International Journal of Innovative Computing, Information and Control
Volume **8**, Number **1(A)**, January **2012**

ICIC International ©2012 ISSN 1349-4198
pp. **219–232**

AN OPTIMIZED FACE RECOGNITION SYSTEM USING STABLE ORTHOGONALIZATION

Imtiaz Ahmed Sajid, Mohammad Mansoor Ahmed and Mohammad Sagheer

Department of Electronic Engineering
Mohammad Ali Jinnah University
Islamabad Expressway, Kahuta Road, Zone-V, Islamabad, Pakistan
{ imtiazsajid; mansoor; sagheer }@jinnah.edu.pk

Received June 2010; revised October 2010

Abstract. *This article presents a new architecture to implement an optimized face recognition system based on reduced dimensionality of covariance matrix. The leading components of reduced matrix are computed by using modified Gram-Schmidt orthogonalization (MGSO). Use of MGSO in fast principle component analysis has improved the convergence and accuracy of a face recognition system especially for high dimensional images. To optimize the performance of a face recognition system, receiver operating characteristics are simulated with least square fitting. Convex optimization is applied for optimum selection of system variables. It is demonstrated that the proposed technique, compared with decomposition, provides better discriminating power in Eigen space.*

Keywords: Fast principal component analysis (PCA), Face recognition, Modified Gram-Schmidt orthogonalization, Eigen space

1. **Introduction.** Principal component analysis (PCA) is one of the holistic algorithms which has been used in many biometric applications, like face recognition and object identification [1-9]. Recently, serially-connected dual 2D PCA has been used for efficient face recognition [10]. PCA may be used in fisher or linear discriminators [11], whereas its prime focus is to evaluate Eigen values in low-dimensional space commonly known as subspace. While doing so it tries to maintain original sample space characteristics [12].

PCA evaluates Eigen values by employing decomposition method. However, decomposition method requires more time than orthogonalization [13]. Fast PCA (FPCA) uses Gram-Schmidt orthogonalization (GSO) method to evaluate Eigen values. It is worth mentioning that GSO based Eigen values calculation technique uses orthogonalization in low-dimensional space which is time-efficient, because it reduces complexity of decomposition from $O(n^3)$ to $O(n^2)$ [13-15].

It has been shown by many researchers that the use of orthogonalization for the extraction of feature vectors in face space produces better results compared with decomposition [16-18] because it has better discriminating power. However, orthogonalization in GSO is sensitive towards rounding effects due to finite precision of a machine, which may cause instability in FPCA [13-15,19].

It has been demonstrated by numerous numerical experiments that GSO may produce a set of non-orthogonal vectors which may lead to a non-convergence state of FPCA [20-22]. In modified GSO (MGSO) vectors are normalized prior to the orthogonalization process which minimizes the generation of non-orthogonal vectors. This property of MGSO could be used in FPCA to overcome its instability for the evaluation of Eigen values.

For a time-efficient decision support system, the selection of leading Eigen vectors (LEVs), which are generated by MGSO, is critical. FPCA has the capability to compute efficiently the desired number of LEVs, whereas decomposition continues till the diagonal of a decomposed matrix is filled. Proper selection of LEVs plays an important role in maximizing the variance [2,12,23]. It is reported that the variance defined by three LEVs, usually occupies more than 90% of the total space [24]. Therefore, further inclusion of LEVs may not give considerable improvement in a face recognition system. However, the selection of optimum number of LEVs is not a trivial decision [25] and is dependent upon sample space and system reliability.

It is an established fact that the first LEV occupies maximum portion of variance and the second LEV shares the second highest and so on [25]. In general, the recognition accuracy in Eigen space is dependent upon the number of LEVs used in reduced subspace [23]. However, the system performance starts saturating after a certain number of LEVs [1,2,12]. On the other hand, recognition/decision time of a system increases by increasing the number of LEVs. Thus, improvement in a system accuracy, within the permitted decision time, requires an optimum selection of LEVs.

The accuracy of a face recognition system, apart from the optimum number of LEVs, also depends upon threshold value used to generate receiver operating characteristic (ROC) curves. ROC curve is a graphical representation of false acceptance rate (FAR) and false rejection rate (FRR), which is used commonly to assess the performance of a face recognition system [3,26,27]. To minimize the system error and to enhance its accuracy, it is imperative to know minimum error value on ROC for a given number of LEVs.

In this article, FPCA has been modified by incorporating MGSO to improve its stability and accuracy for a face recognition system. A technique has been developed to evaluate the optimum number of LEVs by using modified FPCA. Further, a mechanism is suggested to achieve the best combination of system variables for its improved reliability and accuracy.

2. **Architecture.**

Architecture of the proposed system illustrates storage and flow of data amongst various components as shown in Figure 1. Architecture is divided into three layers as shown in Figure 1. Covariance matrix and the desired number of LEVs are provided as inputs to MGSO module in the learning layer of the architecture to evaluate Eigen values. Eigen faces are then computed after extracting the feature vectors (Eigen vectors) which are then stored in a database (DB).

In optimization layer of Figure 1, an adaptive classifier provides an initial p to the threshold balancer on the basis of Eigen faces and sample test images. After getting a new p from the threshold balancer, the adaptive classifier defines FAR, FRR, decision time and corresponding number of LEVs which are subsequently stored in a database for simulation purposes. The sum of FAR and FRR (S_p) is plotted against p and the same is simulated by the least square fitting module to generate convex functions. Whereas, the minimizer provides minimum error by applying line search algorithm (LSA) on the generated convex functions.

Maximum allowable time for recognition and/or minimum acceptable accuracy of a system is referred to as application constraints. On the basis of these constrains and by using a correction factor, the optimizer of Figure 1 provides best available h and p for a given application.

Finally, in the testing layer on the basis of optimized h and p, features vectors are extracted and stored in DB. On the other hand, feature vectors of an input image is also extracted and compared with stored feature vectors. This enables the system to accept/reject an input image.

