IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
6-8 September 2007, Dortmund, Germany

A New Incremental Face Recognition System

Youness Aliyari Ghassabeh [1], Aboozar Ghavami [2], Hamid Abrishami Moghaddam [3]
1) K. N. Toosi University of Technology, Tehran, Iran, y_aliyari@ieee.org
2) Sharif University of Technology, Tehran, Iran, aboozaer@gmail.com
3) K. N. Toosi University of Technology, Tehran, Iran, moghaddam@saba.kntu.ac.ir

Abstract— *In this paper, we present new adaptive linear discriminant analysis (LDA) algorithm and apply them for adaptive facial feature extraction. Adaptive nature of the proposed algorithm is advantageous for real world applications in which one confronts with a sequence of data such as online face recognition and mobile robotics. Application of the new algorithm on feature extraction from facial image sequences is given in three steps: i) adaptive image preprocessing, ii) adaptive dimension reduction and iii) adaptive LDA feature estimation. Steps 1 and 2 are done simultaneously and outputs of stage 2 are used as a sequence of inputs for stage3. The proposed system was tested on Yale and PIE face databases. Experimental results on these databases demonstrated the effectiveness of the proposed system for adaptive estimation of feature space for online face recognition.*

Keywords- *Incremental face recognition system, Adaptive linear discriminant analysis, Adaptive dimension reduction, Feature extraction.*

I. INTRODUCTION

Choosing an appropriate set of features is critical when designing pattern classification systems under the frame work of supervised learning. Ideally, it is desirable to use only features having high separability power while ignoring the rest. There has been an increased interest on deploying feature selection in applications such as face and gesture recognition [1]. Most effort in the literature have been focused mainly on developing feature extraction methods [2-4] and employing powerful classifiers such as probabilistic [5], hidden Markov models (HMMs) [6] neural networks (NNs) [7] and support vector machine (SVM) [8].

Feature extraction for face representation is one of the central issues to face recognition system (FR systems). FR is a high dimensional pattern recognition problem. Even low resolution images generate huge dimensional feature space (4096 dimensions in the case of 64×64 pixels face images). In addition to the problems of large computational complexity and memory usage, this high dimensionality makes it very difficult to obtain statistical models of the input space using well-defined parametric models. Among various solutions to the problem [9] the most successful seems to be those appearance-based approaches, which generally operate directly on images or appearances of face objects and process the image as two-dimensional patterns. The main trend in feature extraction has been representing the data in a lower dimensional space computed through a linear or non-linear transformation satisfying certain properties. Statistical techniques have been widely used for face recognition and in facial analysis to extract the abstract features of the face patterns. Principal component analysis (PCA) [2] and linear discriminant analysis (LDA) [3] are two main techniques used for data reduction and feature extraction in the appearance-based approaches. Eigen-faces [2] and fisher-faces [3, 10] built based on these two techniques, have been proved to be very successful. LDA algorithm selects features that are most effective for class separability while PCA selects features important for class representation. A study in [11] demonstrated that PCA might outperform LDA when the number of samples per class is small and in the case of training set with a large number of samples, the LDA still outperform the PCA. Compared to the PCA method, the computation of the LDA is much higher [12] and PCA is less sensitive to different training data sets. However, simulations reported in [13] demonstrated an improved performance using the LDA method compared to the PCA approach.

The typical implementation of these two techniques assumes that a complete dataset of training samples is given in advance, and learning is carried out in one batch. However, when we conduct PCA /LDA learning over datasets in real-world applications, we often confront difficult situations where a complete set of training samples is not given in advance. Actually in most cases such as on-line face recognition and mobile robotics, data are presented as a stream. Even if a large amount of face images are available when constructing a face recognition system, all the variations that will happen in future can not be considered in advance, thus high recognition performance in practical situations can hardly be expected with only a static data set. A solution to this problem is to make face recognition systems learn continuously to adapt to incoming training samples. This can be done by embedding an adaptive learning ability into a face recognition system. Different adaptive versions of PCA and LDA are introduced by researchers and some of them are used to construct adaptive FR systems. Hall *et al*. [14] proposed incremental PCA (IPCA) based on the updating of the covariance matrix through a residue estimation procedure. Sanger [15] derived another algorithm for adaptive principal component analysis (APCA), using generalized Hebbian learning. The same works have done for adaptive LDA, Mao and Jain [16] proposed a two layer network for LDA, each of which was a

PCA network. Chaterjee and Roychowdhurry [17] presented an adaptive algorithm and a self organizing LDA network for adaptive optimal feature extraction. Abrishami Moghaddam *et al*. [18] derived accelerated convergence adaptive algorithm for LDA based on based on steepest descent, conjugate direction and Newton-Raphson methods. Furthermore, based on adaptive algorithm mentioned above several incremental FR systems were developed [19, 20].

