We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

An Extension of Principal Component Analysis

Hongchuan Yu and Jian J. Zhang
National Centre for Computer Animation, Bournemouth University
U.K.

1. Introduction

Principal component analysis (PCA), which is also known as Karhunen-Loeve (KL) transform, is a classical statistic technique that has been applied to many fields, such as knowledge representation, pattern recognition and image compression. The objective of PCA is to reduce the dimensionality of dataset and identify new meaningful underlying variables. The key idea is to project the objects to an orthogonal subspace for their compact representations. It usually involves a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables, which are called principal components. The first principal component accounts for as much of the variability in the dataset as possible, and each succeeding component accounts for as much of the remaining variability as possible. In pattern recognition, PCA technique was first applied to the representation of human face images by Sirovich and Kirby in [1,2]. This then led to the well-known Eigenfaces method for face recognition proposed by Turk and Penland in [3]. Since then, there has been an extensive literature that addresses both the theoretical aspect of the Eigenfaces method and its application aspect [4-6]. In image compression, PCA technique has also been widely applied to the remote hyperspectral imagery for classification and compression [7,8]. Nevertheless, it can be noted that in the classical 1D-PCA scheme the 2D data sample (e.g. image) must be initially converted to a 1D vector form. The resulting sample vector will lead to a high dimensional vector space. It is consequently difficult to evaluate the covariance matrix accurately when the sample vector is very long and the number of training samples is small. Furthermore, it can also be noted that the projection of a sample on each principal orthogonal vector is a scale. Obviously, this will cause the sample data to be over-compressed. In order to solve this kind of dimensionality problem, Yang et al. [9,10] proposed the 2D-PCA approach. The basic idea is to directly use a set of matrices to construct the corresponding covariance matrix instead of a set of vectors. Compared with the covariance matrix of 1D-PCA, one can note that the size of the covariance matrix using 2D-PCA is much smaller. This improves the computational efficiency. Furthermore, it can be noted that the projection of a sample on each principal orthogonal vector is a vector. Thus, the problem of over-compression is alleviated in the 2D-PCA scheme. In addition, Wang et al. [11] proposed that the 2D-PCA was equivalent to a special case of the block-based PCA, and emphasized that this kind of block-based methods had been used for face recognition in a number of systems.

For the multidimensional array cases, the higher order SVD (HO-SVD) has been applied to face recognition in [12,13]. They both employed a higher order tensor form associated with people, view, illumination, and expression dimensions and applied the HO-SVD to it for face recognition. We formulated them into the N-Dimensional PCA scheme in [14]. However, the presented ND-PCA scheme still adopted the classical single directional decomposition. Besides, due to the size of tensor, HO-SVD implementation usually leads to a huge matrix along some dimension of tensor, which is always beyond the capacity of an ordinary PC. In [12,13], they all employed small sized intensity images or feature vectors and a limited number of viewpoints, facial expressions and illumination changes in their "tensorface", so as to avoid this numerical challenge in HO-SVD computation.

Motivated by the above-mentioned works, in this chapter, we will reformulate our ND-PCA scheme presented in [14] by introducing the multidirectional decomposition technique for a near optimal solution of the low rank approximation, and overcome the above-mentioned numerical problems. However, we also noted the latest progress – Generalized PCA (GPCA), proposed in [15]. Unlike the classical PCA techniques (i.e. SVD-based PCA approaches), it utilizes the polynomial factorization techniques to subspace clustering instead of the usual Singular Value Decomposition approach. The deficiency is that the polynomial factorization usually yields an overabundance of monomials, which are used to span a high-dimensional subspace in GPAC scheme. Thus, the dimensionality problem is still a challenge in the implementation of GPCA. We will focus on the classical PCA techniques in this chapter.

The remainder of this chapter is organized as follows: In Section 2, the classical 1D-PCA and 2D-PCA are briefly revisited. The ND-PCA scheme is then formulated by using the multidirectional decomposition technique in Section 3, and the error estimation is also given. To evaluate the ND-PCA, it is performed on the FRGC 3D scan facial database [16] for multi-model face recognition in Section 4. Finally, some conclusions are given in Section 5.

2. 1D- AND 2D-PCA, AN OVERVIEW

1D-PCA

Let a sample $X \in R^n$. This sample is usually expressed in a vector form in the case of 1D-PCA. Traditionally, principal component analysis is performed on a square symmetric matrix of the cross product sums, such as the Covariance and Correlation matrices (i.e. cross products from a standardized dataset), i.e.

$$\begin{cases} Cov = E\left\{(X - \bar{X})(X - \bar{X})^T\right\} \\ Cor = (X - X_0)(Y - Y_0)^T \end{cases} \quad (1)$$

where, $\bar{X}$ is the mean of the training set, while X_0, Y_0 are standard forms. Indeed, the analysis of the Correlation and Covariance are different, since covariance is performed within the dataset, while correlation is used between different datasets. A correlation object has to be used if the variances of the individual samples differ much, or if the units of measurement of the individual samples differ. However, correlation can be considered as a special case of covariance. Thus, we will only pay attention to the covariance in the rest of this chapter.

