different number of selected eigenvectors

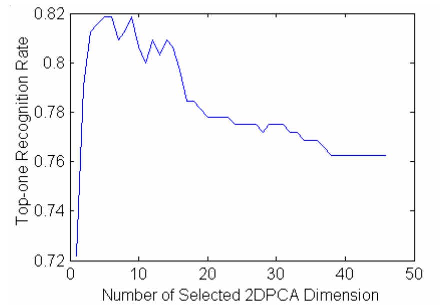

Figure 7. The performance of 2DPCA with different number of selected eigenvectors

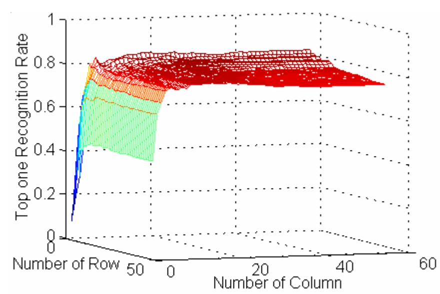

Figure 8. The performance of CSA with different number of selected eigenvectors

Table 4. The top-one recognition rate of Fisherface, PCA+LDA and CSA+LDA on ORL database

	G4/P6	G3/P7	G2/8
Fisherface	87.9%	86.4%	71.9%
PCA+LDA	93.8%	91.4%	82.2%
CSA+LDA	97.1%	94.6%	84.1%

Table 5. The top-one recognition rate of Fisherface, PCA+LDA and CSA+LDA on PIE_I database

	G4/P6	G3/P7	G2/8
Fisherface	88.1%	82.1%	53.8%
PCA+LDA	93.4%	91.6%	71.6%
CSA+LDA	97.1%	98.2%	81.2%

Table 6. The top-one recognition rate of Fisherface, PCA+LDA and CSA+LDA on PIE-II database

	G4/P16	G5/P15	G6/14
Fisherface	76.9%	83.6%	88.9%
PCA+LDA	79.1%	84.7%	91.5%
CSA+LDA	85.9%	88.7%	94.1%

The above experiments demonstrate the following interesting points: 1) although PCA, 2DPCA and CSA are all unsupervised approaches, different image representation may lead to different recognition performance due to the *curse of the dimension dilemma* and *small sample size problem* existed in face recognition problem; 2) Coupled Subspaces Analysis algorithm treats all row/column vectors of all the image matrices, thus it alleviates the above two problems. Moreover, CSA removes the redundancy among both row and column vectors, thus it can obtain a much lower dimensional matrix representation than 2DPCA; and 3) LDA can be used directly after CSA, and CSA+LDA shows much better performance compared with the classical Fisherface algorithm.

4.2. Experiments for Facial Expression Recognition

In this part, we conduct the facial expression recognition experiments on JAFFE 11 database. JAFFE database contains 213 images of female facial expression. Ten expressers were asked to pose seven different facial expressions (anger, disgust, fear, happiness, neutral, sadness and surprise). In this experiment, each image is resized to 64*64 according to the locations of the two eyes. Histogram equilibrium is also applied as the preprocessing method. As in 6, the analysis is performed on the different images, obtained by subtracting the average neutral images from each of the other expression images, so we only need to recognize six expressions. A similar leave-one-out strategy is applied in the experiments, *i.e.* each individual is in turn used for testing and other individuals are used as training samples. The experimental results for each individual are shown in Table 7. We can find that, in most cases, the proposed CSA is superior to PCA and 2DPCA algorithms; and CSA+LDA consistently outperforms Fisherface and also performs better than PCA+LDA with only one exception.

Table 7. The top-one recognition rate of different algorithms on JAFFE database

	PCA	2DPCA	CSA
0	65.0%	70.0%	65.0%
1	52.6%	68.4%	68.4%
2	63.2%	68.4%	89.5%
3	47.1%	52.9%	58.8%
4	61.1%	72.2%	83.3%
5	44.4%	61.1%	61.1%
6	64.7%	47.1%	52.9%
7	38.9%	44.4%	55.6%
8	38.9%	33.3%	44.4%
9	63.2%	78.9%	68.4%

	Fisherface	PCA+LDA	CSA+LDA
0	70.0%	75.0%	85.0%
1	73.7%	79.0%	94.7%
2	89.5%	100.0%	100.0%
3	58.8%	76.5%	76.5%
4	72.2%	88.9%	94.4%
5	83.3%	83.3%	88.9%
6	70.6%	82.4%	88.2%
7	72.2%	100.0%	88.9%
8	72.2%	83.3%	83.3%
9	100.0%	100.0%	100.0%

5. Conclusions

In this paper, we have developed a novel algorithm, called Coupled Subspaces Analysis (CSA), for face recognition and facial expression analysis. In this algorithm, the images are directly treated as 2D matrices, and an optimal matrix reconstruction criterion is proposed to reconstruct the original image matrices using two coupled subspaces. To acquire the solution, an iterative procedure is presented to learn these two subspaces. The proposed CSA effectively utilized the intrinsic spatial structure information and overcomes the two problems for face recognition, *i.e.* the *curse of dimensionality* dilemma and *small sample size problem*. Moreover, with a deep analysis of PCA and 2DPCA algorithms, we proved that PCA and 2DPCA are the simplified special cases of our proposed algorithm. To the best of knowledge, CSA is the first work to study the subspace learning problem to reconstruct the original samples by integrating two coupled subspaces.

References

1. M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component Analysis," IEEE Trans. on Neural Networks, pp. 1450-1464, Vol. 13, No. 6, Nov. 2002.