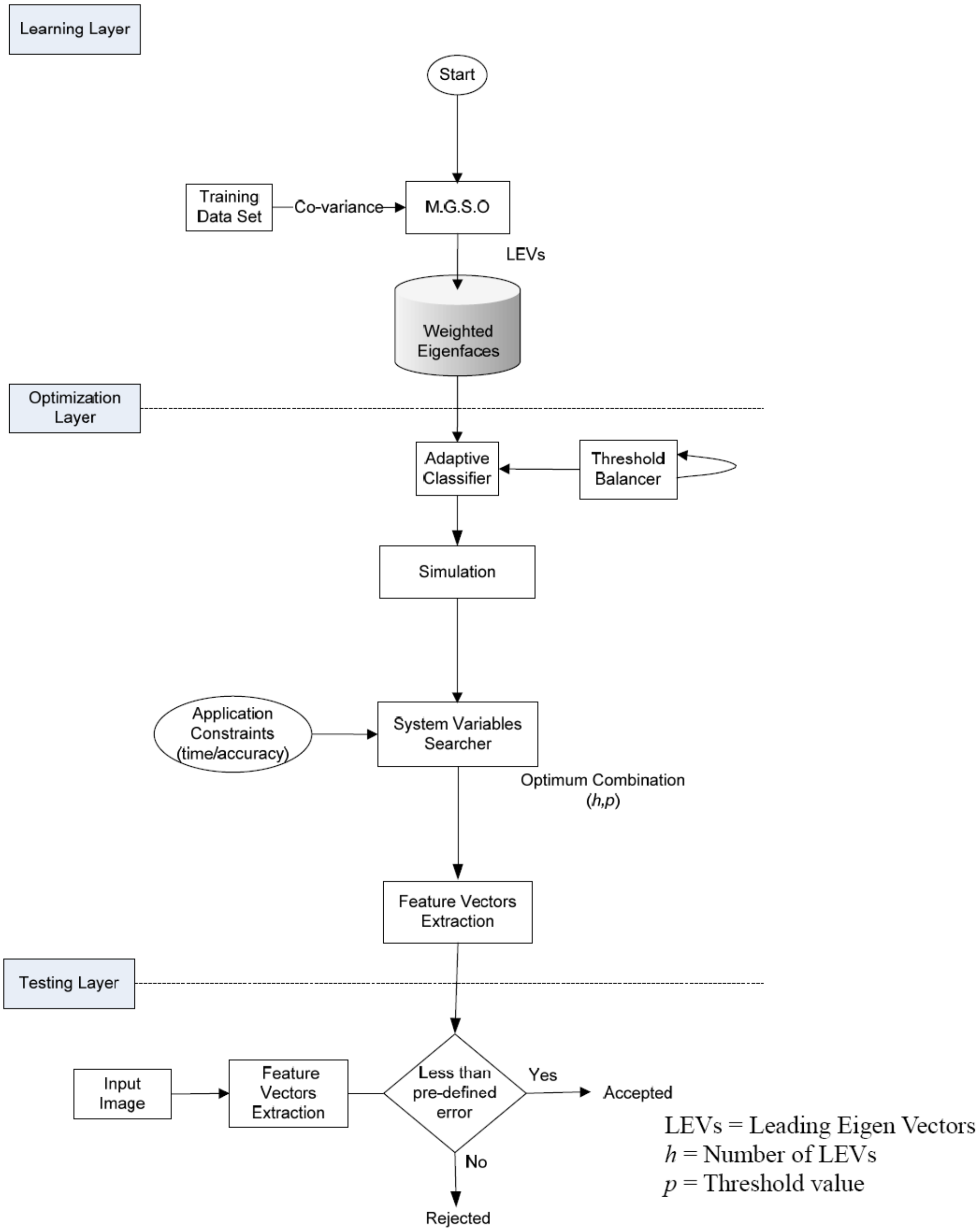

FIGURE 1. A three layer architecture for optimum combination of decision time, minimum error rate and leading eigenvectors

3. **Statistical View of the Proposed System.** Let there be d training images, which are represented by the matrices $I_1, I_2, \ldots, I_d$ of order $r \times c$ each. Variation in illumination degrades face recognition performance [2,28]. To mild the lighting effects $I_1, I_2, \ldots, I_d$ matrices are preprocessed using $I'_i = \dfrac{(I_i - m_i) \times ds_i}{s_i} + dm_i$, where $i = 1, 2, \ldots, d$, m_i is the mean, s_i is the standard deviation, dm_i is the desired mean and ds_i is the desired standard deviation [29]. After preprocessing, matrices are converted into column vectors $J_1, J_2, \ldots, J_d$ each of length rc. Let $\overline{J} = \dfrac{1}{d}\sum_{i=1}^{d} J_i$ and $A_{rc\times d} = [A_1 \quad A_2 \quad A_3 \quad \ldots \quad A_d]$, where $A_i = J_i - \overline{J}$. Then, the covariance matrix is $C_{rc\times rc} = AA^T$. To find the significant

Eigen vectors of C, $L_{d\times d}$ is constructed as: $L = A^T A$ where $d \ll rc$ [12]. Suppose $\lambda_1, \lambda_2, \ldots, \lambda_d$ are the Eigen values (in descending order) of L corresponding to the Eigen vectors $u_1, u_2, \ldots, u_d$, respectively. The LEVs (h) are then computed by employing the following algorithm:

3.1. **Algorithm for computing the leading Eigen vectors.** In this algorithm, the iteration counter is denoted by IC, the maximum number of allowed iterations is represented by M and the number of non-zero columns of a matrix U by $\beta(U)$.

Step (i) I*nitialize a zero matrix U of the order of $d \times h$. Set $IC = 0$, assign a positive integral value to M. Choose a sufficiently small tolerance $\varepsilon > 0$, e.g., (10^{-2}) and $h < d$.*

Step (ii) Update U as suggested in Step (viii) (ignore this step while calculating the 1st leading Eigen vector).

Step (iii) Randomly select a vector ψ_o of size d.

Step (iv) Normalization of ψ_o is computed as $\psi_1 = \dfrac{\psi_o}{\|\psi_o\|}$, where $\|\psi_o\|$ is Euclidian norm of a vector.

Step (v) Compute $\psi_2 = L\psi_1$.

Step (vi) For calculating the first leading Eigen vector, set $\psi_3 = \psi_2$, otherwise, $\psi_3 = \psi_2 - \left(\sum_{j=1}^{\beta(U)} \left(\psi_2^T u_j\right) u_j\right)$, where u_j is the j^{th} column of U.

Step (vii) Normalization of ψ_2 and ψ_3 are computed as $\psi_4 = \dfrac{\psi_3}{\|\psi_3\|}$ and $\psi_2 = \dfrac{\psi_2}{\|\psi_2\|}$, where $\|\psi_2\|$ and $\|\psi_3\|$ are Euclidian norms of ψ_2 and ψ_3, respectively.

Step (viii) (a) If $\left|\psi_4^T \psi_2 - 1\right| < \varepsilon$, go back to Step (ii) and replace the 1st zero column of U by ψ_4. If there is a zero column in U repeat Step (iii) to Step (viii) with a different choice of ψ_o in Step (iii), otherwise, go to Step (x).

