Efficient tensor based face recognition

Santu Rana, Wanquan Liu, Mihai Lazarescu and Svetha Venkatesh
Dept. of Computing, Curtin University of Technology
GPO Box U1987, Perth, WA 6845, Australia.
{santu.rana, wanquan, m.lazarescu, svetha}@cs.curtin.edu.au

Abstract

This paper addresses the limitation of current multilinear PCA based techniques, in terms of prohibitive computational cost of testing and poor generalisation in some scenarios, when applied to large training databases. We define person-specific eigenmodes to obtain a set of projection bases, wherein a particular basis captures variation across lightings and viewpoints for a particular person. A new recognition approach is developed utilizing these bases. The proposed approach performs on a par with the existing multilinear approaches, whilst significantly reducing the complexity order of the testing algorithm.

1 Introduction

Facial image recognition across different lightings, viewpoints and expressions is a challenging task. Linear methods [8] are usually inaccurate and non-linear methods [6] lead to high computational cost. Recently, multilinear models [10] [9] have been proposed as a solution, with higher accuracy than the linear models and lower computational costs than the non-linear methods. With this framework, facial images are organised as a data tensor, with different factors of variation (*i.e.* lighting, viewpoint *etc.*) modeled as different modes of the tensor. Factorization of the tensor with Higher Order SVD [2], a generalisation of SVD for higher order matrices, separates subspaces corresponding to different factors of variation and a core tensor that contains the information on how different factors interact with each other.

A decomposition approach has been developed in [9] to provide a recognition algorithm that performs better than the linear models. However, only the person-mode decomposition is utilized effectively, whilst decomposition in other modes is used optionally, only for dimensionality reduction. More precisely, they use *lighting-viewpoint specific eigenmodes* to generate a set of projection basis for recognition. In [7], the authors argue that the previous recognition approach is not suitable for face recognition under novel lightings and viewpoints as the projection bases do not contain information on the variation of lighting and viewpoints. Instead of *lighting-viewpoint specific eigenmodes* as in [9], they proposed the eigenmodes encompassing all the modes of variation. These *multilinear eigenmodes* are able to capture the variations across persons, lightings and viewpoints and provide an intuitive basis for face image representation while at the same time being effective for recognition. The recognition in this framework requires projection of a test image into the *multilinear eigenmode* space and then comparing the coefficients with the stored coefficients vectors of all the training images.

However, further investigation of their algorithm reveals the following shortcomings,

- Though the space of *multilinear eigenmodes* provides a general framework for representation, it is inefficient for testing on a large database, as the testing involves comparison with the coefficients vectors of all the training images, resulting in a testing complexity of $O(N_T)$, where N_T is the total number of training images.
- When there are multiple modes of variation with very few examples in each of them, existing recognition algorithm have trouble dealing with the sparseness of training exemplar in the high-dimensional *multilinear eigenmode* space (the size of the space increases in a multiplicative way with the number of modes). This leads to poor generalisation by the classifier and results in a poor performance.

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

Based on these observations, we propose a new recognition approach to overcome the shortcomings outlined above. More precisely, we define *person-specific eigenmodes*, that capture the variation across lightings and viewpoints for a particular person. These *person-specific eigenmodes* are used as components of a person-specific basis. A set of bases is defined with one particular basis corresponding to one particular person. Each of these bases spans a limited subspace of the whole image space, therefore, projections on these bases result in loss of information. However, projecting an image onto the same person-specific basis the image belongs to, results in the smallest loss of information. We use the reconstruction error to quantify the loss of information. The basis, which gives the lowest reconstruction error, reveals the identity of the test image. The proposed method, can be thought of as a compartmentalization of the general *multilinear eigenmode* space corresponding to different persons. Testing involves projection to N_p (number of persons in the database) number of bases and comparison of N_p number of reconstruction error, resulting in a testing complexity of only $O(N_p)$. As $N_p << N_T$, we overcome the inefficiency of [7], outlined above. Moreover, as the individual *person-specific eigenmode* spaces are much smaller than the space of *multilinear eigenmodes*, we overcome the generalisation problem as well.

