Non-negative Principal Component Analysis for Face Recognition

Zhang Yan, Yu Bin

Open Science Index, Electronics and Communication Engineering Vol:4, No:12, 2010 waset.org/Publication/14158

***Abstract*—**Principle component analysis is often combined with the state-of-art classification algorithms to recognize human faces. However, principle component analysis can only capture these features contributing to the global characteristics of data because it is a global feature selection algorithm. It misses those features contributing to the local characteristics of data because each principal component only contains some levels of global characteristics of data. In this study, we present a novel face recognition approach using non-negative principal component analysis which is added with the constraint of non-negative to improve data locality and contribute to elucidating latent data structures. Experiments are performed on the Cambridge ORL face database. We demonstrate the strong performances of the algorithm in recognizing human faces in comparison with PCA and NREMF approaches.

***Keywords*—**classification, face recognition, non-negative principle component analysis (NPCA)

I. Introduction

FACE recognition is an important research problem which can be formulated as verifying or determining of the person in the input image in the given database of face images. Biometric-based technologies include identification based on physiological characteristics (such as face, fingerprints, finger geometry, hand veins, palm, iris, retina, ear and voice) and behavior traits (such as gait, signature and keystroke dynamics) [1]. Face recognition appears to offer several advantages over other biometric methods such as it can be done passively without any explicit action or participation on the part of the user since face images can be acquired from a distance by a camera, therefore it is a totally non-intrusive and does not carry any health risks. There have been numerous practical applications used for two primary tasks: verification (one-to-one matching) and identification (one-to-many matching). A few of the applications can be outlined as: security (access control, airports/seaports, ATM machines and border checkpoints [2, 3]; network security [4]; email authentication on multimedia workstations); surveillance; general identity verification (national IDs, passports, drivers' licenses); criminal justice systems; video indexing etc. In addition to these applications, the underlying techniques in the current face recognition technology have also been modified and used for related applications such as gender classification [5-7], expression recognition [8, 9] and facial feature recognition and tracking [10].

In [11], a formal method of classifying faces was first proposed. Great progress has been made and the methods fall into two main categories: feature-based and holistic [12-14].

Feature-based approaches first identify and extract distinctive facial features such as the eyes, mouth, nose, etc., as well as other c marks, and then compute the geometric relationships among those facial points, thus reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition techniques are then employed to match faces using these measurements. Wiskott etc. proposed the elastic bunch graph matching method which is based on Dynamic Link Structures [15]. Recent variations of this approach replace the Gabor features by a graph matching strategy [16] and HOGs (Histograms of Oriented Gradients) [17]. The major disadvantage of these approaches is the difficulty of automatic feature detection and the fact that the implementer of any of these techniques has to make arbitrary decisions about which features are important [18].

Numerous variations on and extensions to the standard eigenfaces and the Fisherfaces approaches have been suggested since their introduction. As a well-established dimension reduction technique, Principal Component Analysis (PCA) projects data in an orthogonal subspace generated by the eigenvectors of the data covariance matrix. Some recent advances in PCA-based algorithms include multi-linear subspace analysis [19], symmetrical PCA [20], two-dimensional PCA [21], weighted modular PCA [22], Kernel PCA [23], and diagonal PCA [24]. However, PCA can only capture these features contributing to the global characteristics of data because it is a global feature selection algorithm. It misses those features contributing to the local characteristics of data because each principal component only contains some levels of global characteristics of data [25]. This global feature selection mechanism not only leads to difficulty in interpreting each principal component but also prevents subtle data local latent structure discovery in the following classification.

One important reason for the global nature of PCA is that data representation in the classic PCA is not "purely additive," i.e., each principal component consists of both negative and positive entries. The linear combination in the PCA subspace mixes with both positive and negative weights which are likely to cancel each other partially in the data representation. In fact,

Zhang Yan is with Qingdao University of Science and Technology, Qingdao, 266061, China (e-mail: zy@qust.edu.cn).

Yu Bin is with the College of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao, 266061, China (e-mail: yubin@qust.edu.cn).

