N-Dimensional Principal Component Analysis

Hongchuan YU and Jian J. ZHANG

NCCA, Bournemouth University

Abstract In this paper, we first briefly introduce the multidimensional Principal Component Analysis (PCA) techniques, and then amend our previous N-dimensional PCA (ND-PCA) scheme by introducing multidirectional decomposition into ND-PCA implementation. For the case of high dimensionality, PCA technique is usually extended to an arbitrary n-dimensional space by the Higher-Order Singular Value Decomposition (HO-SVD) technique. Due to the size of tensor, HO-SVD implementation usually leads to a huge matrix along some direction of tensor, which is always beyond the capacity of an ordinary PC. The novelty of this paper is to amend our previous ND-PCA scheme to deal with this challenge and further prove that the revised ND-PCA scheme can provide a near optimal linear solution under the given error bound. To evaluate the numerical property of the revised ND-PCA scheme, experiments are performed on a set of 3D volume datasets.

1 Introduction

Principal component analysis (PCA) is a classical statistic technique that has been applied to many fields. Nevertheless, it can be noted that in the classical PCA 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 scalar. This causes the sample data to be over-compressed. In order to overcome this kind of dimensionality problems, [5] and [6] separately proposed their individual PCA schemes for 2D case. For the case of high dimensionality, the higher order SVD (HO-SVD) has been applied to face recognition [2,8]. 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 [1]. 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 [2,8], 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 paper we will reformulate our ND-PCA scheme by introducing the multidirectional decomposition for a near optimal solution of the low rank approximation and overcome the above-mentioned numerical problems.

2 Overview of Multidimensional PCA Techniques

[5] firstly presented a 2D-PCA scheme by using the single dimensional decomposition technique for 2D case. It has been noted that 2D-PCA only considers between column (or row) correlations [4]. In order to improve the accuracy of the low rank approximation,

[6] presented a 2D-SVD scheme, that is, SVD is applied respectively to two covariance matrices as follows,

$$\begin{cases} F = \sum_i (X_i - \bar{X})(X_i - \bar{X})^T = U \Lambda_F U^T \\ G = \sum_i (X_i - \bar{X})^T (X_i - \bar{X}) = V \Lambda_G V^T \end{cases}, \quad (1)$$

where $X_i \in R^2$ denotes a sample, $\bar{X}$ denotes the mean of a set of samples, and Λ_F, Λ_G denote the diagonal matrices of the eigenvalues respectively. Let U_k contains the first *k* principal eigenvectors of *F* and V_s contains the first *s* principal eigenvectors of *G*. The low-rank approximation of *X* can be expressed as, $\begin{cases} \hat{X} = U_k M V_s^T + \bar{X} \\ M = U_k^T (X - \bar{X}) V_s \end{cases}$. It is clear that the 2D-SVD employs the 2-directional decomposition, i.e. both U_k and V_s appear in $\hat{X}$, while the 2D-PCA only employs the classical single directional decomposition. It is proven that the 2D-SVD can obtain a near-optimal solution compared to the 2D-PCA in [6].

For the case of high dimensionality, we presented the ND-PCA scheme in [1], in which a difference tensor was used instead of the covariance tensor as follows,

$$D = \left((X_1 - \bar{X}), ..., (X_M - \bar{X}) \right), \quad (2)$$

where $X_i \in R^{I_1 \times ... I_i \times ... \times I_N}$ and $D \in R^{I_1 \times ... MI_i \times ... \times I_N}$, i.e. *N*-order tensors $(X_m - \bar{X}), m = 1, ..., M$ are stacked along the *i*th dimension in the tensor *D*. Furthermore, applying HO-SVD to *D* generate n-mode singular vectors contained in $U^{(n)}, n = 1, ..., N$. (For HO-SVD computation, refer to [7] please.) Accordingly, our ND-PCA scheme in [1] is expressed as,

$$\begin{cases} \hat{X} = Y_n \times_n U_k^{(n)} + \bar{X} \\ Y_n = (X - \bar{X}) \times_n U_k^{(n)T} \end{cases}, \quad (3)$$