In this paper we introduce a new incremental FR system using the adaptive LDA algorithms introduced in [21]. We use these algorithms to construct an incremental FR (IFR) system for on-line face recognition applications. . The new IFR system consists of three independent steps: i) adaptive image preprocessing, ii) adaptive dimension reduction and iii) adaptive LDA feature estimation. All steps in this system use a sequence or stream of data for training stage; therefore the need to keep a large amount of sample data for training phase is reduced. Memory size and complexity reduction provided by the new adaptive feature extraction algorithms make it appropriate for different real time pattern recognition applications. Simulation results approved ability of proposed IFR system for effective LDA feature extraction and face classification during on-line training.

II. FEATURE EXTRACTION METHODS

LDA searches the directions for maximum discrimination of classes in addition to dimensionality reduction. To achieve this goal, within-class and between-class matrices are defined [22]. In LDA, the optimum linear transform is composed of $p(\leq n)$ eigenvectors of $\mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}$. The computation of the eigenvector matrix $\mathbf{W}_{LDA}$ of $\mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}$ is equivalent to the solution of the generalized eigenvalue problem $\mathbf{\Sigma}\mathbf{W}_{LDA} = \mathbf{\Sigma}_{\mathbf{w}}\mathbf{W}_{LDA}\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is the generalized eigenvalue matrix. Under assumption of positive definite matrix $\mathbf{\Sigma}_{\mathbf{W}}$, there exists a symmetric $\mathbf{\Sigma}_w^{-1/2}$ such that the problem can be reduced to a symmetric eigenvalue problem:

$$\mathbf{\Sigma}_{\mathbf{w}}^{-1/2}\mathbf{\Sigma}\mathbf{\Sigma}_{\mathbf{w}}^{-1/2}\mathbf{\Psi} = \mathbf{\Psi}\mathbf{\Lambda} \tag{1}$$

where $\mathbf{\Psi} = \mathbf{\Sigma}_w^{1/2}\mathbf{W}_{LDA}$. PCA algorithm can be used to find a subspace whose basis vectors correspond to the maximum variance directions in the original n dimensional space. PCA transfer function is composed of significant eigenvectors of covariance matrix [22]. The following equation can be used for incremental estimation of covariance matrix:

$$\mathbf{\Sigma}_k = \mathbf{\Sigma}_{k-1} + \eta_k(\mathbf{x}_k\mathbf{x}_k^t - \mathbf{\Sigma}_{k-1}) \tag{2}$$

where $\mathbf{\Sigma}_k$ is estimation of the covariance matrix at k-th iteration, $\mathbf{x}_k$ is the incoming input vector and η_k is the learning rate. Author in [15] proved that the following equation will converge to a matrix, whose rows are eigenvectors of covariance matrix:

$$\mathbf{T}_{k+1} = \mathbf{T}_k + \gamma_k(\mathbf{T}_k\mathbf{x}_k\mathbf{x}_k^t - LT[\mathbf{T}_k\mathbf{x}_k\mathbf{x}_k^t\mathbf{T}_k^t]\mathbf{T}_k) \tag{3}$$

where LT[.] makes its input matrix lower triangular, $\mathbf{x}_k$ is input vector at k-*th* iteration and rows of $\mathbf{T}_k$ converge to eigenvectors of covariance matrix ordered by decreasing eigenvalues. There are different adaptive estimations of the mean vector. The following equation has been used in [17, 18]:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \eta_k(\mathbf{x}_{k+1} - \mathbf{m}_k) \tag{4}$$

where η_k is learning rate.