Open Science Index, Electronics and Communication Engineering Vol:4, No:12, 2010 waset.org/Publication/14158

it is more likely that weights from local features are partially canceled out than the weights from global features for their frequencies. Therefore, the partial cancellations lead to PCA's loss of data locality in the feature selection [26].

Imposing non-negative constraints on PCA, restricting each entry of a principal component to be non-negative, can entirely prevent the partial cancellations because each input data variable is represented by only additive components, therefore, it will improve data locality and contribute to elucidating latent data structures. In fact, that is the reason why non-negative matrix factorization has been shown to be a useful decomposition for multivariate data. Moreover, non-negativity also causes each principal component to be sparse, which contributes to the intuitive interpretation of each principal component [27].

In the following sections, we formulate the non-negative principal component analysis (NPCA) as a constraint optimization problem and present a NPCA-based face recognition algorithm. In section II, we introduce the PCA and demonstrate the NPCA. Furthermore, we present a NPCA-based face recognition algorithm. Finally, we discuss the generalizations and limitations of the NPCA-based face recognition algorithm and conclude the paper.

II. Methods

A. *Principal Component Analysis*

Principal component analysis (Karhunen-Loeve or Hotelling transform) - PCA , providing a powerful tool for data analysis and pattern recognition which is often used in signal and image processing, belongs to linear transform based on the statistical techniques.

Principle component analysis in signal processing can be described as a transform of a given set of n input vectors (variables) with the same length K formed in the n-dimensional vector $X=[x_1,x_2,\cdots x_n]^T$ into a vector y according to

$$y = A(X - m_x) \tag{1}$$

This point of view enables to form a simple formula (1) but it is necessary to keep in the mind that each row of the vector X consists of K values belonging to one input. The vector m_x in Eq. (1) is the vector of mean values of all input variables defined by relation

$$m_x = E\{X\} = \frac{1}{K}\sum_{k=1}^{K} X_k \tag{2}$$

Matrix A in Eq. (1) is determined by the covariance matrix C_x . Rows in the A matrix are formed from the eigenvector e of C_x ordered according to corresponding eigenvalues in descending order. The evaluation of the C_x matrix is possible according to relation

$$C_x = E\{(X - m_x)(X - m_x)^T\} = \frac{1}{K}\sum_{k=1}^{K} X_k X_k^T - m_x m_x^T \tag{3}$$

As the vector X of input variables is n-dimensional it is obvious that the size of C_x is $n\times n$. The elements $C_x(i,i)$ lying in its main diagonal are the variances

$$C_x(i,i) = E\{(X_i - m_i)^2\} \tag{4}$$

of X and the other values $C_x(i,j)$ determine the covariance between input variables X_i, X_j .

$$C_x(i,j) = E\{(X_i - m_i)(X_j - m_j)\} \tag{5}$$

The rows of A in Eq. (1) are orthonormal so the inversion of PCA is possible according to relation

$$X = A^T y + m_x \tag{6}$$

The kernel of PCA defined by Eq. (1) has some other interesting properties resulting from the matrix theory which can be used in the signal and image processing to fulfil various goals as mentioned below.

Matrix factorization is described as follows: given a matrix X , with m rows and n columns, and a positive integer γ , $\gamma \le rank(X)$, find matrix factors W and H such that

$$X \approx WH = \sum_a W_{ia} H_{aj} \tag{7}$$

where W , H are matrices with rows m,r and columns r,n respectively. $\approx$ means that you may choose your own objective function to measure the distance between X and WH and minimize it to find W and H .
The objective function of PCA is

$$D(X \| WH) = \sum_{i,j}(X_{i,j} - (WH)_{i,j})^2 + \alpha\sum_{i\neq j}\left|(W^T W)_{i,j}\right| \tag{8}$$

subject to

$$(W^T W)_{i,i} = 1, \qquad \alpha > 0 \tag{9}$$

The second term is used to make the bases W near to orthogonal bases. It is evident that

$$D(X \| WH) > \sum_{i,j}(X_{i,j} - (WH)_{i,j})^2 \tag{10}$$

According to the properties of the singular value decomposition, from Eq. (10), we have that if $W = U$, $H = \sum V^T$ where U and V are the first γ eigenvectors of the XX^T and $X^T X$ matrices respectively, $\Sigma = diag(\lambda_1,\cdots,\lambda_\gamma)$, $\lambda_1,\cdots,\lambda_\gamma$ are the first γ positive eigenvalues of XX^T , then (W,H) is the minimum solutions of Eq. (8).