After the construction of the covariance matrix, Eigen Value Analysis is applied to *Cov* of Eq.(1), i.e. $Cov = U\Sigma U^T$. Herein, the first k eigenvectors in the orthogonal matrix U corresponding to the first k largest eigenvalues span an orthogonal subspace, where the major energy of the sample is concentrated. A new sample of the same object is projected in this subspace for its compact form (or PCA representation) as follows,

$$\alpha = U_k^T (X - \bar{X}), \tag{2}$$

where, U_k is a matrix consisting of the first k eigenvectors of U, the projection α is a k-dimensional vector, which calls the k principal components of the sample X. The estimate of a novel representation of X can be described as,

$$\tilde{X} = U_k \alpha + \bar{X}. \tag{3}$$

It is clearly seen that the size of the covariance matrix of Eq.(1) is very large when the sample vectors are very long. Due to the large size of the covariance matrix and the relatively small number of training samples, it is difficult to estimate the covariance matrix of Eq.(1) accurately. Furthermore, a sample is projected on a principal vector as follows,

$$\alpha_i = u_i^T (X - \bar{X}), \quad \alpha_i \in \alpha, u_i \subset U_k, i = 1...k.$$

It can be noted that the projection α_i is a scale. Thus, this usually causes over-compression, i.e. we will have to use many principal components to approximate the original sample X for a desired quality. We call these above-mentioned numerical problems as "curse of dimensionality".

2D-PCA

In order to avoid the above mentioned problem, Yang et al. in [10] firstly presented a 2D-PCA scheme for 2D array cases in order to improve the performance of the PCA-style classifiers, that is, SVD is applied to the covariance matrix of, $G = \sum_i (X_i - \bar{X})^T (X_i - \bar{X})$, to get $G = V\Lambda V^T$, where $X_i \in R^{n \times m}$ denotes a sample, $\bar{X}$ denotes the mean of a set of samples, and V is the matrix of the eigenvectors and Λ is the matrix of the eigenvalues. The low-rank approximation of sample X is described as,

$$\begin{cases} \tilde{X} = YV_k^T + \bar{X} \\ Y = (X - \bar{X})V_k \end{cases} \tag{4}$$

where V_k contains the first k principal eigenvectors of G. It has been noted that 2D-PCA only considers between column (or row) correlations [11].

In order to improve the accuracy of the low rank approximation, Ding et al. in [17] presented a 2D-SVD scheme for 2D cases. The key idea is to employ the 2-directional decomposition to the 2D-SVD scheme, that is, two covariance matrices of,

$$\begin{cases} F = \sum_i (X_i - \bar{X})(X_i - \bar{X})^T = U\Lambda_F U^T \\ G = \sum_i (X_i - \bar{X})^T (X_i - \bar{X}) = V\Lambda_G V^T \end{cases},$$

are considered together. Let U_k contain the first k principal eigenvectors of F and V_s contain the first s principal eigenvectors of G. The low-rank approximation of X can be expressed as,

$$\begin{cases} \tilde{X} = U_k M V_s^T + \bar{X} \\ M = U_k^T (X - \bar{X}) V_s \end{cases}.$$
(5)

Compared to the scheme Eq.(5), the scheme Eq.(4) of 2D-PCA only employs the classical single directional decomposition. It is proved that the scheme Eq.(5) of 2D-SVD can obtain a near-optimal solution compared to 2D-PCA in [17]. While, in the dyadic SVD algorithm [18], the sample set is viewed as a 3 order tensor and the HO-SVD technique is applied to each dimension of this tensor except the dimension of sample number, so as to generate the principal eigenvector matrices U_k and V_s as in the 2D-SVD.

3. N-DIMENSIONAL PCA

For clarity, we first introduce Higher Order SVD [19] briefly, and then formulate the N-dimensional PCA scheme.

3.1 Higher Order SVD

A higher order tensor is usually defined as $A \in R^{I_1 \times ... \times I_N}$, where N is the order of A, and $1 \leq i_n \leq I_n$, $1 \leq n \leq N$. In accordance with the terminology of tensors, the column vectors of a 2-order tensor (matrix) are referred to as 1-mode vectors and row vectors as 2-mode vectors. The n-mode vectors of an N-order tensor A are defined as the I_n-dimensional vectors obtained from A by varying the index i_n and keeping the other indices fixed. In addition, a tensor can be expressed in a matrix form, which is called matrix unfolding (refer to [19] for details).

Furthermore, the n-mode product, $\times_n$, of a tensor $A \in R^{I_1 \times ... \times I_n ... \times I_N}$ by a matrix $U \in R^{J_n \times I_n}$ along the n-th dimension is defined as,

$$(A \times_n U)_{i_1,...,i_{n-1},j_n,i_{n+1},...,i_N} = \sum_{i_n} a_{i_1,...,i_n,...,i_N} u_{j_n,i_n} .$$

In practice, n-mode multiplication is implemented first by matrix unfolding the tensor A along the given n-mode to generate its n-mode matrix form $A_{(n)}$, and then performing the matrix multiplication as follows,

$$B_{(n)} = U A_{(n)} .$$

After that, the resulting matrix $B_{(n)}$ is folded back to the tensor form, i.e. $A \times_n U = \text{fold}_n \left(U \text{unfold}_n(A) \right)$. In terms of n-mode multiplication, Higher Order SVD of a tensor A can be expressed as,

$$A = S \times_1 U^{(1)} \times_2 ... \times_N U^{(N)} ,$$
(6)

where, $U^{(n)}$ is a unitary matrix of size $I_n \times I_n$, which contains n-mode singular vectors. Instead of being pseudo-diagonal (nonzero elements only occur when the indices $i_1 = ... = i_N$), the tensor S (called the core tensor) is all-orthogonal, that is, two subtensors $S_{i_n=a}$ and $S_{i_n=b}$ are orthogonal for all possible values of n, a and b subject to $a \neq b$. In addition, the Frobenius-norms $s_i^{(n)} = \left\| S_{i_n=i} \right\|_F$ are n-mode singular values of A and are in

decreasing order, $s_1^{(n)} \geq \ldots \geq s_{I_n}^{(n)} \geq 0$, which correspond to n-mode singular vectors $u_i^{(n)} \subset U^{(n)}, i = 1, \ldots, I_n$ respectively. The numerical procedure of HO-SVD can be simply described as,

$$\text{unfold}_n(A) = U^{(n)} \Sigma^{(n)} V^{(n)T}, n = 1, \ldots, N ,$$

where, $\Sigma^{(n)} = diag\left(s_1^{(n)}, \ldots, s_{I_n}^{(n)}\right)$ and $V^{(n)}$ is another orthogonal matrix of SVD.