III. INCREMENTAL LDA FEATURE EXTRACTION

We use two adaptive training algorithms in cascade for extracting optimal LDA features. The first algorithm called $\mathbf{\Sigma}^{-1/2}$ algorithm is for the computation of the square root of inverse covariance matrix. The second algorithm is an APCA (eq. 3) as introduced by [15]. We prove the convergence of the cascade architecture as an ALDA for feature selection.

We define the cost function $J(\mathbf{w})$ with parameter $\mathbf{w}$, $J:\ \Re^{n\times n} \to \Re$ as follows:

$$J(\mathbf{W}) = \frac{tr(\mathbf{W}^3\mathbf{\Sigma})}{3} - tr(\mathbf{W}) \tag{5}$$

Where $\mathbf{\Sigma}$ is the covariance matrix and $\mathbf{w}$ is a symmetric, positive definite matrix. It is not difficult to prove that $\mathbf{\Sigma}^{-1/2}$ is a critical point (matrix) of the cost function given in (5) and Hessian matrix of (5) by substituting $\mathbf{w}$ with $\mathbf{\Sigma}^{-1/2}$ will become a positive definite matrix [21]. Therefore cost function (5) has a minimum that occurs in $\mathbf{\Sigma}^{-1/2}$ [23]. Using the gradient descent optimization method [24] and by substituting $\mathbf{\Sigma}$ with $\mathbf{x}_{k+1}\mathbf{x}_{k+1}^t$, we get the following equation for adaptive computation of the $\mathbf{\Sigma}^{-1/2}$:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \eta_k(\mathbf{I} - (\mathbf{W}_k^2\mathbf{x}_{k+1}\mathbf{x}_{k+1}^t + \mathbf{x}_{k+1}\mathbf{x}_{k+1}^t\mathbf{W}_k^2 + \mathbf{W}_k\mathbf{x}_{k+1}\mathbf{x}_{k+1}^t\mathbf{W}_k)/3)) \tag{6}$$

It is quite easy to prove that if initial condition $\mathbf{W}_0$ meets the $\mathbf{W}_0\mathbf{\Sigma} = \mathbf{\Sigma}\mathbf{W}_0$ (for example $\mathbf{W}_0$ chosen equal to identity matrix), then (6) will be simplified to the following equation [21]:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \eta_k(\mathbf{I} - \mathbf{W}_k\mathbf{x}_{k+1}\mathbf{x}_{k+1}^t\mathbf{W}_k) \tag{7}$$

It is clear that (7) has less computational cost compared with (6), and can be used for incremental estimation of the $\mathbf{\Sigma}^{-1/2}$ instead of equation (6). As mentioned in section II, the LDA features are significant eigenvectors of $\mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}$. For adaptive computation of them, we combine two algorithms (Eq. (3) and Eq. (7)) in cascade and show that this architecture asymptotically computes LDA features. Let $\mathbf{m}_k^i$ denote the estimated mean vector of class $i(i = 1,2,...,L)$ at *k-th* iteration and $\omega(\mathbf{x}_k)$ denote the class of $\mathbf{x}_k$. The training sequence $\{\mathbf{y}_k\}$ for $\mathbf{\Sigma}^{-1/2}$ algorithm is defined by

$\mathbf{y}_k = \mathbf{x}_k - \mathbf{m}_k^{\omega(\mathbf{x}_k)}$. With the arrival of every training sample $\mathbf{x}_k$, $\mathbf{m}_k^i$ is updated according to its class using (4). It is easy to show that the correlation of the sequence $\{\mathbf{y}_k\}$ is the within-class scatter matrix $\mathbf{\Sigma}_W$. Therefore, we have the following equation:

$$\lim_{k\to\infty} E[(\mathbf{x}_k - \mathbf{m}_k^{\omega(x_k)})(\mathbf{x}_k - \mathbf{m}_k^{\omega(x_k)})^t] = \lim_{k\to\infty} E[\mathbf{y}_k \mathbf{y}_k^t] = \mathbf{\Sigma}_W \quad (8)$$