B. *Non-negative Principal Component Analysis*

The objective function of NPCA is:

$$D(X \| WH) = \sum_{i,j}(X_{i,j} - (WH)_{i,j})^2 + \alpha\sum_{i\neq j}\left|(W^T W)_{i,j}\right| \tag{11}$$

subject to

$$(W^T W)_{i,i} = 1, W_{i,j} \ge 0, H_{i,j}, \alpha \ge 0 \tag{12}$$

which is used to make the base W near to orthogonal bases, so that there is little redundancy information in W . α is chosen to make the scalar of the second term be near to the first term in Eq. (11).

We can derive the iterative formulae by so-called auxiliary functions [28]. The auxiliary function $G_H(H,H')$ with respect to H, H' and $G_W(W,W')$ with respect to W,W' are:

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:4, No:12, 2010

Open Science Index, Electronics and Communication Engineering Vol:4, No:12, 2010 waset.org/Publication/14158

$$\begin{aligned} G_H(H,H') = & \frac{1}{2}\sum_{i,j}(X_{i,j}-(WH')_{i,j})^2 \\ & +\sum_j\sum_k(H_{k,j}-H'_{k,j})\sum_i(X_{i,j}-(WH')_{i,j})(-W_{i,k}) \\ & +\frac{1}{2}\sum_i\sum_j(H_{i,j}-H'_{i,j})^2(W^TWH')_{i,j}/H'_{i,j} \\ & +\alpha\sum_{i,j,i\neq j}(W^TW)_{i,j} \end{aligned} \quad (13)$$

$$\begin{aligned} G_W(W,W') = & \frac{1}{2}\sum_{i,j}(X_{j,i}-(W'H)_{j,i})^2 \\ & +\sum_j\sum_k(W_{j,k}-W'_{j,k})\sum_i(X_{j,i}-(W'H)_{j,i})(-H_{k,i}) \\ & +\frac{1}{2}\sum_i\sum_j(W_{j,i}-W'_{j,i})^2(HH^TW'^T)_{i,j}/W'_{j,i} \\ & +\alpha\sum_{i,j,k}\left[\frac{1}{2}(W_{k,i}-W'_{k,i})^2+\frac{1}{2}(W_{k,j}-W'_{k,j})^2 \right. \\ & \left. -W'_{k,i}W'_{k,j}+W'_{k,i}W_{k,j}+W_{k,i}W'_{k,j}\right] \end{aligned} \quad (14)$$

According to the property that the minimum of the auxiliary function is just the minimum of the corresponding objective function, then the iterative formulae are: $H^{n+1}=\arg\min_H G_H(H,H^n)$ and $W^{n+1}=\arg\min_W G_W(W,W^n)$. In fact, we can obtain the local minimum of $G_H(H,H^n)$ and $G_W(W,W^n)$ through some different calculation. Thus iterative formulae are as follows:

$$H^{n+1}_{i_0,j_0}=H^n_{i_0,j_0}(W^{nT}X)_{i_0,j_0}/(W^{nT}W^nH^n)_{i_0,j_0} \quad (15)$$

$$\begin{aligned} W^{n+1}_{i_0,j_0}= & (2*\alpha*r*W^n_{i_0,j_0}-\alpha\sum_{k\neq j_0}W^n_{i_0,j_0}+(XH^{nT})_{i_0,j_0}) \\ & /(2*r*\alpha+(W^nH^nH^{nT})_{i_0,j_0}/W^n_{i_0,j_0}) \end{aligned} \quad (16)$$

It is reasonable to require the corresponding dimension reduction data of the matrix to be positive or at least nonnegative to maintain data locality in the feature selection since the matrices themselves can be nonnegative in many applications.