where $U_k^{(n)}$ denotes the matrix of n-mode *k* principal vectors. It can be noted that the proposed ND-PCA scheme still adopted the classical single directional decomposition, i.e. only $U_k^{(n)}$ is used in $\hat{X}$. However, unfolding a tensor along some dimensions in the HO-SVD implementation usually leads to a huge matrix, which is always beyond the capacity of an ordinary PC, such as, unfolding *D* of Eq.(2) along the 1st dimension will generate a matrix of size $I_1 \times (I_2 \cdot ... \cdot MI_i \cdot I_{i+1} \cdot ... \cdot I_N)$ in the HO-SVD computation. The size of the unfolded matrix depends upon the number of samples *M* and the sample size $(I_1 \times ... \times I_N)$.

3 Reformulating ND-PCA Scheme

Introducing the multidirectional decomposition to Eq.(3) yield,

$$\begin{cases} \hat{X} = Y \times_1 U_{k_1}^{(1)} \times_2 ... \times_N U_{k_N}^{(N)} + \bar{X} \\ Y = (X - \bar{X}) \times_1 U_{k_1}^{(1)T} \times_2 ... \times_N U_{k_N}^{(N)T} \end{cases}, \quad (4)$$

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* of Eq.(2) in HO-SVD usually generates an overly large matrix.

First, we consider the case of unfolding *D* of Eq.(2) along the *i*th dimension, which generates a matrix of size $MI_i \times (I_{i+1} \cdot ... \cdot I_N \cdot I_1 \cdot ... \cdot I_{i-1})$. We prefer a unitary matrix $U^{(i)}$ of size $I_i \times I_i$ to that of the size $MI_i \times MI_i$. This can be achieved by reshaping the unfolded matrix as follows.

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

Then, consider the generic case. Since the size of each dimension $I_1, ..., I_N$ 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, $M=(M_1,...,M_m)$, maintains its own singular value σ_i, i.e. $M_iM_i^T = U_i diag(\sigma_i^2,0,...,0)U_i^T$, while the singular values of M are $s_1,...,s_{\min(m,n)}$, i.e. $M = Vdiag(s_1,...,s_{\min(m,n)})U^T$, then $\sum_{i=1}^{\min(m,n)} \sigma_i^2 = \sum_{i=1}^{\min(m,n)} s_i^2$.

Proof:

Let $n > m$. Because,

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

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

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

where v_i is the ith column of V, we thus have,

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

The 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 $M_iM_i^T = U_i diag(s_{1i}^2,...s_{ri}^2,...,0)U_i^T$. It can be noted that there are more than one non-zero singular values $s_{1i} \geq ... \geq s_{ri} \geq 0$. If we let $rank(M_iM_i^T)=1$, the approximation of $M_iM_i^T$ can be written as $M_iM_i^T \approx U_i diag(s_{1i}^2,0,...,0)U_i^T$. In terms of the lemma, we can also approximate it as $M_iM_i^T \approx M_{1i}M_{1i}^T = u_{1i}\sigma_{1i}^2u_{1i}^T$, 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 by sorting these singular values $\{\sigma_i\}$ and partition it into 2 block submatrices $\hat{M}=(M_1,M_2)$ (assume $m \geq n$ below), so that M_1 contains the columns corresponding to the first k biggest singular values while M_2 contains others. Note that $\hat{M}$ is different from the original M because of a column permutation (denoted as *Permute*). Applying SVD to each M_i respectively yields,

$$\hat{M} = (U_1,U_2)\begin{pmatrix}\Lambda_1 & \\ & \Lambda_2\end{pmatrix}\begin{pmatrix}V_1^T & \\ & V_2^T\end{pmatrix}. \qquad (5)$$

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

$$\hat{M} \approx \hat{M}' = (U_1,U_2)\begin{pmatrix}\Lambda_1 & \\ 0 & \end{pmatrix}\begin{pmatrix}V_1^T & \\ & V_2^T\end{pmatrix}. \qquad (6)$$

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.(6). 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,

$$\begin{cases}\hat{X} = Y \times_1 U_{k_1}^{(1)} ... \times_i U_{k_i}^{(i)} ... \times_N U_{k_N}^{(N)} + \bar{X} \\ Y = (X-\bar{X}) \times_1 U_{k_1}^{(1)T} ... \times_N U_{k_N}^{(N)T} \\ U_{k_i}^{(i)} \text{ is from Eq.(6)}\end{cases}. \qquad (7)$$

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

$$\varepsilon = \left\|\hat{X}-\hat{X}'\right\|_F = \left\|Y-Y'\right\|_F. \qquad (8)$$

Furthermore, we can give out the following proposition.