Suppose the sequence $\{\mathbf{z}_k\}$ is defined by, $\mathbf{z}_k = \mathbf{x}_k - \mathbf{m}_k$. Where $\mathbf{m}_k$ is the estimated mixture mean value in *k-th* iteration. We train the $\mathbf{\Sigma}^{-1/2}$ algorithm by the sequence $\{\mathbf{y}_k\}$ and use $\mathbf{W}_k$ in (7) to create the new sequence $\{\mathbf{u}_k\}$ as follows, $\mathbf{u}_k = \mathbf{W}_k \mathbf{z}_k$. The sequence $\{\mathbf{u}_k\}$ is used to train the algorithm (3). As mentioned before, the matrix $\mathbf{T}_k$ in the algorithm (3) converges to the eigenvectors of the covariance matrix of the input vectors, ordered by decreasing eigenvalues. Hence, (3) will converge to the eigenvectors of $E(\mathbf{u}_k \mathbf{u}_k^t)$. It is quite easy to show:

$$\lim_{k\to\infty} E(\mathbf{u}_k \mathbf{u}_k^t) = \mathbf{\Sigma}_W^{-1/2} \mathbf{\Sigma} \mathbf{\Sigma}_W^{-1/2} \quad (9)$$

Our aim is to estimate the eigenvectors of $\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}$. Suppose $\mathbf{\Phi}$ and $\mathbf{\Lambda}$ denote the eigenvector and eigenvalue matrices corresponding to $\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}$. Following equations are held [22]:

$$\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{\Sigma}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda} \quad (10)$$

$$\mathbf{\Sigma}_W^{-1/2}\mathbf{\Sigma}\mathbf{\Sigma}_W^{-1/2}\mathbf{\Psi} = \mathbf{\Psi}\mathbf{\Lambda} \quad (11)$$

where $\mathbf{\Psi} = \mathbf{\Sigma}_\mathbf{w}^{1/2}\mathbf{\Phi}$. From (10, 11), it is concluded that the eigenvector matrix of $\mathbf{\Sigma}_W^{-1/2}\mathbf{\Sigma}\mathbf{\Sigma}_W^{-1/2}$ is equal to $\mathbf{\Psi}$. In the other words, the matrix $\mathbf{T}^t$ in the second algorithm converges to $\mathbf{\Psi}$ and the following equation is held:

$$\lim_{k\to\infty} \mathbf{T}_k^t = \Psi = \mathbf{\Sigma}_W^{1/2}\mathbf{\Phi}^t \quad (12)$$

By multiplying the outputs of the first and second algorithms as $k \to \infty$, we will have:

$$\lim_{k\to\infty} \mathbf{W}_k \mathbf{T}_k^t = \mathbf{\Sigma}_W^{-1/2}\mathbf{\Sigma}_W^{1/2}\mathbf{\Phi}^t = \mathbf{\Phi}^t \quad (13)$$

Therefore, the combination of the first and second algorithms will converge to the desired matrix $\mathbf{\Phi}$, whose columns are eigenvectors of $\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}$. As described in the previous sections, by choosing a $p \times n$ random matrix as the initial value of (3), the final result of (7) in cascade with (3) will converge to a $p \times n$ matrix, composed of p significant eigenvectors of the $\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}$ ordered by decreasing eigenvalues. By the definition given in section II, these eigenvectors are used as the LDA features.

IV. New Incremental Face Recognition System

In this section, we introduce a new IFR system constructed based on the proposed adaptive algorithm. This system included three parts: *i)* adaptive preprocessing, *ii)* adaptive dimension reduction, *iii)* adaptive LDA feature estimation.

A. Adaptive Preprocessing

In this step, every input image is histogram equalized to the range of values from 0 to 255 to spread the energy of all intensity pixel values in the image. In addition, we normalized all of input images to make them have the same energy. Then we used (4) to estimate the mean image and subtracted it from every input image. In all experiments described in this paper, we cropped every input image to a 40×40 image; therefore final vectorized image has a size of 1600×1. Figure 1, demonstrates this process in a diagram.

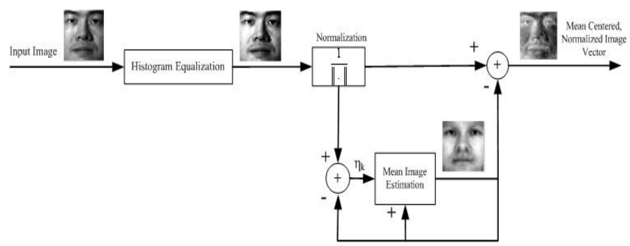

Fig. 1. Adaptive preprocessing block diagram.