III. Experiments

The experiments are performed on the Cambridge ORL face database, which contains 40 distinct persons. Each person has ten different images of the size of 112×92, taken at different times. There are variations in facial expressions such as open/closed eyes, smiling/nonsmiling, and facial details such as glasses/no glasses. All the images were taken against a dark homogeneous background with the subjects in an up-right, frontal position, with tolerance for some side movements. There is also some variations in scale. We show four individuals in the ORL face images in Fig. 1.

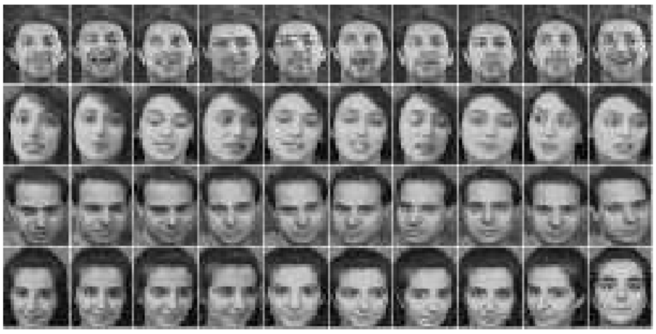

Fig. 1 Four individuals in the ORL face database.

Fig. 2, Fig. 3 and Fig. 4 shows the basis images of the training set learnt by PCA, Non-negative Relative Entropy Matrix Factorization (NREMF) [29, 30], and NPCA approaches respectively. They show 49 basis images of each approach on the ORL face database. It can be seen that the basis of all methods are additive except for PCA. Moreover, the greater number of basis image is, the more localization is learnt in NMF-based approaches.

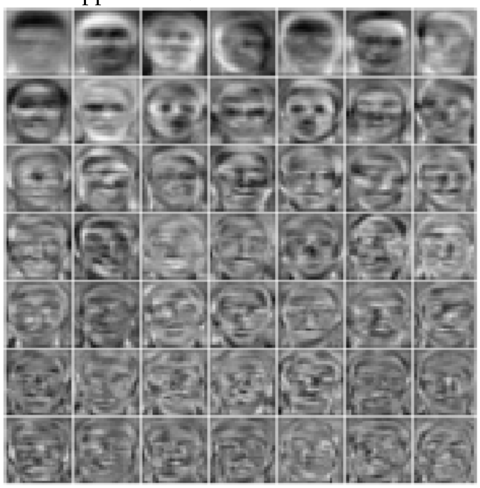

Fig. 2 The basis images of PCA on ORL database results.

Fig. 3 The basis images of NREMF on ORL database results.

Open Science Index, Electronics and Communication Engineering Vol:4, No:12, 2010 waset.org/Publication/14158

Fig. 4 The basis images of NPCA on ORL database results.

We randomly select n (n=2, 4, 6) images from each person for training, while the rest of images of each individual for testing. The average accuracies of training samples ranging from 2 to 5 are recorded in Table . The recognition accuracies of PCA, NREMF, and NPCA are 34.1%, 66.7%, and 67.2%, respectively, with 2 training images. The performance for each method is improved when the number of training images increases. The recognition accuracies of PCA, NREMF, and NPCA are 92.8%, 93.5%, and 96.3%, respectively, with 6 training images. The recognition ratio of the test set reached 100% in each approach.

TABLE I
THE AVERAGE ACCURACIES OF TRAINING SAMPLES

n	Recognition Ratio		
	PCA	NREMF	NPCA
2	0.341	0.667	0.672
4	0.875	0.910	0.945
6	RATR[a]=1 RATE[b]=0.928	RATR=1 RATE=0.935	RATR=1 RATE=0.963

[a]The abbreviation for the recognition accuracies of the training set.
[b]The abbreviation for the recognition accuracies of the test set.