Proposition:

$\hat{X}$ of Eq.(7) is a near optimal approximation to sample X in a least-square sense.

(For proof, please refer to appendix.)

4 Experiments

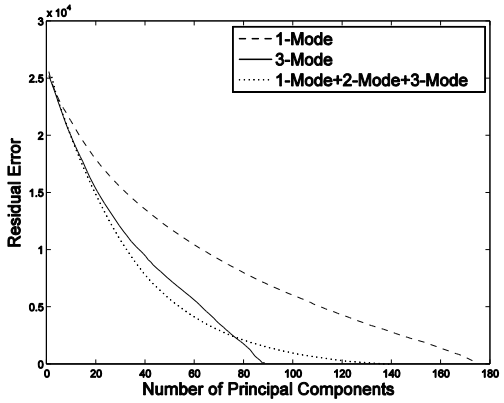
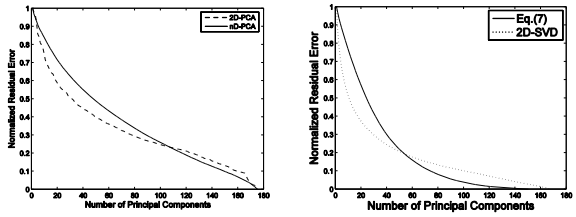


Fig.3. Comparison of the reconstruction through 1-mode, 3-mode and 1-mode+2-mode+3-mode principal subspace respectively. The scheme of Eq.(7) converges quicker than that of Eq.(3).

The proposed ND-PCA approach was performed on a 3D range database of human faces used for the Face Recognition Grand Challenge [3]. In order to establish an analogy with a 3D volume dataset or higher dimensional solid dataset, we embedded each 3D range dataset into a 3D array and mapped the pixels of the corresponding 2D face image to the voxels 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.

The experiment is to test the quality of the reconstructed sample. Within our 3D volume dataset, we got the 1-mode, 2-mode and 3-mode singular vectors, which can span three independent orthogonal spaces respectively. Our objective is to test which manner leads to the best reconstruction quality based on these three spaces. To this end, we first compare the residual errors of reconstructions by performing Eq.(7) on the 1-mode, 3-mode and 1-mode+2-mode+3-mode principal subspaces respectively. (Note that when Eq.(7) is performed on 1-mode or 3-mode principal subspaces, it will degenerate into Eq.(3).) The residual errors and conclusion are shown in Fig.3. To $U^{(1)}$ and $U^{(3)}$, as their dimensions are different, the ranges of principal component numbers k are different too. If the curve of 3-mode (solid curve) is quantified to the same length



a. single direction. b. multi-direction.

Fig.5 Comparison of the reconstructions by using single directional decomposition and multidirectional composition in terms of the normalized residual errors.

of row coordinate as the curve of 1-mode (dashed line) in Fig.3, there isn't substantial difference compared to the 1-mode test, i.e. the curve of 3-mode is similar to that of 1-mode. The reconstructed results based on Eq.(3) are not affected by the difference between the different n-mode principal component subspaces.

Furthermore, in the test of 1-mode+2-mode+3-mode principal component subspace, the numbers of principal components k are increased each time by 2 for both $U^{(1)}$ and $U^{(2)}$ while increased by 1 for $U^{(3)}$, and the maximum k are set to 180 for $U^{(1)}$ and $U^{(2)}$ while 90 for $U^{(3)}$.