Experimentation shows that in terms of recognition accuracy the proposed method performs on a par with the existing multilinear based methods in most cases, and much better in one specific case. In terms of computational cost of testing, it is around 50 times faster for the largest database and around 3 times faster for the smallest database tested.

2 Background

2.1 Tensor Model of Face database

Let us assume that the face database contains images with variations in lighting and viewpoint only. The tensor representation of the database is presented by,

$$T(i_p, i_l, i_v) = I_{P_{i_p}, L_{i_l}, V_{i_v}} \tag{1}$$

where, $I_{P_{i_p}, L_{i_l}, V_{i_v}}$ is the image vector of i_p'th person at i_l'th lighting and i_v'th viewpoint. T is a tensor of order 4 and $T \in \mathbb{R}^{N_p \times N_l \times N_v \times N_x}$, where, N_p is the number of persons, N_l and N_v represent the number of lighting and viewpoint instances respectively and N_x is the size of the image vector.

2.2 Multilinear PCA

Multilinear PCA for an ensemble of images is performed by computing the Higher Order SVD [2] of the tensor T. HOSVD of tensor T yields four orthogonal subspaces, wherein each subspace corresponds to one mode of variation. This is represented as follows:

$$T = S \times_1 U^P \times_2 U^L \times_3 U^V \times_4 U^X \tag{2}$$

where, $\times_k$ denotes *mode-k* product. The columns of U^P, U^L, U^V and U^X define the person, lighting, viewpoint and the pixel subspaces respectively. The columns in U^X represent traditional *eigenfaces* and the columns of U^P, U^L and U^V, represent the $N'_p(\leq N_p)$, $N'_l(\leq N_l)$ and $N'_v(\leq N_v)$ dominant eigenvectors of the person, lighting and viewpoint subspaces respectively. We refer to them as *eigen-person*, *eigen-lighting* and *eigen-viewpoint* respectively. S is called the core tensor.

2.3 Existing Approach: MPCA-JS

MPCA-JS [7] uses *multilinear eigenmodes* as the basis for face image representation and recognition. From (2) let $A = S \times_4 U^X$ and if $\mathcal{A}_{person}$ denotes the unfolding of the tensor $\mathcal{A}$ in the person mode, then

$$\mathcal{A}_{person} = \begin{bmatrix} I_{P_1^e L_1^e V_1^e} & .. & I_{P_1^e L_{N'_l}^e V_{N'_v}^e} \\ I_{P_2^e L_1^e V_1^e} & .. & .. \\ .. & .. & .. \\ I_{P_{N'_p}^e L_1^e V_1^e} & .. & I_{P_{N'_p}^e L_{N'_l}^e V_{N'_v}^e} \end{bmatrix} \tag{3}$$

where, $I_{P_i^e L_j^e V_k^e}$ is one specific *multilinear eigenmode*. These *multilinear eigenmodes* are used as the components for their projection basis. The coefficient vectors of projection for the training images are calculated separately using U^P, U^L, U^V and stored for testing. Testing involves projecting the test image on the basis and comparing them with the stored coefficient vectors, resulting in a testing complexity of $O(N_p \times N_l \times N_v)$.

3 Proposed Approach

Motivated by our desire to reduce the complexity of the testing algorithm for face recognition using tensor framework, we present a new algorithm utilising *person-specific eigenmodes* as the basis for projection. We derive our approach starting from the multilinear PCA decomposition of face datatensor T, as in (2) . Let us denote, $\mathcal{B} = S \times_1 U^P$

and if $\mathcal{B}_{person}$ denotes the unfolding of the tensor $\mathcal{B}$ in the person mode, then

$$\mathcal{B}_{person} = \begin{bmatrix} I^e_{P_1 L^e_1 V^e_1} & .. & I^e_{P_1 L^e_{N'_l} V^e_{N'_v}} \\ I^e_{P_2 L^e_1 V^e_1} & .. & .. \\ .. & .. & .. \\ I^e_{P_{N_p} L^e_1 V^e_1} & .. & I^e_{P^e_{N_p} L^e_{N'_l} V^e_{N'_v}} \end{bmatrix} \quad (4)$$