3.2 Formulating N-dimensional PCA

For the multidimensional array case, we first employ a difference tensor instead of the covariance tensor as follows,

$$D = \left((X_1 - \bar{X}), \ldots, (X_M - \bar{X})\right) , \tag{7}$$

where $X_i \in R^{I_1 \times \ldots I_i \times \ldots \times I_N}$ and $D \in R^{I_1 \times \ldots MI_i \times \ldots \times I_N}$, i.e. N-order tensors $(X_n - \bar{X}), n = 1, \ldots, M$ are stacked along the ith dimension in the tensor D. Then, applying HO-SVD of Eq.(6) to D will generate n-mode singular vectors contained in $U^{(n)}, n = 1, \ldots, N$. According to the n-mode singular values, one can determine the desired principal orthogonal vectors for each mode of the tensor D respectively. Introducing the multidirectional decomposition to Eq.(7) will yield the desired N-dimensional PCA scheme as follows,

$$\begin{cases} \tilde{X} = Y \times_1 U_{k_1}^{(1)} \times_2 \ldots \times_N U_{k_N}^{(N)} + \bar{X} \\ Y = (X - \bar{X}) \times_1 U_{k_1}^{(1)T} \times_2 \ldots \times_N U_{k_N}^{(N)T} \end{cases} \tag{8}$$

where $U_{k_i}^{(i)}$ denotes the matrix of i-mode k_i principal vectors, $i = 1, \ldots N$. The main challenge is that unfolding the tensor D in HO-SVD usually generates an overly large matrix.

First, we consider the case of unfolding D along the ith dimension, which generates a matrix of size $MI_i \times (I_{i+1} \cdot \ldots \cdot I_N \cdot I_1 \cdot \ldots \cdot I_{i-1})$. We prefer a unitary matrix $U^{(i)}$ of size $I_i \times I_i$ to one of the sizes $MI_i \times MI_i$. This can be achieved by reshaping the unfolded matrix as follows.

Let A_j be a $I_i \times (I_{i+1} \cdot \ldots \cdot I_N \cdot I_1 \cdot \ldots \cdot I_{i-1})$ matrix and $j = 1, \ldots M$. The unfolded matrix is expressed as $A = \begin{pmatrix} A_1 \\ \ldots \\ A_M \end{pmatrix}$. Reshaping A into a $I_i \times M(I_{i+1} \cdot \ldots \cdot I_N \cdot I_1 \cdot \ldots \cdot I_{i-1})$ matrix $\tilde{A} = (A_1, \ldots, A_M)$, we can obtain an unitary matrix $U^{(i)}$ of size $I_i \times I_i$ by SVD.

Then, consider the generic case. Since the sizes of dimensions $I_1, \ldots, I_N$ may be very large, this still leads to an overly large matrix along some dimension of sample X. Without loss of generality, we assume that the sizes of dimensions of sample X are independent of each other.

Now, this numerical problem can be rephrased as follows, for a large sized matrix, how to carry out SVD decomposition. It is straightforward to apply matrix partitioning approach to the large matrix. As a start point, we first provide the following lemma.

Lemma:

For any matrix $M \in R^{n \times m}$, if each column M_i of M, $M = (M_1, \ldots, M_m)$, maintain its own singular value σ_i, i.e. $M_i M_i^T = U_i diag(\sigma_i^2, 0, \ldots, 0) U_i^T$, while the singular values of M are $s_1, \ldots, s_{\min(m,n)}$, i.e. $M = V diag(s_1, \ldots, s_{\min(m,n)}) U^T$, then $\sum\limits_{i=1}^{\min(m,n)} \sigma_i^2 = \sum\limits_{i=1}^{\min(m,n)} s_i^2$.

Proof:

Let $n > m$. Because,

$$MM^T = \sum_{i=1}^{m} M_i M_i^T = \sum_{i=1}^{m} u_i \sigma_i^2 u_i^T = (u_1, \ldots, u_m) diag(\sigma_1^2, \ldots, \sigma_m^2)(u_1, \ldots, u_m)^T,$$

where u_i is the first column of each U_i, while the SVD of MM^T,

$$MM^T = V diag(s_1^2, \ldots, s_m^2, 0, \ldots, 0) V^T = \sum_{i=1}^{m} v_i s_i^2 v_i^T,$$

where v_i is the ith column of V. We have,

$$tr(MM^T) = \sum_{i}^{m} \sigma_i^2 = \sum_{i}^{m} s_i^2, \qquad \textbf{End of proof.}$$

This lemma implies that each column of M corresponds to its own singular value. Moreover, let M_i be a submatrix instead of column vector, $M_i \in R^{n \times r}$. We have,

$$M_i M_i^T = U_i diag(s_{1i}^2, \ldots s_{ri}^2, \ldots, 0) U_i^T.$$

It can be noted that there are more than one non-zero singular values $s_{1i} \geq \ldots \geq s_{ri} \geq 0$. If we let $rank(M_i M_i^T) = 1$, the approximation of $M_i M_i^T$ can be written as $M_i M_i^T \approx U_i diag(s_{1i}^2, 0, \ldots, 0) U_i^T$. In terms of the lemma, we can also approximate it as $M_i M_i^T \approx M_{1i} M_{1i}^T = u_{1i} \sigma_{1i}^2 u_{1i}^T$, where M_{1i} is a column of M_i corresponding to the biggest singular value σ_{1i} of column vector. On this basis, M_{1i} is regarded as the principal column vector of the submatrix M_i.