IV. CONCLUSION

In this study, we present a novel face recognition approach using non-negative principal component analysis. We demonstrate the strong performances of the algorithm in recognizing human faces in comparison with PCA and NREMF approaches.

Since feature selection algorithms are widely employed in analyzing all types of expression data, it would be interesting to further investigate the potentials of the NPCA-related techniques in the classifications and biomarker captures for the proteomics, SNP, and array-based CGH data [31, 32]. Furthermore, we are also interested in investigating the applications of the sibling algorithm of NPCA: non-negative independent component analysis [33, 34] and the other classic or novel feature selection algorithm such as kernel independent component analysis [35], convex and seminonnegative matrix factorizations [36] in microarray data analysis. These will be further investigated in future work.

ACKNOWLEDGMENT

This work was supported by grants from the National Natural Science Foundation of China (No. 3057l059).

REFERENCES

[1] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: Personal Identification in Networked Security*, Kluwer Academic Publishers, 1999.
[2] K. Kim, "Intelligent Immigration Control System by Using International Symposium on Neural Networks," in *Proc. The International Symposium on Neural Networks*, Chongqing, 2005, pp. 147–156.
[3] J. N. K. Liu, M. Wang, and B. Feng, "iBotGuard: an Internet-based intelligent robot security system using invariant face recognition against intruder," *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews*, vol. 35, pp. 97-105, 2005.
[4] H. Moon, "Biometrics Person Authentication Using Projection-Based Face Recognition System in Verification Scenario," in *Proc. Of International Conference on Bioinformatics and its Applications*, Hong Kong, 2004, pp. 207–213.
[5] K. Balci, V. Atalay, "PCA for Gender Estimation: Which Eigenvectors Contribute?," in *Proc. Of 16th International Conference on Pattern Recognition*, vol. 3, Qubec, 2002, pp. 363–366.
[6] B. Moghaddam, M. H. Yang, "Learning Gender with Support Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 707-711, 2002.
[7] R. Brunelli, T. Poggio, "HyperBF Networks for Gender Classification," in *Proc. Of DARPA Image Understanding Workshop*, 1992, pp. 311–314.
[8] A. Colmenarez, B. J. Frey, T. S. Huang, "A Probabilistic framework for embedded face and facial expression recognition," in *Proc. Of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, Ft. Collins, CO, 1999, pp. 1592–1597.
[9] Y. Shinohara, N. Otsu, "Facial Expression Recognition Using Fisher Weight Maps," in *Proc. Of 16th IEEE International Conference on Automatic Face and Gesture Recognition*, vol. 100, 2004, pp. 499–504.
[10] F. Bourel, C. C. Chibelushi, A. A. Low, "Robust Facial Feature Tracking," in *Proc. Of British Machine Vision Conference*, Bristol, 2000, pp. 232–241.
[11] F. Galton, "Personal Identification and Description," *Nature*, pp. 173–177, June 21, 1888.
[12] R. Brunelli, T. Poggio, "Face Recognition: features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, 1993, pp. 1042–1052.
[13] M. A. Grudin, "On Internal Representations in Face Recognition Systems," *Pattern Recognition*, vol. 33, pp. 1161–1177, 2000.
[14] B. Heisele, P. Ho, J, Wu, and T. Poggio, "Face Recognition: Component-based versus global approaches," *Computer Vision and Image Understanding*, vol. 91, pp. 6-21, 2003.
[15] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, 1997, pp. 775-77.
[16] H. Shin, S. D. Kim, and H.C. Choi, "Generalized Elastic Graph Matching for Face Recognition," *Pattern Recognition Letters*, vol. 28, pp. 1077-1082, 2007.
[17] A. Albiol, D. Monzo, A. Martin, J. Sastre, "Face Recognition using HOG-EBGM," *Pattern Recognition Letters*, vol. 29, pp. 1537-1543, 2008.
[18] R. Cendrillon and B. C. Lowel, "Real-time Face Recognition using Eigenfaces," in *Proceedings of the SPIE International Conference on Visual Communications and Image Processing*, vol. 4067, 2000, pp. 269–276.
[19] M. A. O. Vasilescu, D. Terzopoulos, "Multilinear Subspace Analysis of Image Ensemblers," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2003, pp. 93–99.
[20] Q. Yang, X. Q. Ding, "Symmetrical Principal Component Analysis and Its Application in Face Recognition," *Chinese Journal of Computers*, vol. 26, pp. 1146-1151, 2003.
[21] J. Meng, W. Zhang, "Volume measure in 2DPCA-based Face Recognition," Pattern Recognition Letters, vol. 28, pp. 1203-1208, 2007.
[22] A. P. Kumer, S. Das, V. Kamakoti, "Face Recognition using weighted modular principle component analysis," *Neural Information Processing*,