(a) If $\left|\psi_4^T \psi_2 - 1\right| \geq \varepsilon$, go to Step (ix).

Step (ix) Update IC by $IC = IC + 1$, if $IC < M$, then repeat Step (v) to Step (viii) with $\psi_1 = \psi_4$ in Step (v), otherwise, go back to Step (iii).

Step (x) Compute Eigen values of L by $U^T LU$.

The matrix U is computed by using the above algorithm containing h LEVs, which is used to populate u_j vectors as discussed in [12]. Assuming Au_j are the Eigen vectors of a matrix C. Now define

$$t_j = \frac{Au_j}{\|Au_j\|} \text{ for } j = 1, 2, 3, \ldots, h \text{ and } T_{rcxh} = [t_1\, t_2\, \ldots\, t_h]\,.$$

Then, collect the projection of each of the h leading Eigen vectors on A_i in a matrix $W_{d\times h}$ as:

$$W(i, j) = A_i^T t_j; \quad i = 1, 2, 3, \ldots, d \text{ and } j = 1, 2, 3, \ldots, h.$$

Let N be a column vector of length rc, representing an input image. Define three column vectors D_N, W_N, V_N and a number r as:

$$D_N = N - \overline{J}; \tag{1}$$

$$W_N(j) = (D_N)^T t_j; \tag{2}$$

$$V_N(i) = \left\|W_N - \left(i^{\text{th}} \text{ column of } W^T\right)\right\|; \tag{3}$$

$$r = \max(V_N) - \min(V_N). \tag{4}$$

Suppose there are k input images $N_1, N_2, \ldots, N_k$ and half of which are unknown. Each image is represented by a column vector of length rc. Repeat (1) to (4) for each N_1, N_2,

..., N_k image to get $r_1, r_2, \ldots, r_k$ and let $R = [r_1 \; r_2 \; \ldots \; r_k]$. Calculate the mean of R and denote it by Z.

$$\text{Define } Z_p = pZ, \text{ where } p = 0, h', 2h', \ldots, 2 \text{ for some } h' > 0. \tag{5}$$

For each p, define a row matrix H_P of length k, by

$$H_p(i) = \begin{cases} 1 & \text{if } r_i < Z_p \\ 0 & \text{otherwise} \end{cases}$$

where $i = 1, 2, 3, \ldots, k$.

Associated with each p, define two numbers F_{RP} and F_{AP} for any of the following two cases:

Case-I: When k is even,

$$F_{RP} = \text{Number of zeros in } Hp(i); \quad i = 1, 2, \ldots, \frac{k}{2}.$$
$$F_{AP} = \text{Number of ones in } Hp(i); \quad i = \frac{k}{2} + 1, \frac{k}{2} + 2, \ldots, k.$$

Case-II: When k is odd,

$$F_{RP} = \text{Number of zeros in } Hp(i); \quad i = 1, 2, \ldots, \frac{k-1}{2}.$$
$$F_{AP} = \text{Number of ones in } Hp(i); \quad i = \frac{k+1}{2}, \frac{k+1}{2} + 1, \ldots, k.$$

For each p, define a false rejection rate (FRR_p) and a false acceptance rate (FAR_p) as follows:

$$FRR_p = \begin{cases} \dfrac{2F_{RP}}{k} \times 100, & \text{if } k \text{ is even} \\ \dfrac{2F_{RP}}{(k-1)} \times 100, & \text{if } k \text{ is odd} \end{cases}$$
$$FAR_p = \begin{cases} \dfrac{2F_{AP}}{k} \times 100, & \text{if } k \text{ is even} \\ \dfrac{2F_{AP}}{(k+1)} \times 100, & \text{if } k \text{ is odd} \end{cases} \tag{6}$$

Denote the sum of the above two by S_p, i.e.,

$$S_p = FRR_p + FAR_p. \tag{7}$$

For obvious reasons, accuracy of results depends upon h and p. We will now implement our scheme for different choices of h and p to select their optimum combination to achieve the required accuracy for a given time.

Let us assign different values to $h_i = h_1, h_2, \ldots, h_q$. To achieve global minima, S_p has been simulated by using least square method. Denote the minimum value of Q_i by: $\min(Q_i)$; $i = 1, 2, \ldots, q$ and the value of p for which Q_i attains its minimum value by p_i; $i = 1, 2, \ldots, q$. Suppose τ_i; $i = 1, 2, \ldots, q$, is a time taken for decision making when h takes values h_i; $i = 1, 2, \ldots, q$. Now write a matrix $F_{q \times 4}$ as follows:

$$F = [H \; P \; \min(Q) \; \tau] \tag{8}$$

where

$$H = [h_1 \, h_2 \, \ldots \, h_q]^T, \quad P = [p_1 \, p_2 \, \ldots \, p_q]^T, \quad \tau = [\tau_1 \, \tau_2 \, \ldots \, \tau_q]^T \text{ and}$$
$$\min(Q) = [\min(Q_1) \; \min(Q_2) \; \ldots \; \min(Q_q)]^T$$

The following methodology is devised to provide optimum selection of h and p for a given time and accuracy constraints.

(a) ***Case-I***: When the system is required to reach a decision in a time not exceeding τ^*, interchange different rows of F in such a way that the 4^{th} column of F is converted into ascending order. Denote the resulting matrix by G_1. Let the first m entries of the 4^{th} column of G_1 not exceed τ^*. Find the smallest of the first m entries of the 3^{rd} column of G_1. Let this be the n^{th} one $(n \leq m)$. Then, the n^{th} entries of the 1^{st} and the 2^{nd} columns of G_1 are the optimum choices of h and p, respectively. If all the entries of the 4^{th} column of G_1 are greater than τ^*, then the system cannot meet the requirements. In this case, the first entry of the first and the second columns of G_1 is the best choice of h and p, respectively.

(b) ***Case-II***: When the system has to make a decision subject to a maximum error Q^*, interchange the rows of F so that the 3^{rd} column of F is converted into the ascending order. Let the new matrix be denoted by G_2 and also the first n entries of the third column of G_2 are $\leq Q^*$. Let m^{th} entry of the 4^{th} column of G_2 be the smallest of the first n^{th} entries of this column. In this case, the m^{th} entries of the 1^{st} and the 2^{nd} column of G_2 represent optimum combination of h and p, respectively. If none of the entries of the 3^{rd} column of $G_2 \leq Q^*$, then the first entry of the first and second columns of G_2 is the best choice of h and p, respectively.