We can rearrange the matrix $M \in R^{n \times m}$ by sorting these singular values $\{\sigma_i\}$ and partition it into t block submatrices, $M = (M_1, \ldots, M_t)$, where $M_i \in R^{n \times m_i}, i = 1, \ldots, t, m = \sum\limits_{i}^{t} m_i$. Indeed, the principal eigenvectors are derived only from some particular submatrices rather than the others as the following analysis. (For computational convenience, we assume $m \geq n$ below.)

In the context of PCA, the matrix of the first k principal eigenvectors is preferred to a whole orthogonal matrix. Thus, we partition M into 2 block submatrices $\tilde{M} = (M_1, M_2)$ in terms of the sorted singular values $\{\sigma_i\}$, so that M_1 contains the columns corresponding to the first k biggest singular values while M_2 contains the others. Note that $\tilde{M}$ is different from the original M because of a column permutation (denoted as *Permute*). Applying SVD to each

M_i respectively yields,

$$\tilde{M} = (U_1, U_2) \begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix} \begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}. \tag{9}$$

Thus, matrix $\tilde{M}$ can be approximated as follows,

$$\tilde{M} \approx \tilde{M}' = (U_1, U_2) \begin{pmatrix} \Lambda_1 & \\ & 0 \end{pmatrix} \begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}. \tag{10}$$

In order to obtain the approximation of M, the inverse permutation of *Permute* needs to be carried out on the row-wise orthogonal matrix of $\begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}$ given in Eq.(10). The resulting matrix is the approximation of the original matrix M. The desired principal eigenvectors are therefore included in the matrix of U_1.

Now, we can re-write our ND-PCA scheme as,

$$\begin{cases} \tilde{X} = Y \times_1 U_{k_1}^{(1)} ... \times_i U_{k_i}^{(i)} ... \times_N U_{k_N}^{(N)} + \bar{X} \\ Y = (X - \bar{X}) \times_1 U_{k_1}^{(1)T} ... \times_N U_{k_N}^{(N)T} \\ U_{k_i}^{(i)} \text{ is from Eq.(10)} \end{cases} \tag{11}$$

For comparison, the similarity metric can adopt the Frobenius-norms between the reconstructions of two samples X and X' as follows,

$$\varepsilon = \left\| \tilde{X} - \tilde{X}' \right\|_F = \left\| Y - Y' \right\|_F. \tag{12}$$

Furthermore, we can provide the following proposition,

Proposition:
$\tilde{X}$ of Eq.(11) is a near optimal approximation to sample X in a least-square sense.
Proof.
According to the property 10 of HO-SVD in [19], we assume that the n-mode rank of $(X - \bar{X})$ be equal to $R_n (1 \leq n \leq N)$ and $(\tilde{X} - \bar{X})$ be defined by discarding the smallest n-mode singular values $\sigma_{I_n'+1}^{(n)}, ..., \sigma_{R_n}^{(n)}$ for given I_n'. Then, the approximation $\tilde{X}$ is a near optimal approximation of sample X. The error is bounded by Frobenius-norm as follows,

$$\left\| X - \tilde{X} \right\|_F^2 \leq \sum_{i_1=I_1'+1}^{R_1} \sigma_{i_1}^{(1)2} + ... + \sum_{i_N=I_N'+1}^{R_N} \sigma_{i_N}^{(N)2}. \tag{13}$$

This means that the tensor $(\tilde{X} - \bar{X})$ is in general not the best possible approximation under the given n-mode rank constraints. But under the error upper-bound of Eq.(13), $\tilde{X}$ is a near optimal approximation of sample X.

Unfolding $(X - \bar{X})$ along ith dimension yields a large matrix which can be partitioned into two submatrices as shown in Eq.(9), i.e.

$$\tilde{M} = (M_1, M_2) = (U_1, U_2) \begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix} \begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}.$$

Let $\tilde{M}' = (U_1, U_2) \begin{pmatrix} \Lambda_1 & \\ & 0 \end{pmatrix} \begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}$ as shown in Eq.(10). Consider the difference of $\tilde{M}$ and $\tilde{M}' \in R^{n \times m}$ as follows,

$$\tilde{M} - \tilde{M}' = (U_1, U_2) \begin{pmatrix} 0 & \\ & \Lambda_2 \end{pmatrix} \begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix},$$

where $U_i \in R^{n \times n}, V_i \in R^{m_i \times m_i}, \Lambda_i \in R^{n \times m_i}, i = 1, 2$. It can be noted that the 2-norm of $\begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}$ is 1, and that of $\begin{pmatrix} 0 & \\ & \Lambda_2 \end{pmatrix}$ is $\max\{\sigma : \sigma \in \Lambda_2\}$. Since

$$(U_1, U_2) = U_1 (I_{n \times n}, I_{n \times n}) \begin{pmatrix} I_{n \times n} & \\ & U_1^T U_2 \end{pmatrix},$$

we can note that the 2-norm of both the orthogonal matrix U_1 and $\begin{pmatrix} I_{n \times n} & \\ & U_1^T U_2 \end{pmatrix}$ are 1, and that of $(I_{n \times n}, I_{n \times n})$ is $\sqrt{2}$ because of identity matrix $I_{n \times n}$. Therefore, we have,

$$\left\| \tilde{M} - \tilde{M}' \right\|_2^2 \leq 2 \max^2 \{\sigma : \sigma \in \Lambda_2\}, \tag{14}$$

in a 2-norm sense.