B. Adaptive Dimension Reduction

As mentioned in the previous section, we cropped every input image to 40×40 image; as a result the input of this stage is a preprocessed 1600×1 vector. We used (2) to estimate the covariance matrix. Then, we computed 60 significant eigenvectors of the estimated covariance matrix. In this stage, we will have 60 eigenfaces, by projection of every input image on these eigenafeces, they will converted to reduced size 60 × 1 vectors. Figure 2 demonstrated dimension reduction stage in a block diagram.

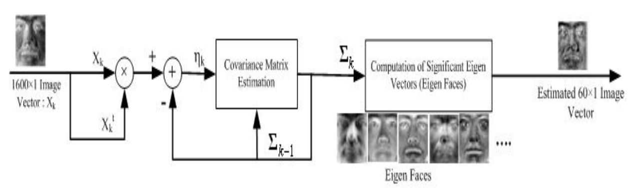

Fig. 2. Adaptive dimension reduction block diagram.

C. Incremental LDA Feature Estimation

In this part, we use the new adaptive algorithm presented in (7) for estimation of the square root of inverse within scatter matrix. Estimated $\mathbf{\Sigma}_W^{-1/2}$ is combined with (3). As proved in the previous section, combination of (7) with (3) will converge to $\boldsymbol{\Sigma}_w^{1/2}\boldsymbol{\Phi}$. If we multiply it with estimated $\boldsymbol{\Sigma}_w^{-1/2}$, the result will be the estimated LDA features. Figure 3 illustrates this process where algorithms in (7) and (3) work simultaneously. As illustrated, the adaptive nature of this structure makes it appropriate for adaptive LDA feature extraction. At the end of this part, every input image is projected on the estimated LDA features to achieve high separability in addition to dimension reduction.

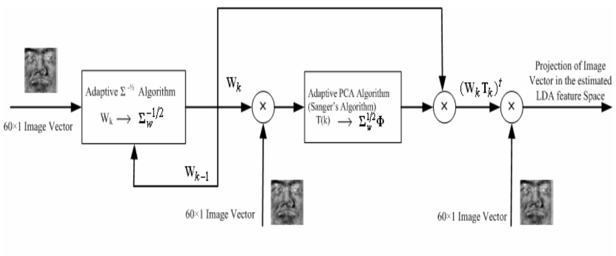

Fig. 3. Block diagram for adaptive linear discriminant analysis.

Fig. 4. 10 samples images from each subject from PIE database

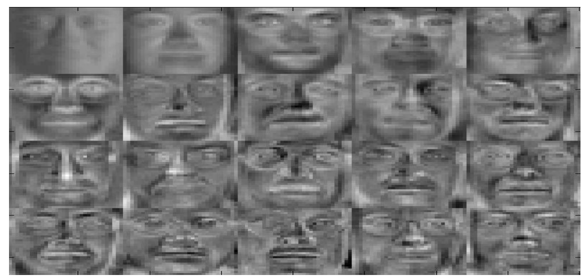

Fig. 5. Estimated 20 most significant eifgenfaces fromPIE database.

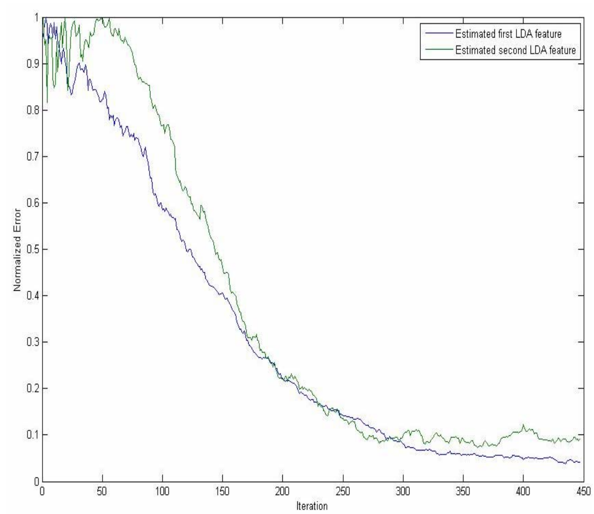

Fig. 6. Convergence of the first and second significant LDA features using proposed IFR system.