(c) ***Case-III***: Consider the case when the system is made to take a decision by comparing simultaneously the constraints discussed in (a) and (b). First, apply (a) and denote the resulted values of h and p by h_a and p_a, respectively, and then apply (b) to get h_b and p_b. Denote the average of h_a and h_b by $\bar{h}$ and that of p_a and p_b by $\bar{p}$. The average $\bar{h}$ and $\bar{p}$ represent optimum values of h and p.

4. **Implementation.** The data sets, which were used for training and testing purposes, contain male and female images of 180×200 resolution [30]. It has more than 150

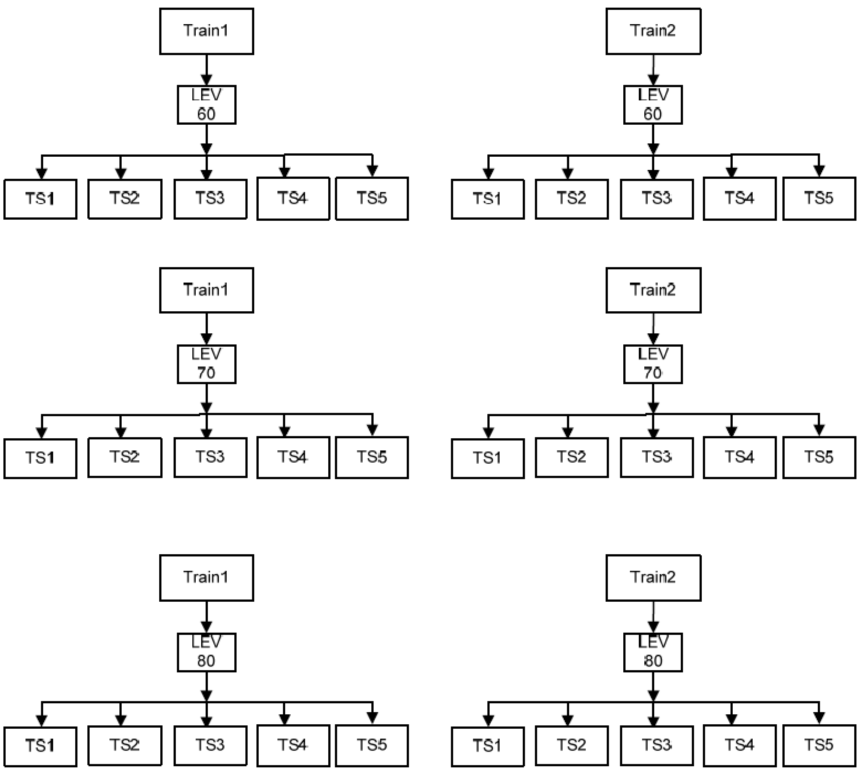

FIGURE 2. Different combination of training and testing data sets

subjects with gray background having minor variation in head turn and tilt. The images were acquired in speech mode with no variation in hair style. However, there was some change in lighting and face position. Over thirty combinations of training and testing data sets were designed as shown in Figure 2. Selection of images in training and testing data sets were random to assess the effectiveness of the proposed technique.

The test data sets had two categories of images. First category was associated with subjects who were present in the training data set but with different facial characteristics. The second category of images had subjects who were not present in the training data.

Assuming that TS1 represents data set having 100 images, TS2 for 200 images and so on. Since each data set was constructed randomly from the image library thus, the probability of two data sets to be 100% identical is rare. On the other hand, Train1 denotes training data set having 450 images and Train2 represents 225 images. Based on the chosen testing and training data sets, six sub trees were constructed for different LEVs as illustrated in Figure 2.

Experiments were then performed for each combination by employing both decomposition and the proposed techniques. A face which is not related to the training subjects and is recognized at verification is treated as false acceptance. The number of false acceptance when divided by the total number of false attempt is called false acceptance rate (FAR). Similarly, a face which is rejected while its subject is used for training is known as false rejection. The number of false rejection when divided by the total number of acceptance is referred to as false acceptance rate (FAR). In face recognition, the same value of FAR and FRR is called equal error rate (EER) which depends on the threshold value of the classifier. EER usually depicts minimum error rate (MER) of a system. However, it is quite possible, in some cases, that the value of MER may be lower than EER.

ROC curves for 225 and 450 training images are generated by applying (6) as shown in Figures 3 and 4, respectively. Both figures clearly demonstrate that the proposed technique has relatively better MER than the decomposition method. The results of the experiments have been summarized in Tables 1 and 2. Comparing the magnitude of MER and EER, it is evident that the proposed technique is significantly better than the decomposition method.

By using (7), the sum of FAR and FRR for various threshold values has been evaluated and shown in Figure 5. The plot also represents least square fitting of experimental data. The simulated curve has a convex profile which defines MER for a given number of LEVs. Whereas, MER variation as a function of LEVs for different threshold values is shown in Figure 6. The plot of Figure 6(a) represents MER of a system designed by using the decomposition method with threshold value as a variable whereas, Figure 6(b) represents the same for the proposed technique and the decision time of the proposed as a function of LEVs is presented in Figure 7.

Since the training sets involve 225 and 450 images, consequently, the maximum Eigen vectors of the reduced covariance matrices are 225 and 450 for Train1 and Train2 respectively [12]. In the proposed system, LEVs are computed in descending order because the subtracting term in the proposed algorithm of Step (vi) is minimum for the first leading Eigen vector. System performance usually improves by increasing number of LEVs, and however, after certain number of LEVs, the system performance saturates. It is observed that the first LEV is of the order of 11, second of the order of 10 and from three to eleven they are of the order of 9. Since for one combination of LEVs, there is an associated ROC curve, a range of ROC curves have been generated starting from 1.4% to 40% of the total Eigen values. It is observed that below 1.4% MER is unacceptably high and by increasing LEVs above 40% does not give a tangible improvement in the system performance.

5. **Results and Discussion.** On the basis of generated ROC curves, MER, the value of p, decision time and number of LEVs are gathered, called offline mode of operation. It is observed that the magnitudes of FAR and FRR are reciprocal in nature, as shown in Figures 3 and 4. Ideally, a point should exist on each ROC curve where both the errors (FAR and FRR) are the same and minimum, i.e., (EER). However, most of the time, it is not the case. Therefore, MER is proposed to characterize a system. If there are more than one MERs, preference will be given to that MER which has lower FAR, to minimize the security risk of a recognition system.