Substituting Eq.(14) into Eq.(13) yields the error upper-bound of $\tilde{X}$ as follows,

$$\left\| X - \tilde{X} \right\|_F^2 \leq 2 \left(\max^2 \left\{ \sigma^{(1)} : \sigma^{(1)} \in \Lambda_2^{(1)} \right\} + \dots + \max^2 \left\{ \sigma^{(N)} : \sigma^{(N)} \in \Lambda_2^{(N)} \right\} \right). \tag{15}$$

This implies that the approximation $\tilde{X}$ of Eq.(11) is a near optimal approximation of sample X under this error upper bound.   **End of proof**.

Remark: So far, we formulated the ND-PCA scheme, which can deal with overly large matrix. The basic idea is to partition the large matrix and discard non-principal submatrices. In general, the dimensionality of eigen-subspace is determined by the ratio of sum of singular values in the subspace to the one of the whole space for solving the dimensionality reduction problems [20]. But, for an overly large matrix, we cannot get all the singular values of the whole matrix here, because of discarding the non-principal submatrices. An alternative is to iteratively determine the dimensionality of eigen-subspace by using reconstruction error threshold.

4. EXPERIMENTS AND ANALYSIS

The proposed ND-PCA approach was performed on a 3D range database of human faces used for the Face Recognition Grand Challenge [16]. In order to establish an analogy with a 3D volume dataset or multidimensional solid array, each 3D range dataset was first mapped to a 3D array and the intensities of the corresponding pixels in the still face image were regarded as the voxel values of the 3D array. For the sake of memory size, the reconstructed volume dataset was then re-sampled to the size of 180×180×90. Figure 1 shows an example of the still face image, corresponding range data and the reconstructed 3D model.

Experiment 1. This experiment is to test the rank of the singular values. In our gallery, eight samples of each person are available for training. Their mean-offset tensors are aligned together along the second index (x axis) to construct a difference tensor $D \in R^{180 \times 1440 \times 90}$. We applied HO-SVD of Eq.(6) to D to get the 1-mode and 3-mode singular values of D, which are depicted in Fig.2. One can note that the numbers of 1-mode and 3-mode singular values are different, and they are equal to the dimensionalities of indices 1 and 3 of D respectively (i.e. 180 for 1-mode and 90 for 3-mode). This is a particular property of higher order tensors, namely the N-order tensor A can have N different n-mode ranks but all of them are less than the rank of A, $rank_n(A) \leq rank(A)$. Furthermore, the corresponding n-mode singular vectors constitutes orthonormal basis which can span independent n-mode orthogonal subspaces respectively. Therefore, we can project a sample to an arbitrary n-mode orthogonal subspace accordingly. In addition, one can also note that the magnitude of the singular values declines very quickly. This indicates that the energy of a sample is only concentrated on a small number of singular vectors as expected.
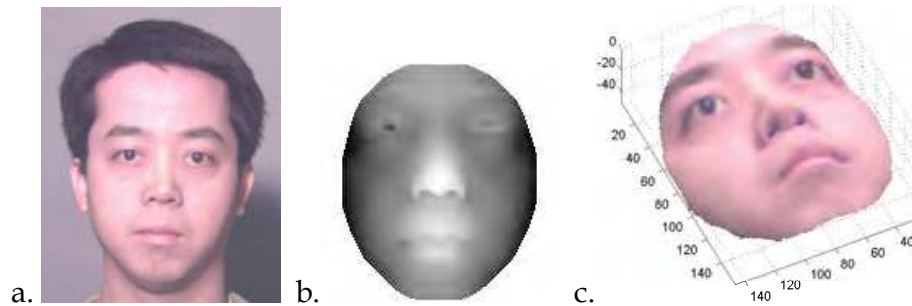

Fig. 1. The original 2D still face image (a), range data (b) and reconstructed 3D model (c) of a face sample.
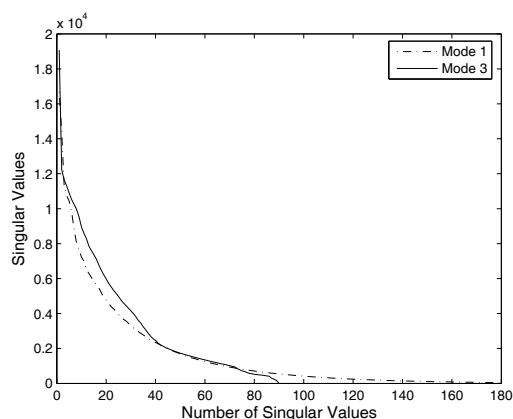

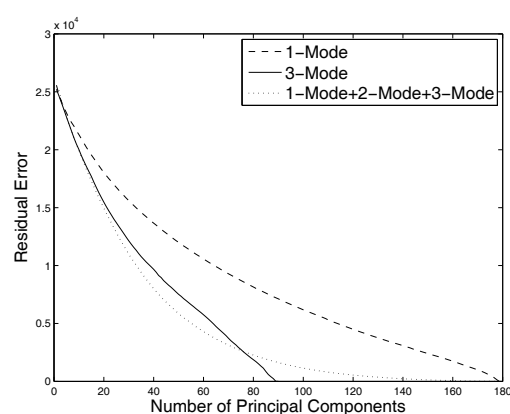Fig. 2. The singular values in decreasing order.

Fig. 3. Comparison of the reconstruction through 1-mode, 3-mode and 1-mode+2-mode+3-mode principal subspace respectively. ND-PCA with multidirectional decomposition converges quicker than ND-PCA with single directional decomposition.