To compare the multidirectional decomposition with the single dimensional ones, we show the reconstructed results of the single directional decomposition (i.e. 2D-PCA and ND-PCA scheme of Eq.(3)) in Fig.5a and the multidirectional decomposition (i.e. 2D-SVD and ND-PCA scheme of Eq.(6-7)) in Fig.5b. The residual errors of reconstruction are normalized to the range of [0,1]. One can note that the multidirectional decomposition performs better than the single directional decomposition in the case of a small number of principal components. But then Fig.5a (or Fig.5b) also seems to show that 2D-PCA (or 2D-SVD) performs a

little better than ND-PCA scheme of Eq.(3) (or Eq.(6-7)) 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 schemes. This is because the reconstructed 3D volume dataset is a sparse 3D array (i.e. only the voxel values on the face surface are not equal to zero but all the others are equal to zero), 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 in Fig.5a (or Fig.5b) would not noticeably appear.

5 Conclusions

In this paper, we amended our previous ND-PCA approach in [1] by introducing the multidimensional decomposition technique. The novelties of this paper include, 1) introducing the multidirectional decomposition and overcoming the numerical difficulty of large matrix SVD decomposition; 2) giving out the estimation of error bound. The experimental results indicate that the revised ND-PCA scheme could effectively improve the accuracy of reconstruction. In future work, we will apply the ND-PCA scheme to the multimodal face data fusion and recognition and develop a practical prototypical system.

Appendix

Proof.

According to the property 10 of HO-SVD in [10], we assume that the n-mode rank of $(X-\bar{X})$ be equal to $R_n (1 \le n \le N)$ and $(\hat{X}-\bar{X})$ 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 $\hat{X}$ is a near optimal approximation of sample X. The error is bounded by Frobenius-norm as follows,

$$\left\| X-\hat{X} \right\|_F^2 \le \sum_{i_1=I_1'+1}^{R_1} \sigma_{i_1}^{(1)2} + ... + \sum_{i_N=I_N'+1}^{R_N} \sigma_{i_N}^{(N)2} . \qquad \text{(A1)}$$

This means that the tensor $(\hat{X}-\bar{X})$ is in general not the best possible approximation under the given n-mode rank constraints. But under the error upper-bound of Eq.(A1), $\hat{X}$ 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.(5), i.e.

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

Let $\hat{M}' = (U_1, U_2) \begin{pmatrix} \Lambda_1 & \\ & 0 \end{pmatrix} \begin{pmatrix} V_1^T & \\ & V_2^T \end{pmatrix}$ as shown in Eq.(6). Consider the difference of $\hat{M}$ and $\hat{M}' \in R^{n \times m}$ as follows,

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

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 $\max\{\sigma : \sigma \in \Lambda_2\}$. As

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

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,

$$\left\| \hat{M} - \hat{M}' \right\|_2^2 \le 2 \max^2 \{\sigma : \sigma \in \Lambda_2\}, \qquad \text{(A2)}$$

in a 2-norm sense.

Substituting Eq.(A2) into Eq.(A1) yields the error upper-bound of $\hat{X}$ as follows,

$$\left\| X-\hat{X} \right\|_F^2 \le 2\left(\max^2 \left\{ \sigma^{(1)} : \sigma^{(1)} \in \Lambda_2^{(1)} \right\} + ... + \max^2 \left\{ \sigma^{(N)} : \sigma^{(N)} \in \Lambda_2^{(N)} \right\} \right).$$

This implies that the approximation $\hat{X}$ of Eq.(7) is a near optimal approximation of sample X under this error upper bound. **End of proof**.

References

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

[2]. H. Wang and N. Ahuja: Facial Expression Decomposition, Proc. of IEEE 9th Int'l Conf. on Computer Vision (2003), Vol.2.

[3]. P. Phillips and P. Flynn et al.: Overview of the Face Recognition Grand Challenge, Proc. of CVPR2005, Vol.1.

[4]. L. Wang and X. Wang et al.: The equivalence of two-dimensional PCA to line-based PCA, Pattern Recognition Letters, Vol.26, No.1, pp.57-60.

[5]. J. Yang and D. Zhang et al.: 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.

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

[7]. L.D. Lathauwer, B.D. Moor and J. Vandewalle: A Multilinear Singular Value Decomposition, SIAM J. on Matrix Analysis and Applications, Vol.21, No.4, pp.1253-1278.

[8]. M. Vasilescu, and D. Terzopoulos: Multilinear subspace analysis of image ensembles, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (2003), Vol.2.