vol. 3316, *Lecture Notes In Computer Science*: Springer Berlin / Heidelberg, 2004, pp. 362-367.

[23] N. Sun, H. Wang, Z. Ji, C. Zou, L. Zhao, "An Efficient algorithm for Kernel Two-dimensional Principal Component Analysis," *Neural Computing & Applications*, vol. 17, pp. 59-64, 2008.

[24] D. Zhang, Z. Zhoua, S. Chen, "Diagonal Principal Component Analysis for Face Recognition," *Pattern Recognition*, vol. 39, pp. 140-142, 2006.

[25] D. Lee, H. Seung, "Learning the Parts of Objects by Non-Negative Matrix Factorization," *Nature*, vol. 401, pp. 1788-1791, 1999.

[26] X. Han, "Nonnegative Principal Component Analysis for Cancer Molecular Pattern Discovery," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 7, pp. 537-549, 2010.

[27] P. Hoyer, "Non-Negative Matrix Factorization with Sparseness Constrains," J. Machine Learning Research, vol. 5, pp. 1457-1469, 2004.

[28] A. Dempster, N. Laird, D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Royal statistical Society B*, Vol. 39, pp. 1-38, 1977.

[29] D. D. Lee and H. S. Seung, "Learning The Parts of Objects by Nonnegative Matrix Factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[30] D. D. Lee and H. S. Seung, "Algorithms for Nonnegative Matrix Factorization," in *Proceedings of Neural Information Processing Systems*, 2000, pp. 556–562.

[31] C. Li et al., "Major Copy Proportion Analysis of Tumor Samples Using SNP Arrays," *BMC Bioinformatics*, vol. 9, no. 204, 2008, doi: 10.1186/1471-2105-9-20.

[32] J. Liu, S. Ranka, and T. Kahveci, "Classification and Feature Selection Algorithms for Multi-Class CGH Data," *Bioinformatics*, vol. 24, pp. 186-195, 2008.

[33] M. Plumbley and E. Oja, "A 'Nonnegative PCA' Algorithm for Independent Component Analysis," *IEEE Trans. Neural Networks*, vol. 15, pp. 66-76, Jan. 2004.

[34] M. Plumbley, "Algorithms for Nonnegative Independent Component Analysis," *IEEE Trans. Neural Networks*, vol. 4, pp. 534-543, May 2003.

[35] F. Bach and M. Jordan, "Kernel Independent Component Analysis," *J. Machine Learning and Research*, vol. 3, pp. 1-48, July 2002.

[36] C. Ding, T. Li, and M. Jordan, "Convex and Semi-Nonnegative Matrix Factorizations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, pp. 45-55, Jan. 2010.

Zhang Yan received the BE degree in computer science and technology from Northwestern Polytechnic University, Xi'an Shanxi, China, in 2004, the ME degree in control theory and control engineering from the Qingdao University of Science and Technology, Qingdao, Shandong, China, in 2009. His research interests include wavelet theory, pattern recognition and image processing.

Yu Bin received the ME degree in computational mathematics from Shanghai University, Shanghai, China, in 2005. His research interests include bioinformatics, systems biology and matrix theory.

Open Science Index, Electronics and Communication Engineering Vol:4, No:12, 2010 waset.org/Publication/14158