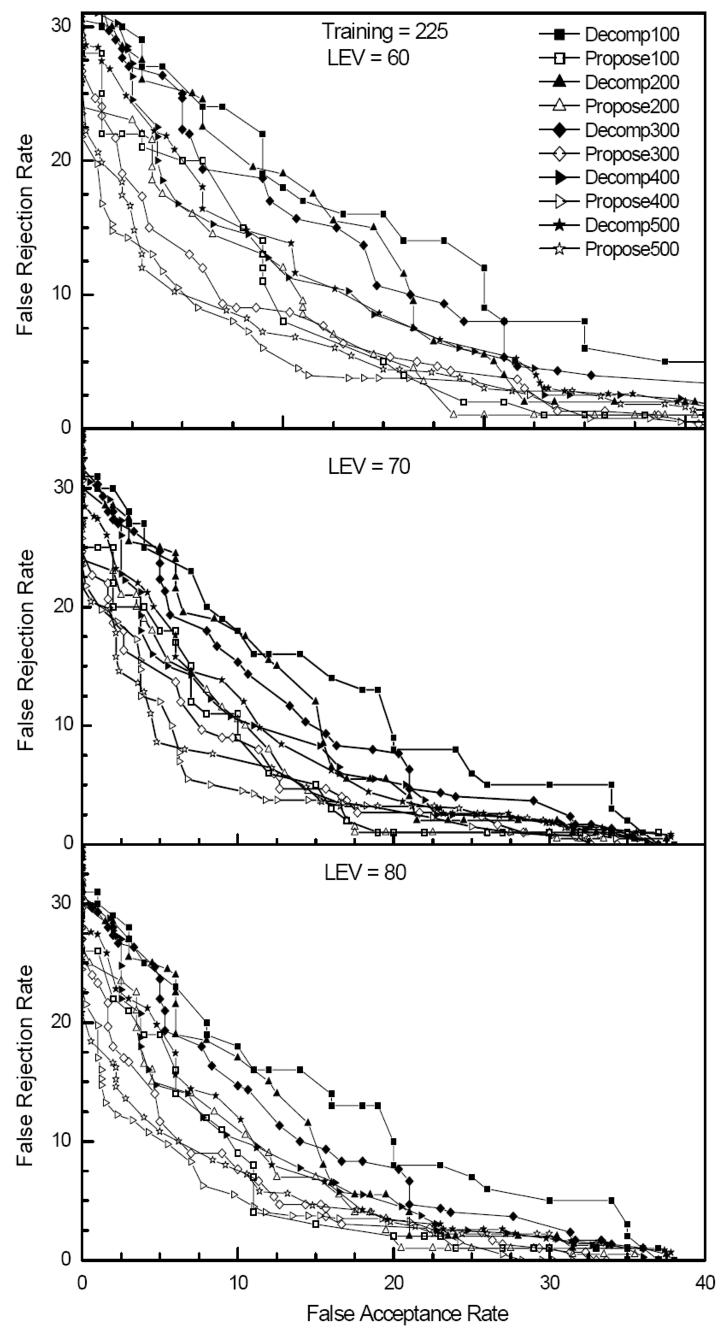

Figure 3. Comparison in false acceptance rate (FAR) vs false rejection rate (FRR) when system is trained with 225 images. Filled symbols represent decomposition method whereas empty symbols show the proposed technique.

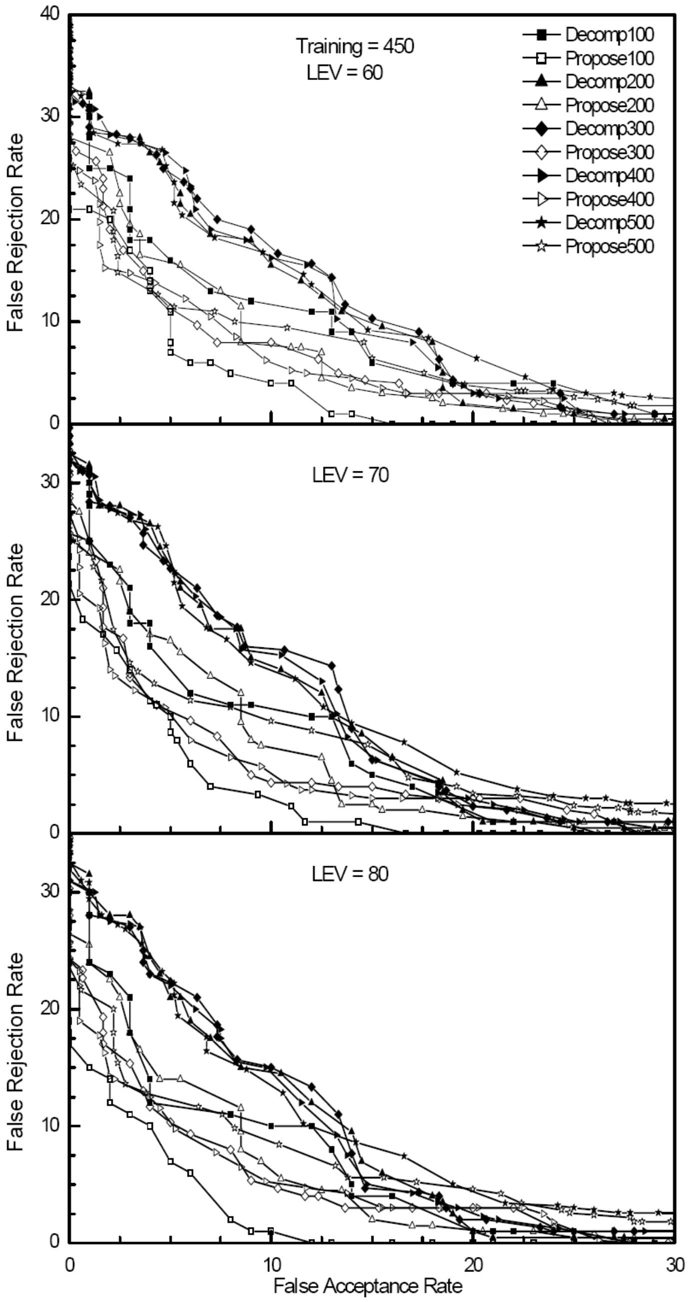

FIGURE 4. Comparison in false acceptance rate (FAR) vs false rejection rate (FRR) when system is trained with 450 images. Filled symbols represent decomposition method whereas empty symbols show the proposed technique.