Experiment 2. This experiment is to test the quality of the reconstructed sample. Within our 3D volume dataset, we have 1-mode, 2-mode and 3-mode singular vectors, which could span three independent orthogonal subspaces respectively. The sample could be approximated by using the projections from one orthogonal subspace, two ones or three ones. Our objective is to test which combination leads to the best reconstruction quality. We designed a series of tests for this purpose. The reconstructed sample using the scheme of Eq.(11) was performed on 1-mode, 3-mode and 1-mode+2-mode+3-mode principal subspaces respectively with a varying number of principal components k. (Note that 1-mode or 3-mode based ND-PCA adopted the single directional decomposition, while 1-mode+2-mode+3-mode based ND-PCA adopted the multidirectional decomposition.) The residual errors of reconstruction are plotted in Fig.3. Since the sizes of dimensions of $U^{(1)}$ and $U^{(3)}$ are different, the ranges of the corresponding number of principal components k are also different. However, k must be less than the size of dimension of the corresponding orthogonal matrix $U^{(1)}$ or $U^{(3)}$. As a result of the differing dimensionalities, the residual error of reconstruction in 3-mode principal subspace converges to zero faster than in 1-mode or 1-mode+2-mode+3-mode principal subspaces. Indeed, if the curve of 3-mode (solid curve) is quantified to the same length of row coordinate as the curve of 1-mode (dashed line) in Fig.3, there is no substantial difference compared to the 1-mode test. This indicates that the reconstructed results are not affected by the difference between the different n-mode principal subspaces. Furthermore, in the test of 1-mode+2-mode+3-mode principal subspaces, the number of principal components k was set to 180 for both $U^{(1)}$ and $U^{(2)}$ while it was set to 90 for $U^{(3)}$. Comparing the curve of 1-mode+2-mode+3-mode (dot line) with that of 1-mode (dashed line) and 3-mode (solid line), one can note that the approximation of 1-mode+2-mode+3-mode principal subspace converges to the final optimal solution more rapidly.

———-·

Remark: In [9,10], the over-compressed problem was addressed repeatedly. [10] gave a comparison of the reconstruction results between the 1D-PCA case and the 2D-PCA case, which is reproduced in Fig.4 for the sake of completeness. It can be noted that the small number of principal components of the 2D-PCA can perform well compared with the large number of principal components of the 1D-PCA. Moreover, consider the cases of single directional decomposition, i.e. 2D-PCA and 1-mode based ND-PCA scheme, and multidirectional decomposition, i.e. 2D-SVD and 1-mode+2-mode+3-mode based ND-PCA. We respectively compared the reconstructed results of the single directional decomposition and the multidirectional decomposition with a varying number of principal components k (i.e. the reconstruction of the volume dataset by using the ND-PCA of Eq.(11) while the reconstruction of the corresponding 2D image respectively by using 2D-PCA of Eq.(4) and 2D-SVD of Eq.(5)). The training set is the same as in the first experiment. The residual errors of reconstruction are normalized to the range of [0,1], and are plotted in Fig.5. One can note that the multidirectional decomposition performs better than the single directional decomposition in the case of a small number of principal components (i.e. comparing Fig.5a with Fig.5b). But then comparing the 2D-PCA with ND-PCA scheme shown in Fig.5a (or 2D-SVD with ND-PCA scheme shown in Fig.5b), one can also note that 2D-PCA (or 2D-SVD) performs a little better than ND-PCA scheme when only a small number of principal components are used. In our opinion, there is no visible difference in the reconstruction quality between 2D-PCA (or 2D-SVD) and ND-PCA scheme with a small number of

singular values. This is because the reconstructed 3D volume dataset is a sparse 3D array (i.e. all voxel values are set to zero except the voxels on the face surface), it is therefore more sensitive to computational errors compared to a 2D still image. If the 3D volume datasets were solid, e.g. CT or MRI volume datasets, this difference between the two curves of Fig.5a or Fig.5b would not noticeably appear.
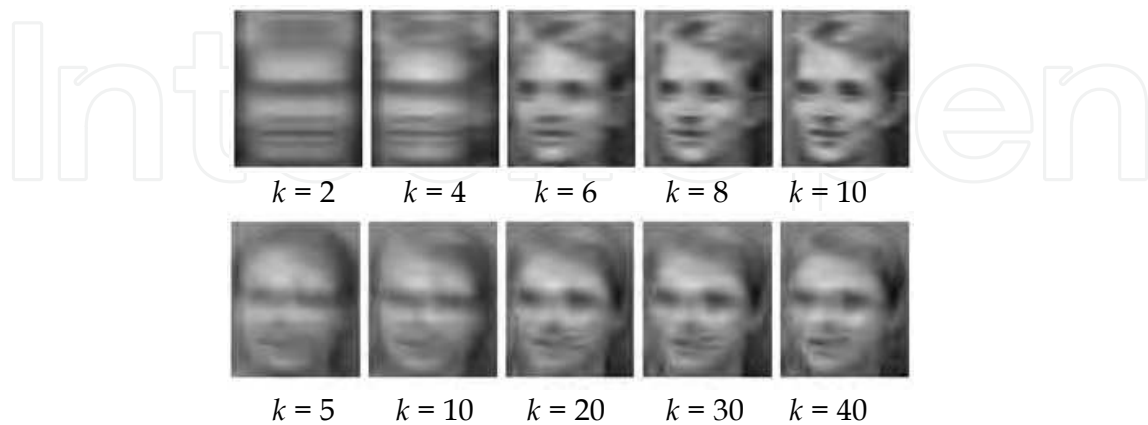

$k = 2$ $k = 4$ $k = 6$ $k = 8$ $k = 10$

$k = 5$ $k = 10$ $k = 20$ $k = 30$ $k = 40$

Fig. 4. Comparison of the reconstructed images using 2D-PCA (upper) and 1D-PCA (lower) from [10].