V. Experimental Results

We applied the proposed new adaptive FR system on YALE and PIE face datasets for online face classification. Before doing any experiment, we cropped the input images to reduce their size to 40×40. In each experiment described in this section, we considered input images as a sequence of random data and simultaneously trained the proposed FR system.

A. Experiment on PIE Database

This database contains images of 68 people under different poses and illuminations with 4 different expressions. In this experiment, we chose 3 random subjects and for each subject 150 images are considered. Figure 4 shows some images from each subject in different posses. We trained our proposed IFR system with these images. (There is total of 450 images). Figure 5 shows first 20 estimated eigenfaces produced from stage 2 of our proposed system. In this experiment, there are only two significant LDA features because we considered 3 subjects. We applied the training images into our algorithm and estimated these two significant features. Figure 6 shows the normalized error between two significant real LDA features (computed using scatter matrices) and estimated ones during the training process. It is clear that after 450 iterations estimated LDA features nearly converge to their real vales. Figure 7 demonstrated the distribution of the face images in two dimensional feature spaces during the training stage. The top left figure, demonstrated the distribution after 100 iteration and other figures demonstrated the distribution of face images after 200, 300 and 450 iteration in the feature space. It can be observed form figure 7 that although overlapping, three subjects are become linearly separable after the online training stage.

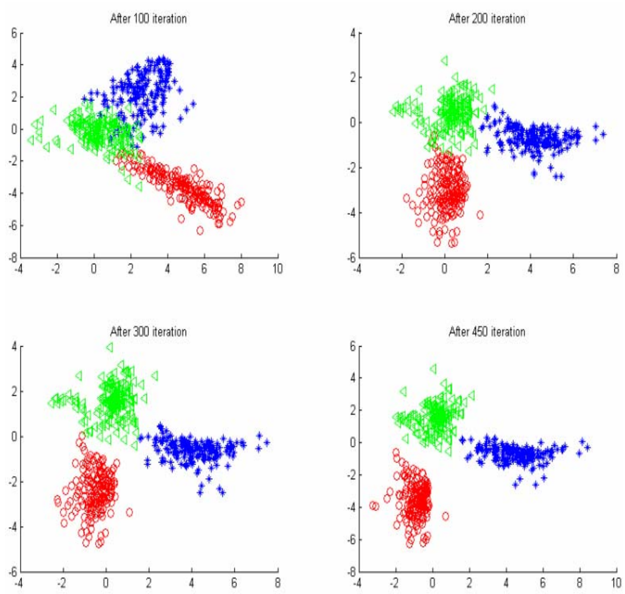

Fig. 7 distribution face images in the estimated LDA feature space, after 100, 200, 300 and 450 iterations.

B. Experiment on Yale Database

This database contains grayscale images of 15 subjects in GIF format. In these experiments, we chose 5 individual subjects and considered 64 images per each subject (total 320 images) containing different illumination and different poses. Figure 8 shows some of selected subjects in different position and illumination. We trained the proposed FR system using these images as a stream. Figure 9 shows first 20 eigenfaces estimated in stage 2 of the proposed IFR system. In this experiment, we considered 5 subjects; therefore there exist 4 LDA features.

Fig. 8. Sample images from 5 subjects in different illumination and posses.

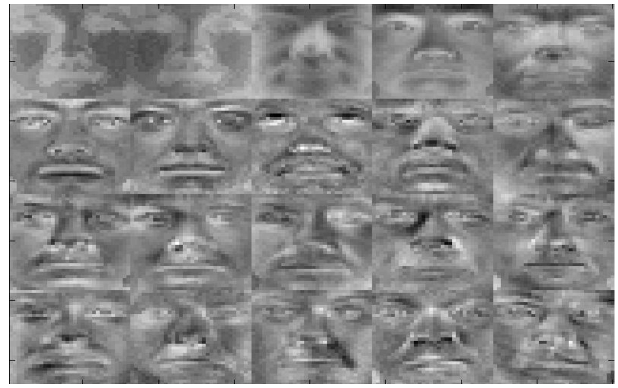

Fig. 9. Estimated 20 significant eigenface as output of stage II (Yale database).