The magnitude of MER has got an inverse relationship with number of LEVs as evident from Figure 6. However, it is observed that MER saturates between 20% to 30% of LEVs. On the other hand, increasing the number of LEVs increases the decision time of a system. Thus, it is necessary to choose an optimum number of LEVs without compromising on the accuracy as well as on the decision time of a system. Further, it is observed that the value of p fluctuates for lower number of LEVs but it stabilizes when the system includes approximately18% of LEVs. This once again suggests that an appropriate number of LEVs is required to define a stable system.

Table 1. Comparison between decomposition and the proposed technique when system is trained with 225 images

LEV 70	Decomposition		Proposed		% Improvement	
Test Set	**MER**	**EER**	**MER**	**EER**	**MER**	**EER**
TS1	18	20	11	12	38.9	40.0
TS2	21	24	16	17	23.8	29.2
TS3	21	25	14	15	33.3	40.0
TS4	21	25	14	14	33.3	44.0
TS5	24	24	17	19	29.2	20.8

Table 2. Comparison between decomposition and the proposed technique when system is trained with 450 images

LEV 70	Decomposition		Proposed		% Improvement	
Test Set	**MER**	**EER**	**MER**	**EER**	**MER**	**EER**
TS1	27	30	19	21	29.6	30.0
TS2	24	27	18	21	25.0	22.2
TS3	25	25	17	18	32.0	28.0
TS4	21	21	12	13	42.9	38.1
TS5	21	21	13	16	38.1	23.8

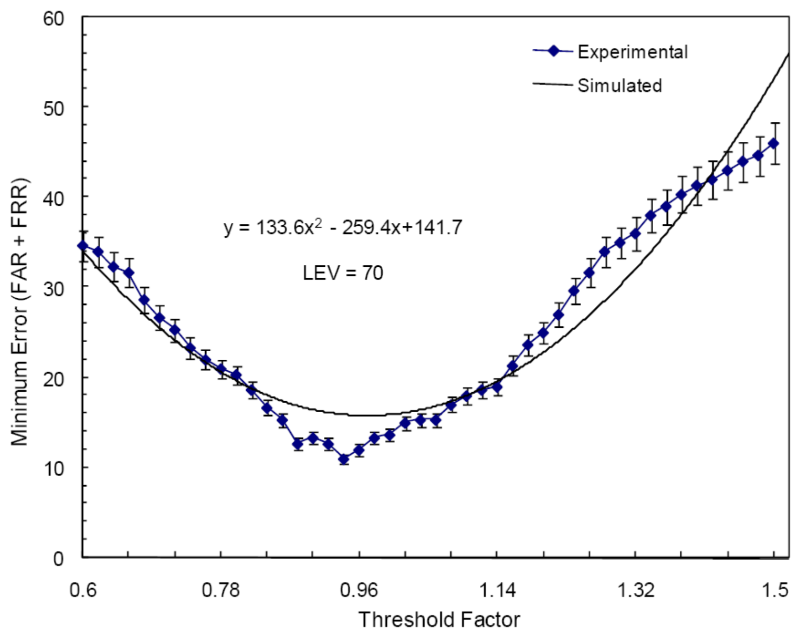

Figure 5. Minimum error as a function of threshold value of the proposed system. Solid line represents the curve fitting by least square method.

From Figure 5, it is evident that MER of the proposed system occurs at $p = 0.95$. Examination of the figure revealed that both the global and the simulated minimum appear at the same value of p. However, the simulated minimum is slightly higher than the experimental one. Figure 8 shows a comparison between experimental and simulated MER as a function of LEVs. An important point to note is that both the curves show a similar profile, but with a finite difference. This demonstrates the validity of simulated data but, of course, with a finite error. To remove the error between the experimental

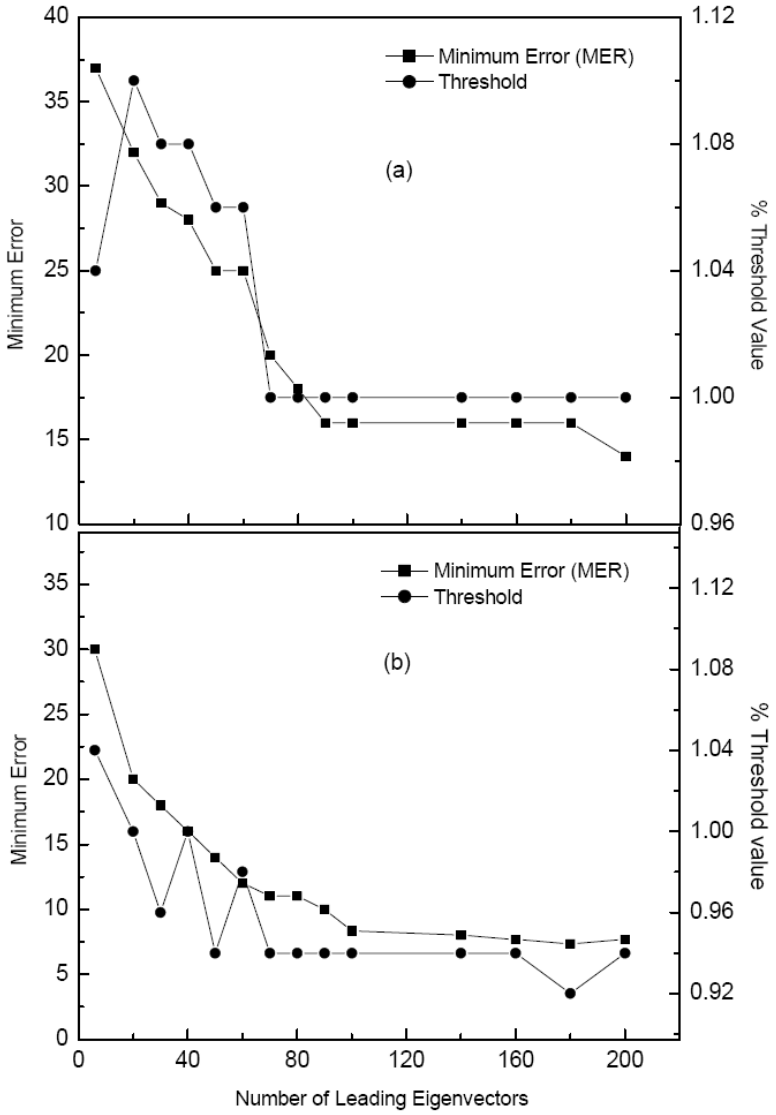

FIGURE 6. Minimum error as a function of leading eigenvectors: (a) for decomposition and (b) for the proposed system

and simulated characteristics following correction factor is proposed:

$$ceil\left(mean\left(\sum_{i=1}^{n}|T_i - E_i|\right)\right)$$

where n represents total data points, T theoretical and E experimental value at i^{th} point respectively. The proposed relationship works well for various training sets as shown in Figure 9. Furthermore, the observed root mean square error, as calculated by (9), between the simulated and the experimental data for various LEVs (6-200) was less than 1.1. This value is well within acceptable error margin for a face recognition system.