a. single direction decomposition.   b. multiple direction decomposition
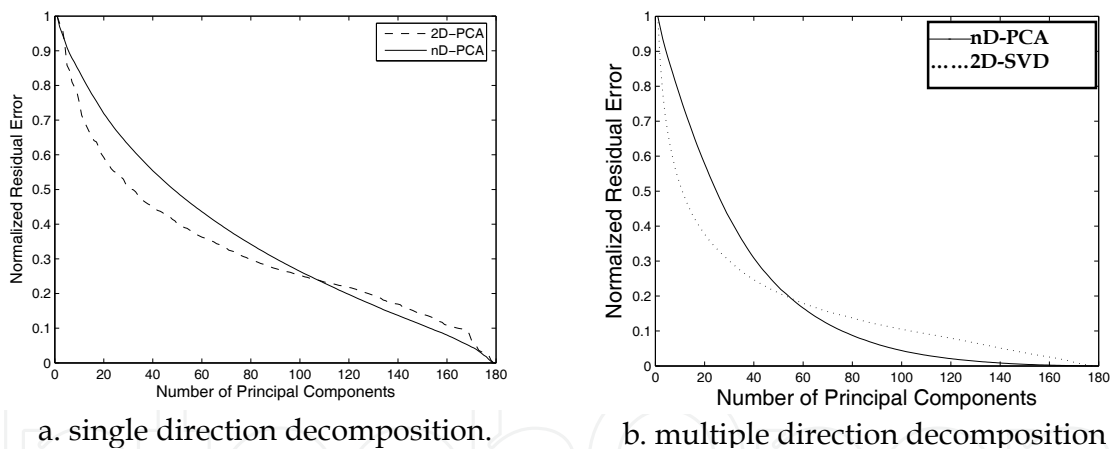
Fig. 5. Comparison of the reconstruction by using single directional decomposition (a), i.e. 2D-PCA and 1-mode based ND-PCA scheme, and multidirectional composition (b), i.e. 2D-SVD and ND-PCA, in terms of the normalized residual errors.

Experiment 3. In this experiment, we compared the 1-mode based ND-PCA scheme with the 1-mode+2-mode+3-mode based ND-PCA scheme on the performance of the face verification using the Receiver Operating Characteristic (ROC) curves [21]. Our objective is to reveal the recognition performance between these two ND-PCA schemes respectively by using the single directional decomposition and the multidirectional decomposition. The whole test set includes 270 samples (i.e. range datasets and corresponding still images), in which there are 6 to 8 samples for one person. All these samples are from the FRGC database and are re-sampled. Two ND-PCA schemes were carried out directly on the reconstructed volume

datasets. Their corresponding ROC curves are shown respectively in Fig.6. It can be noted that the overlapping area of the genuine and impostor distributions (i.e. false probability) in Fig.(6a) is smaller than that in Fig.(6b). Furthermore, their corresponding ROC curves relating to the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) are depicted by changing the threshold as shown in Fig.(6c). At some threshold, the false probability of recognition corresponds to some rectangular area under the ROC curve. The smaller the area under the ROC curve, the higher is the rising of the accuracy of the recognition. For quantitative comparison, we could employ the Equal Error Rate (EER), which is defined as the error rate at the point on ROC curve where the FAR is equal to the FRR. The EER is often used for comparisons because it is simpler to obtain and compare a single value characterizing the system performance. In Fig.(6c), the EER of Fig.(6a) is 0.152 while the EER of Fig.(6b) is 0.224. Obviously, the ND-PCA scheme with multidirectional decomposition can improve the accuracy of face recognition. Of course, since the EERs only give comparable information between the different systems that are useful for a single application requirement, the full ROC curve is still necessary for other potentially different application requirements.
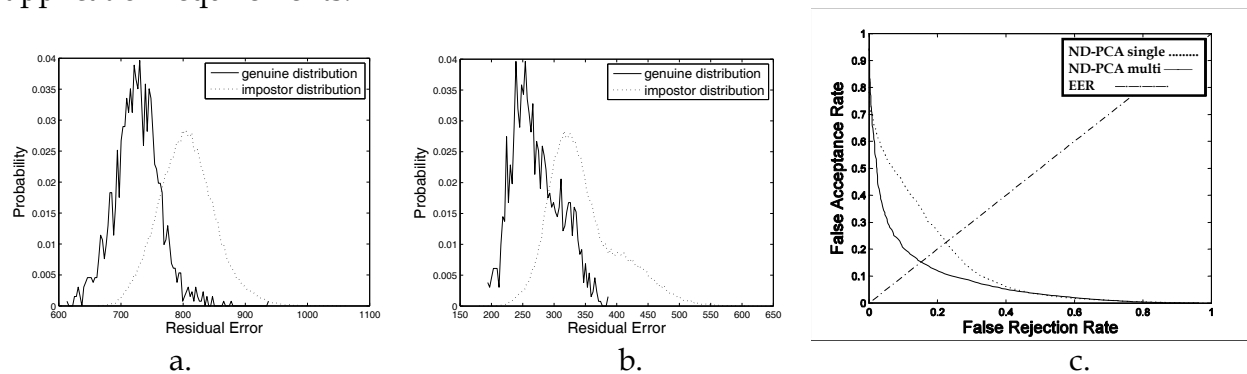
Fig. 6. Comparison of the recognition performance. a) are the genuine and impostor distribution curves of ND-PCA with multidirectional decomposition; b) are the genuine and impostor distribution curves of ND-PCA with single directional decomposition; c) are the ROC curves relating to the False acceptance rate and False rejection rate.