Using the introduced IFR system, we estimated the three significant LDA features and by projection of face images onto them, we reduced the data dimension into three. Figure 10 shows convergence of three estimated LDA features used for dimension reduction. In this figure, we computed the normalized error between estimated LDA feature and real LDA feature (computed using scatter matrices) in each iterations. Figure 11 shows distribution of face images in the estimated three dimensional feature space during the training phase. The top left figure shows distribution after 80 iteration and other three figures demonstrated the distribution of subject images in feature space after 120, 220 and 320 iteration, respectively. it is clear from figure 8 that face images at first iterations are not clearly separable but gradually by adaptive training of IFR system, each subjects separate from others and at the end of process (after 320 iteration) all of the subjects are linearly separable (although overlapping) in three dimensional estimated feature space.

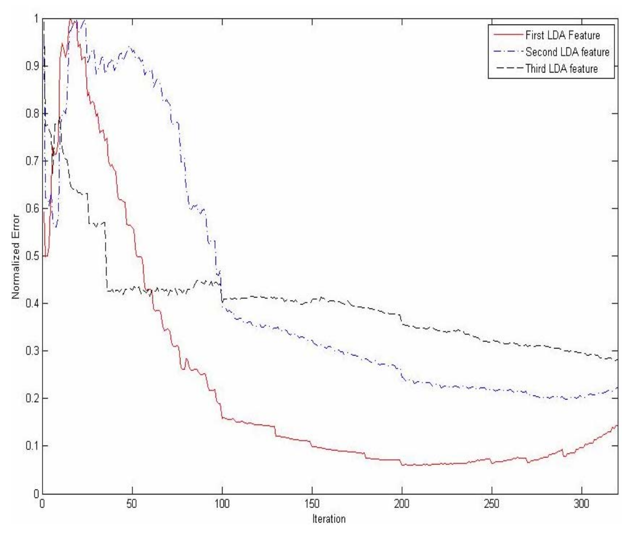

Fig. 10. Convergence of the LDA significant features during the training process using proposed IFR system for YALE datasbe.

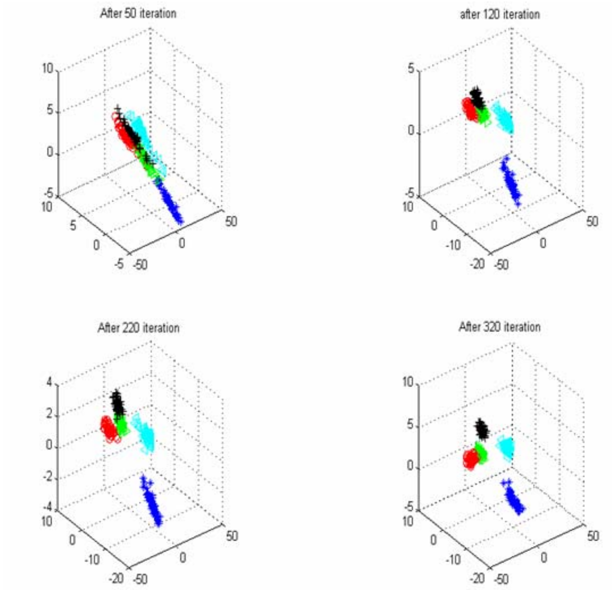

Fig. 11 Distribution of subject images in the estimated three dimensional feature space, after 50, 120, 220 and 320.

VI. Concluding Remarks

In this paper, a new IFR system based on the new adaptive LDA feature extraction algorithm was presented. The new IFR system was considered as a combination of a new adaptive $\Sigma^{-1/2}$ algorithm in cascade with APCA. Convergence of the new adaptive LDA algorithm was proved. We use this algorithm to construct an IFR system for on-line face recognition applications. . The new IFR system consists of three independent steps: i) adaptive image preprocessing, ii) adaptive dimension reduction and iii) adaptive LDA feature estimation Simulation results for LDA feature extraction using YALE and PIE face datasets demonstrated the ability of the proposed algorithm for adaptive optimal feature extraction. The new adaptive algorithm can be used in many fields of online machine learning and pattern recognition applications such as face and gesture recognition and mobile robotics.