$$\sqrt{\frac{\sum_{i=1}^{n}\left(E_i - T_{adj,i}\right)^2}{n}} \tag{9}$$

In an online mode of operation, let F matrix be populated by the information based on Figures 6 and 7 and assuming a maximum allowable decision time of an application

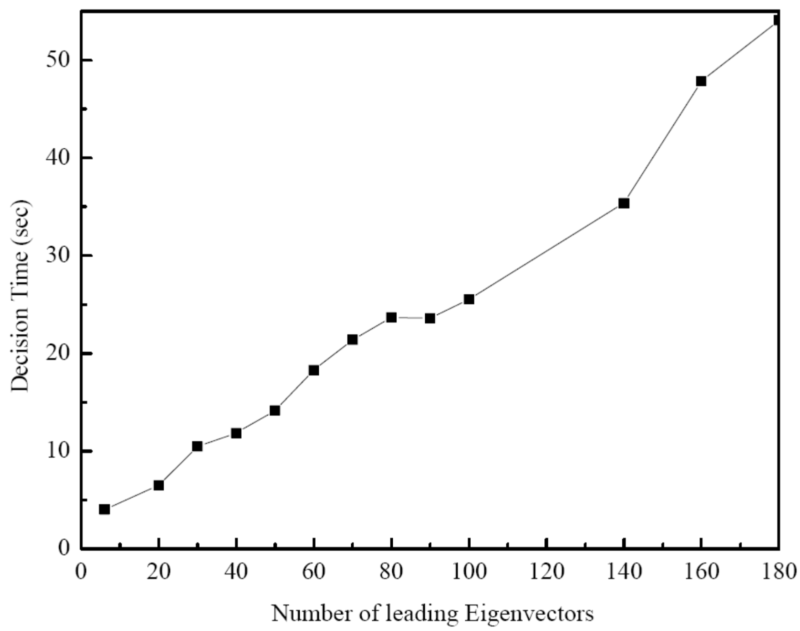

FIGURE 7. Variation of decision time against leading eigenvectors when 450 images are used in training and 100 images in testing

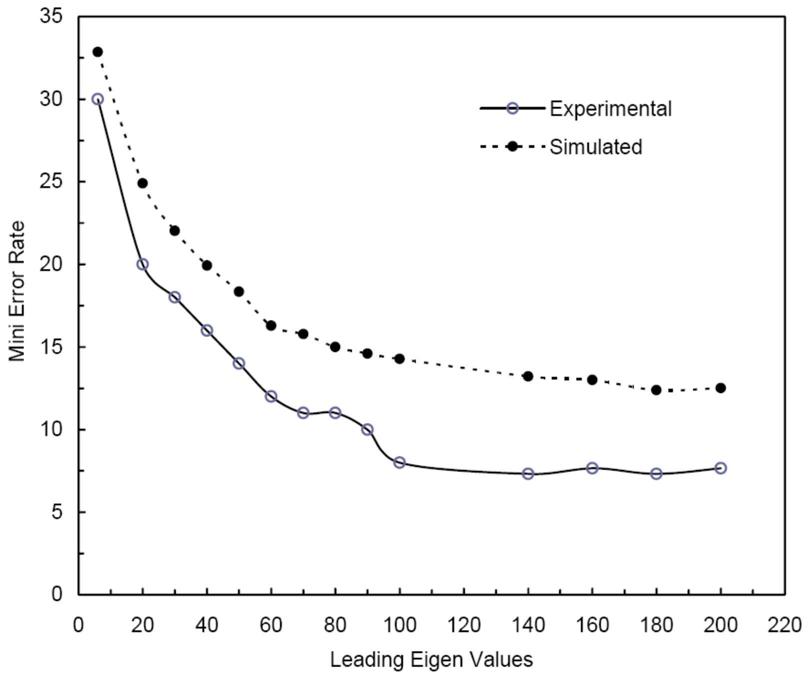

FIGURE 8. Simulated and experimental minimum error vs number of leading eigenvectors

is 0.3 seconds per frame. Then, the system adaptively chooses $h = 100$, such that it provides decision in 0.25 second per frame with an accuracy of 92%, whereas a system defined by the same variable but using decomposition technique provides decision in 0.83 second/frame with an accuracy of 84%.

6. **Conclusions.** A modification in fast PCA has been proposed which improves face recognition in Eigen space. The instability of fast PCA, which is usually observed for

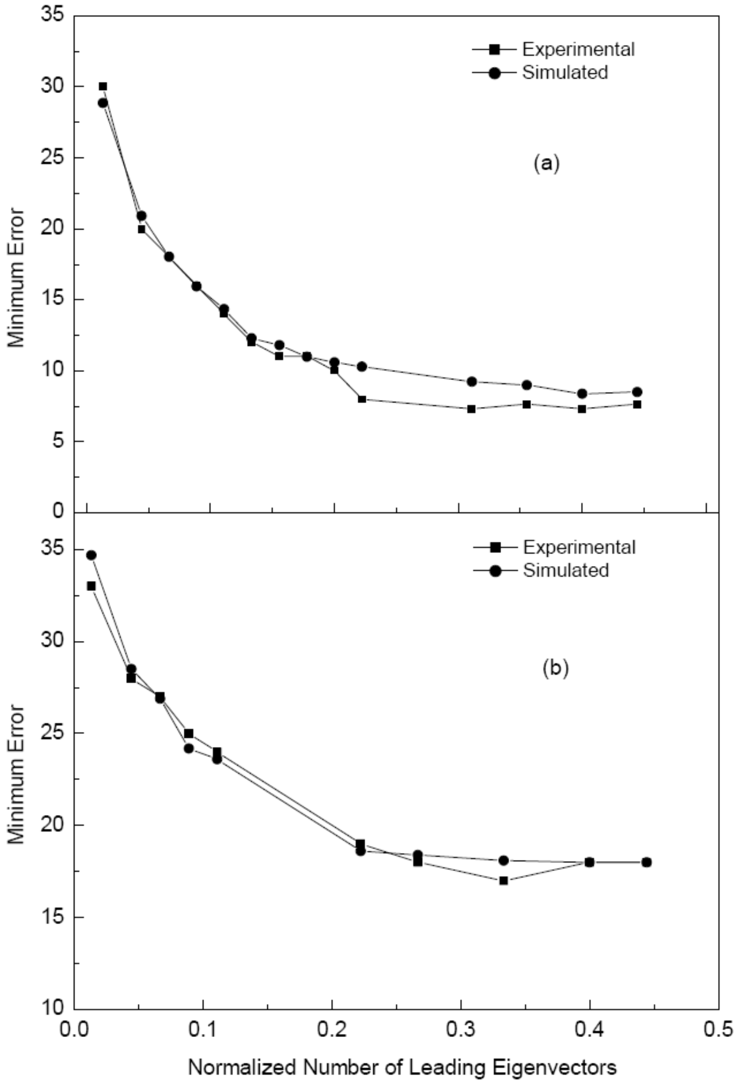

Figure 9. Experimental and simulated data against normalized number of leading Eigen values: (a) for training set of 450 images and (b) for training set of 225 images