where $I^e_{P_{i_p} L^e_{i_l} V^e_{i_v}}$ is the *eigen-image* of the i_p'th person at i_l'th *eigen-lighting* and at i_v'th *eigen-viewpoint*. Let us define,

$$B_i = \begin{bmatrix} I^e_{P_i L^e_1 V^e_1} \\ I^e_{P_i L^e_2 V^e_1} \\ \cdots \\ I^e_{P_i L^e_{N'_l} V^e_{N'_v}} \end{bmatrix} \quad (5)$$

where, B_i for $i = 1, ..., N_p$ are the person-specific bases. For a test image I_T, the projection on the basis B_k will generate a description on how the images of person k at different *eigen-lighting* and *eigen-viewpoint* are combined to create the image. B_k spans only a limited subspace, corresponding to the images of k'th person, out of the whole image space. Therefore, projections on B_k can only preserve a certain fraction of information of the image I_T. However, if I_T belongs to person k then a projection on B_k will preserve the maximum information of the image, out of all $\{B_i\}_{i=1}^{N_p}$. We will use the minimum reconstruction error as the criterion to quantify the information preserved in a certain person-specific base. The algorithm for testing is given in Algorithm 1.

Algorithm 1 Testing algorithm for the proposed method

Testing: I_T is the test image.

1. Calculate $I'_T = I_T \times U^X$.
2. For every $i = 1, ..., N_p$ calculate
   - $c_i = I'_T \times B_i$
   - $I_T^{recon'} = c_i \times B_i^+$ ($^+ \implies$ *pseudoinverse*)
   - $e_i = ||I'_T - I_T^{recon'}||$
     The distance measure we use is Euclidean distance.
3. If e_k is the smallest of $\{e_i\}_{i=1}^{N_p}$ then I_T belongs to the person k.

3.1 Complexity analysis

Complexity analysis of our proposed algorithm shows that we have N_p matrix multiplications and N_p distance calculations. Therefore, our proposed method is of order $O(N_p)$.

4 Experiments and Analysis

We used PEAL [3], YaleB Frontal [4], Extended YaleB [5] and Weizmann face databases for experiments. The Extended YaleB database contains images of 38 persons at 64 different lightings and at 9 different viewing directions. The YaleB frontal is a subset of the Extended YaleB database, which contains images of 38 persons at 64 different lighting conditions, at the frontal viewpoint only. PEAL is a face-database of Chinese nationals at different poses, lightings and expressions, from which we chose images of 20 persons at 20 different lighting conditions only. For the Weizmann face database, we have access to face images of 28 persons at 5 different viewpoints ($0°, \pm17°, \pm34°$), 3 different lightings and 2 different expressions. Prior to the experiments, all the images were cropped and their eye-points were manually aligned. Then all the image vectors were normalized to unity. For HOSVD and other tensor operations, we used the tensor toolbox developed by Kolda *et. al.* in MATLAB^{TM} [1].

For the PEAL dataset, four sets of experiments were performed with randomly chosen 5, 7, 9 and 11 lighting conditions as training. The rest were used for testing. Similarly, for the YaleB Frontal dataset, four sets of experiments were performed with randomly chosen 5, 10, 15 and 20 lighting conditions for training and the rest for testing. For experiments on the Extended YaleB database, 16 representative lighting conditions at 4 representative viewpoints were used for training and the rest for testing. For experiments on Weizmann face database 3 representative viewpoints (at $0°$ and $\pm34°$), at 2 randomly selected lighting conditions and at all the expressions were used for training and the rest for testing. Thus, in both the PEAL and YaleB Frontal database, the test images were at unseen lighting. For both Extended YaleB database and Weizmann database, the test images were either at unseen lighting or unseen viewpoints or both. For the experiments on PEAL and YaleB Frontal, each set of experiments was repeated 20 times and the average results are reported. The performance of the proposed recognition procedure is compared with the PCA [8] and MPCA-JS [7] methods of face recognition. All the timings reported here are for Matlab code running on a Pentium dual core 1.86GHz system, with 2GB RAM.

We have used the same *energy thresholding* as [7] to select *eigenvectors*. The distance measures were also same for PCA and MPCA-JS, with the proposed method using Euclidean distance measure.