2. P. N. Belhumeur, J. Hespanda and D. Kiregeman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” IEEE Trans. PAMI, Vol. 19, No. 7, pp. 711-720, Jul. 1997.

3. I. Buciu, C. Kotropoulos and I. Pitas, “ICA and Gabor Representation for Facial Expression Recognition,” Proceedings of ICIP, 2003.

4. R. Chellappa, C.L. Wilson and S. Sirohey, “Human and Machine Recognition of Faces: A Survey,” Proceedings of the IEEE, Vol. 83, pp. 705-741, May 1995.

5. X. Chen and T.S. Huang, “Facial Expression Recognition: a Clustering based Approach,” Pattern Recognition Letters, pp. 1295-1302, 2003.

6. G. Donato, S. Bartlett, J.C. Hager, P. Ekman and T.J. Sejnowski, “Classifying Facial Actions,” IEEE Trans. PAMI, Vol. 21, No. 10, pp. 974-989,Oct. 1999.

7. I. Essa, and A. P. Pentland, “Coding, Analysis, Interpretation, and Recognition of Facial Expression,” IEEE Trans. PAMI, Vol. 19, No. 7, pp. 757-763, Jul. 1997.

8. B. Fasel, and J. Luettin, “Automatic Facial Expression Analysis: a survey,” Pattern Recognition, Vol. 36, pp. 259-275, 2003.
9. X. F. He, S. C. Yan, Y. X. Hu and H. J. Zhang, “Learning a Locality Preserving Subspace for Visual Recognition,” Proceedings of ICCV, 2003.
10. X. F. He, S. C. Yan, Y. X. Hu, H. F. Liu and H. J. Zhang, "Spectral Analysis for Face Recognition", Proceedings of ACCV, 2004.
11. M. Lyons, S. Akamatsu and J. Gyoba, “Coding Facial Expressions with Gabor Wavelets,” Proceedings of Automatic Face and Gesture Recognition, 1998.
12. M. Lyons, J. Budynek and S. Akamatsu, “Automatic Classification of Single Facial Images,” IEEE Trans. PAMI, Vol. 21, No. 12, pp. 1357-1362, Dec. 2003.
13. A.M. Martinez, “Recognizing Expression Variant Faces From a Single Sample Image per Class,” Proceedings of CVPR, 2003.
14. S. Mitra and Y. Liu, “Local Facial Asymmetry for Expression Classification,” Proceedings of CVPR, 2004.
15. B. Moghaddam, T. Jebara and A. Pentland, “Bayesian Face Recognition,” Pattern Recognition, Vol. 33, pp. 1771-1782, 2000.
16. Olivetti & Oracle Research Laboratory, The Olivetti & Oracle Research Laboratory Face Database of Faces, http://www.cam-orl.co.uk/facedatabase.html.
17. M. Pantic, and L.J.M. Rothkrantz, “Automatic Analysis of Facial Expression: the State of the Art,” IEEE Trans. PAMI, Vol. 22, No. 12, pp. 1424-1445, Dec. 2000.
18. P. Penev and J. Atick, “Local Feature Analysis: A General Statistical Theory for Object Representation,” Network Comput. Neural Syst., pp. 477-500, 1996.
19. P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss, “The Ferret Evaluation Methodology for Face-Recognition Algorithms," IEEE Trans. on PAMI, Vol. 22, No. 10, pp. 1090-1104, Oct. 2000.
20. P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, M. Bone, “Face Recognition Vendor Test: Evaluation Reprot,” http://www.frvt.org, Mar. 2003.
21. T. Sim and S. Baker, “The CMU Pose, Illumination, and Expression Database,” IEEE Trans. PAMI, Vol. 25, No. 12, pp. 1615-1617, Dec. 2003.
22. Y. Tian, T. kanade, and J.F. Cohn, “Recognition Action Units for Facial Expression Analysis,” IEEE Trans. PAMI, Vol. 23, No. 2, pp. 97-115, Feb. 2001.
23. Y. Tian, T. Kanade and J. F. Cohn, “Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity,” Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2002.
24. M. Turk and A. Pentland, “Eigenfaces for Recognition,” J. Of Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86, 1991.
25. L. Wiskott, J.M. Fellous, N. Kruger and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No.7, pp. 775-779, Jul. 1997.
26. D. Xu, S. Yan, L. Zhang, M. Li, W. Ma, Z. Liu and H. Zhang, “Parallel Image Matrix Compression for Face Recognition,” accepted by The 7th International Multimedia Modeling Conference, 2005.
27. D. Xu, S. Yan, L. Zhang, Z. Liu and H. Zhang, “Coupled Subspace Analysis,” Microsoft Research Technical Report.
28. Y. Yacoob and L.S. Davis, “Recognizing Human Facial Expression From Long Image Sequences Using Optical Flow,” IEEE Trans. PAMI, Vol. 18, No. 6, pp. 636-642, Jun. 1996.
29. J. Yang, D. Zhang, A.F. Frangi and J.Y. Yang, “Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recogntion,” IEEE Trans. PAMI, Vol. 26, No. 1, pp. 131-137, Jan. 2004.
30. Y. Zhang and Q. Ji, “Facial Expression Understanding in Image Sequences using Dynamic and Active Visual Information Fusion,” Proceedings of ICCV03.
31. Z. Zhang, “Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiments with a Multi-Layer Perceptron,” International Journal of Pattern Recognition and Artificial Intelligence, pp. 893-911, 1998.
32. W. Zhao, R. Chellappa and P. Phillips, “Face Recognition: A Literature Survey,” Technical Report, 2002.