References

[1] Z. sun, G. Bebis, R. miller, 2004, object detection using feature selection. Pattern recognition, Elsevier Vol. 37, no 11, pp 2165-3176

[2] M. Turk, A. Pentland, "Eigenfaces for face recognition", journal cognitive neuroscience, Vol. 3, No.1 , 1991.pp 71-8

[3] P. Belhumeur, J. Hespanha, D. Kregman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection".IEEE Trans. Pattern Anal. Machine Intell.,Vol 19, pp711-720,1997.

[4] O.Deniz, M. Castrill_on, M. Hern_andez, " Face recognition using independent component analysis and support vector machines" , Pattern Recognition letters, Vol. 24, pp. 2153-2157, 2003.

[5] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition", IEEE Trans. pattern Anal. Machine Intell., Vol. 24, No. 6, pp. 780-788, 2002.

[6] H. Othman, T. Aboulnasr, " A separable low complexity 2D HMM with application to face recognition" IEEE Trans. Pattern. Anal. Machie Inell., Vol. 25, No. 10, pp. 1229-1238, 2003.

[7] M. Er, S. Wu, J. Lu, L.H.Toh, "face recognition with radial basis function(RBF) neural networks", IEEE Trans. Neural Networks, Vol. 13, No. 3, pp. 697-710.

[8] K. Lee, Y. Chung, H. Byun, "SVM based face verification with feature set of small size", electronic letters, Vol. 38, No. 15, pp. 787-789, 2002.

[9] J. R. Solar, P. Navarreto, " Eigenspace-based face recognition: a comparative study of different approaches, IEEE Tran. , Systems man And Cybernetics- part c:Applications, Vol. 35, No. 3, 2005.

[10] J. Lu, K. N. Plataniotis and A.N. Venetsanopoulos," Face recognition using LDA –based algorithms", IEEE Trans. Neural Networks, Vol. 14, No. 1, 2003.

[11] A. M. Martinez, A. C. Kak, "PCA versus LDA", IEEE Trans. Pattern Anal. Machine Intell, Vol. 23, pp. 228-233.

[12] W, Zhao, R. Chellappa, A. Krishnaswamy, "Discriminant analysis of principal components for face recognition", IEEE int. Conf. Automatic face and gesture recognition. Pp. 336-341, 1998.

[13] J.J. Weng, "using discriminant eigenfeatures for image retrieval", IEEE Trans. Pattern Anal. Machine Intell. , Vol. 18, No. 8, pp. 831-836, 1996.

[14] P. Hall, and R. Martin, "Incremental eigen analysis for classification" in proc. Brit. Machine Vision Conf. Vol. 1, pp 286-295.

[15] T.D. Sanger,"optimal unsupervised learning in a single-layer linear feed forward neural network", Neural Networks, Vol. 2, pp. 459-473, 1989.

[16] J. Mao and A. K. Jain, "Discriminant analysis neural networks", In IEEE Int. Conf. on Neural Networks, CA, pp. 300-305, March 1993.

[17] C. Chatterjee, V. P. Roychowdhurry, "On self-organizing algorithm and networks for class separability features", IEEE Trans. Neural Network, Vol. 8, No. 3, pp. 663-678, 1997.

[18] H.Abrishami Moghaddam, M.Matinfar, S.M. Sajad Sadough, Kh. Amiri Zadeh, "Algorithms and networks for accelerated convergence of adaptive LDA", Pattern Recog , Vol. 38, No. 4, pp. 473-483, 2005.

[19] O. Deniz, M. Castrllon, J.Lorenzo, N. Hernandez, "an incremental learning algorithm for face recognition, in M. Tistarelli, J. Bigun, A.K.Jain(Eds), Biometric Authentication ,pp. 1-9, Speringer-Verlog, 2002.

[20] S. Ozawa, S.L. Toh, S. Abe, S. Pang, N. Kasabov, "Incremental learning of feature space and classifier for face recognition", Neural Networks, Vol. 18, pp. 575-584, 2005.

[21] Y. Aliyari Ghassabeh H. Abrishami Moghaddam, "A New Incremental Optimal Feature Extraction Method for On-line Applications ", 4th International Conference on Image Analysis and Recognition (ICIAR 2007), Canada, 2007.

[22] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, *2nd* Edition, Academic Press, New York, 1990.

[23] J.R. Magnus, H. Neudecker, *Matrix Differential Calculus*, John Wiley, 1999.

[24] B.Widrow, S. Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.