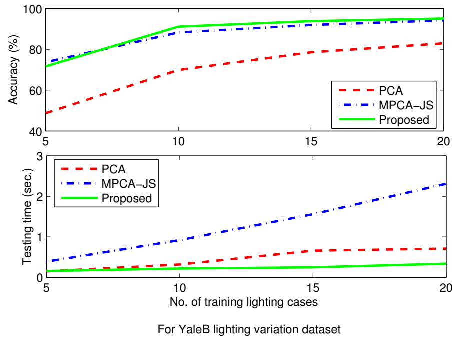

Figure 1: Results on YaleB lighting variation dataset. The testing time is for 100 test images.

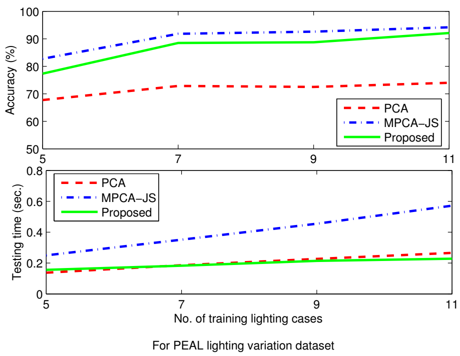

Figure 2: Results on PEAL lighting variation dataset. The testing time is for 100 test images.

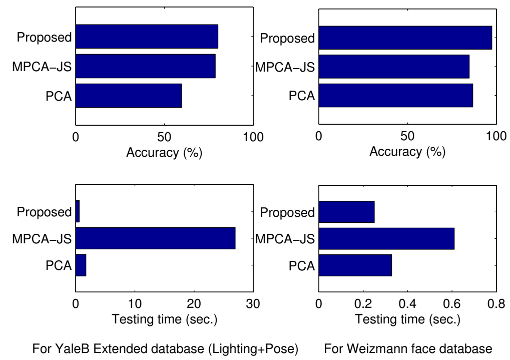

Figure 3: Results on YaleB Extended and Weizmann face database. The testing time is for 100 test images.

Figure 1 and 2 show the results of experiments with only unseen lightings. While for PEAL database the proposed method performed slightly worse, for YaleB database it performed slightly better. However, in both the cases, the proposed method provides significant speed up for testing.

Figure 3 shows the result of experiments for unseen lightings and viewpoints. For Extended YaleB the performance gain in accuracy is slight, however, testing is almost 50 times faster than MPCA-JS. For Weizmann database, the proposed method produced much better results. Interestingly, in this case, MPCA-JS fared worse than PCA, as the database has 4 modes of variation and only a few examples in viewpoint and expression modes, resulting in high-dimensional, but sparse multilinear space. However, the proposed method, by restricting itself into person-specific multilinear space, was able to produce much better performance.

5 Conclusion

In this paper we proposed a novel face recognition method using *person-specific eigenmodes* as the projection bases. This new approach is capable to equal or better the recognition performance of existing multilinear based methods at a significantly lower computational cost for testing.

References

[1] B. W. Bader and T. G. Kolda. Tensor toolbox version 2.2, Sandia National Laboratories.

[2] L. De Lathauwer, B. D. Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[3] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The cas-peal large-scale chinese face database and baseline evaluations. *IEEE SMC Part A*, 38(1):149–161, Jan. 2008.

[4] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE PAMI*, 23(6):643–660, Jun 2001.

[5] K.-C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE PAMI*, 27(5):684–698, May 2005.

[6] Q. Liu, R. Huang, H. Lu, and S. Ma. Kernel-based nonlinear discriminant analysis for face recognition. *J. Comput. Sci. Tech.*, 18(6):788–795, 2003.

[7] S. Rana, W. Liu, M. Lazarescu, and S. Venkatesh. Recognising faces in unseen modes: a tensor based approach. *Proceedings CVPR'08*.

[8] M. Turk and A. Pentland. Face recognition using eigenfaces. *Proceedings CVPR '91*, pages 586–591.

[9] M. Vasilescu and D. Terzopoulos. Multilinear image analysis for facial recognition. *Proceedings ICPR '02*, 2:511–514 vol.2.

[10] M. A. O. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: Tensorfaces. *Proceedings ECCV '02-Part I*, pages 447–460.