5. CONCLUSION

In this chapter, we formulated the ND-PCA approach, that is, to extend the PCA technique to the multidimensional array cases through the use of tensors and Higher Order Singular Value Decomposition technique. The novelties of this chapter include, 1) introducing the multidirectional decomposition into ND-PCA scheme and overcoming the numerical difficulty of overly large matrix SVD decomposition; 2) providing the proof of the ND-PCA scheme as a near optimal linear classification approach. We performed the ND-PCA scheme on 3D volume datasets to test the singular value distribution, and the error estimation. The results indicated that the proposed ND-PCA scheme performed as well as we desired. Moreover, we also performed the ND-PCA scheme on the face verification for the comparison of single directional decomposition and multidirectional decomposition. The experimental results indicated that the ND-PCA scheme with multidirectional decomposition could effectively improve the accuracy of face recognition.

6. References

1. Sirovich, L. and Kirby, M. (1987). Low-Dimensional Procedure for Characterization of Human Faces. *J. Optical Soc. Am.*, Vol. 4, pp. 519-524.

2. Kirby, M. and Sirovich, L. (1990). Application of the KL Procedure for the Characterization of Human Faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, pp. 103-108.

3. Turk, M. and Pentland, A. (1991). Eigenfaces for Recognition. *J. Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86.

4. Sung, K. and Poggio, T. (1998). Example-Based Learning for View-Based Human Face Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 39-51.

5. Moghaddam, B. and Pentland, A. (1997). Probabilistic Visual Learning for Object Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 696-710.

6. Zhao, L. and Yang, Y. (1999). Theoretical Analysis of Illumination in PCA-Based Vision Systems. *Pattern Recognition*, Vol. 32, No. 4, pp. 547-564.

7. Harsanyi, J.C. and Chang, C. (1994). Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Trans. Geoscience Remote Sensing*, Vol. 32, No. 4, pp. 779-785.

8. Sunghyun, L.; Sohn, K.H. and Lee, C. (2001). Principal component analysis for compression of hyperspectral images. *Proc. of IEEE Int. Geoscience and Remote Sensing Symposium*, Vol. 1, pp. 97-99.

9. Yang, J. and Yang, J.Y. (2002). From Image Vector to Matrix: A Straightforward Image Projection Technique—IMPCA vs. PCA. *Pattern Recognition*, Vol. 35, No. 9, pp. 1997-1999.

10. Yang, J.; Zhang, D.; Frangi, A.F. and Yang, J.Y. (2004). Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 1, pp. 131-137.

11. Wang, L.; Wang, X. and Zhang, X. et al. (2005). The equivalence of the two-dimensional PCA to lineal-based PCA. *Pattern Recognition Letters*, Vol. 26, pp. 57-60.

12. Vasilescu, M. and Terzopoulos, D. (2003). Multilinear subspace analysis of image ensembles. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2003)*, Vol. 2, June 2003.

13. Wang, H. and Ahuja, N. (2003). Facial Expression Decomposition. *Proc. of IEEE 9th Int'l Conf. on Computer Vision (ICCV'03)*, Vol. 2, Oct. 2003.

14. Yu, H. and Bennamoun, M. (2006). 1D-PCA 2D-PCA to nD-PCA. *Proc. of IEEE 18th Int'l Conf. on Pattern Recognition*, HongKong, pp. 181-184, Aug. 2006.

15. Vidal, R.; Ma, Y. and Sastry, S. (2005). Generalized Principal Component Analysis (GPCA). *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 12.

16. Phillips, P.J.; Flynn, P.J. and Scruggs, T. et al. (2005). Overview of the Face Recognition Grand Challenge. *Proc. of IEEE Conf. on CVPR2005*, Vol. 1.

17. Ding, C. and Ye, J. (2005). Two-dimensional Singular Value Decomposition (2DSVD) for 2D Maps and Images. *Proc. of SIAM Int'l Conf. Data Mining (SDM'05)*, pp:32-43, April 2005.

18. Inoue, K. and Urahama, K. (2006). Equivalence of Non-Iterative Algorithms for Simultaneous Low Rank Approximations of Matrices. *Proc. of IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR'06)*, Vol.1, pp. 154-159.

19. Lathauwer, L.D.; Moor, B.D. and Vandewalle, J. (2000). A Multilinear Singular Value Decomposition. *SIAM J. on Matrix Analysis and Applications*, Vol. 21, No. 4, pp. 1253-1278.

20. Moghaddam, B. and Pentland, A. (1997). Probabilistic Visual Learning for Object Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 696-710.

21. Jain, A.K.; Ross, A. and Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Trans. on Circuits and Systems For Video Technology*, Vol. 14, No. 1, pp. 4-20.

Face Recognition

Edited by Milos Oravec

This book aims to bring together selected recent advances, applications and original results in the area of biometric face recognition. They can be useful for researchers, engineers, graduate and postgraduate students, experts in this area and hopefully also for people interested generally in computer science, security, machine learning and artificial intelligence. Various methods, approaches and algorithms for recognition of human faces are used by authors of the chapters of this book, e.g. PCA, LDA, artificial neural networks, wavelets, curvelets, kernel methods, Gabor filters, active appearance models, 2D and 3D representations, optical correlation, hidden Markov models and others. Also a broad range of problems is covered: feature extraction and dimensionality reduction (chapters 1-4), 2D face recognition from the point of view of full system proposal (chapters 5-10), illumination and pose problems (chapters 11-13), eye movement (chapter 14), 3D face recognition (chapters 15-19) and hardware issues (chapters 19-20).

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hongchuan Yu and Jian J. Zhang (2010). An Extension of Principal Component Analysis, Face Recognition, Milos Oravec (Ed.), ISBN: 978-953-307-060-5, InTech, Available from: http://www.intechopen.com/books/face-recognition/an-extension-of-principal-component-analysis

INTECH
open science | open minds