high resolution images, has been addressed by incorporating modified Gram-Schmidt orthogonolization. To optimize the system performance least square fitting techniques is used. The proposed system provides an optimum combination of: (a) decision time; (b) minimum error rate and (c) number of leading Eigen vectors. The theoretical trend line is adjusted towards an experimental one by developing a relation which works well on various training sets. A technique is devised to simulate the response of the system. It has been shown, by comparing the simulated and experimental data, that the root mean square error of the proposed system is less than 1.1. It is shown that the system is robust enough and it works reasonably well for images acquired under varying lighting conditions.

REFERENCES

[1] M. A. Martinez and A. C. Kak, PCA versus LDA, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.23, no.2, pp.228-233, 2001.

[2] X. He, S. Yan, Y. Hu, P. Niyogi and H.-J. Zhang, Face recognition using Laplacian faces, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.27, no.3, pp.328-340, 2005.

[3] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, Face recognition: A literature survey, *ACM Computing Surveys (CSUR)*, vol.35, no.4, pp.400-459, 2003.
[4] R. Chellappa, C. L. Wilson and S. Sirohey, Human and machine recognition of faces: A survey, *Proc. of IEEE*, vol.83, no.5, pp.705-740, 1995.
[5] J. Daugman, Face and gesture recognition: Overview, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.675-676, 1997.
[6] S. Pankanti, R. M. Bolle and A. Jain, Biometrics: The future of identification, *Computer*, vol.33, no.2, pp.46-49, 2000.
[7] A. Pentland and T. Choudhury, Face recognition for smart environments, *Computer*, vol.33, no.2, pp.50-55, 2000.
[8] C.-Y. Chang and H.-R. Hsu, Application of principal component analysis to a radial-basis function committee machine for face recognition, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4145-4154, 2009.
[9] S. D. Lin, J.-H. Lin and C.-C. Chiang, Combining scale invariant feature transform with principal component analysis in face recognition, *ICIC Express Letters*, vol.3 no.4(A), pp.927-932, 2009.
[10] Y.-W. Chen, R. Xu and A. Ushikome, Serially-connected dual 2D PCA for efficient face reprefentation and face recognition, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4367-4372, 2009.
[11] D. Zhao, Z. Lin and X. Tang, Laplacian PCA and its applications, *Proc. of IEEE the 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, pp.1-8, 2007.
[12] M. Turk and A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, vol.3, no.1, pp.71-86, 1991.
[13] A. Sharma and K. K. Paliwal, Fast principal component analysis using fixed-point algorithm, *Pattern Recognition Letters*, vol.28, no.10, pp.1151-1155, 2007.
[14] I. Sajid, M. M. Ahmed and I. Taj, Design and implementation of a face recognition system using fast PCA, *IEEE Proc. of the International Symposium on Computer Science and Its Applications*, Hobart, Australia, pp.126-130, 2008.
[15] I. Sajid, M. M. Ahmed and I. Taj, Time efficient face recognition using stable Gram-Schmidt orthonormalization, *International Journal of Signal Processing, Image Processing and Pattern*, vol.2, no.1, pp.35-48, 2009.
[16] D. Cai, X. He, J. Han and H.-J. Zhang, Orthogonal laplacianfaces for face recognition, *IEEE Trans. on Image Processing*, vol.15, no.11, pp.3608-3614, 2006.
[17] R. Krisnamoorthy and R. Bhavani, Face recognition system based on orthogonal polynomials, *Journal of Applied Sciences*, vol.7, no.1, pp.109-114, 2007.
[18] H. Hu, Orthogonal neighborhood preserving discriminant analysis for face recognition, *Journal of Pattern Recognition Society*, vol.41, no.6, pp.2045-2054, 2007.
[19] R. L. Burden and J. D. Faires, *Numerical Analysis*, 7th Edition, Thomson, 1997.
[20] A. Bjorck, Solving linear least squares problems by Gram-Schmidt orthogonalization, *BIT Numerical Mathematics*, vol.7, no.1, pp.1-21, 1967.
[21] L. Giraud, J. Langou, M. Rozloznik and J. van den Eshof, Rounding error analysis of the classical Gram-Schmidt orthogonalization process, *Numerische Mathematik*, vol.101, no.1, 2005.
[22] A. Bjorck, *Numerical Method for Least Squares Problems*, SIAM, 1996.
[23] H. Simon, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 2008.
[24] J. Siwabessy, J. Penrose, R. Kloser and D. Fox, Seabed habitat classification, *Proc. of International Conference on High Resolution Surveys in Shallow Water*, Sydney, 1999.
[25] E. J. Rzempoluch, *Neural Network Data Analysis Using Simulnet*, Springer, 1998.
[26] A. J. O. Toole, P. J. Phillips, F. Jiang, J. Ayyad, N. Penard and H. Abdi, Face recognition algorithms surpass human matching faces over changes in illumination, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.27, no.9, pp.1642-1646, 2007.
[27] S. Yan, D. Xu and X. Tang, Face verification with balanced thresholds, *IEEE Trans. on Image Processing*, vol.16, no.1, pp.262-268, 2007.
[28] J.-B. Li, Nonparametric kernel discriminant analysis for face recognition under varying lighting conditions, *ICIC Express Letters*, vol.4, no.3(B), pp.999-1004, 2010.
[29] N. Mittal and E. Walia, Face recognition using improved fast PCA algorithm, *Proc. of Congress Image and Signal Processing*, pp.554-558, 2008.
[30] L. Spacek, *Collection of Facial Images: Faces94*, http://cswww.essex.ac.uk/mv/allfaces/, 2008.