Performance of LDA Versus PCA in Nonideal Lighting Environment

Hikmat Darwish

Abstract Automatic face recognition (AFR) is used to identify people by processing their photos or snapshots of faces, either in off-line or real-time manners, respectively. However, classical face recognition techniques have been reported to suffer from substantial degradation in performance when person image is subjected to nonideal lighting or some types of occlusion. In real life we may well encounter a certain type of nonideal lighting such as side-shadowing of the face, where substantial part of the face can be totally occluded or masked. In this work, we examine and evaluate the performance of two famous statistical approaches for AFR namely PCA and LDA in terms of face recognition rate (FRR), when both are operating on particular ill-illuminated image exemplified by side-shadowing occlusion with addition of "salt-and-pepper" noise, which is often the encountered case. The two suggested AFR techniques are the well-reputed principal component analysis (PCA) and linear discriminate analysis (LDA). A computer simulation has been executed testing both PCA and LDA and the simulation outcomes indicate much better performance of LDA over PCA in terms of FRR for this particular type of image occlusion.

Keywords Automatic face recognition (AFR) · Principle component analysis (PCA) · Linear discriminate analysis (LDA) · Covariance matrix · Eigen vector · Feature vector · Minimum Euclidian distance · Rate of recognition (ROC)

1 Introduction

In general, AFR can be accomplished through class discrimination which can be realized and implemented through numerous methods and techniques. While some of these techniques are based on deterministic principle, such as the technique of

H. Darwish (✉)
Telecommunication Department, Faculty of Engineering, AlWatania Private University, Hamah, Syria
e-mail: hikmat.darwish@wpu.edu.sy

discrete Fourier transform (DCT) and Gabor filter [1–3], other AFR techniques can be considered to have statistical basis. They are mainly principal component analysis (PCA) [4–6], linear discernment analysis (LDA) [7, 8], support vector machine (SVM) [9], and artificial neural networks (ANN) [10].

In this paper we are interested in applying two different techniques to AFR, namely, PCA and LDA operating on still face images but subjected to gradient-shadowing being a specific image occlusion. The motivation can be justified by the fact that side-shadowing is the occlusion imposed on face images most often. The paper contains basic theoretical analysis of both techniques, illustrating the advantage gained by LDA over PCA technique in terms of enhanced capability of LDA compared with PCA with regard to class separation. Computer simulation evaluates the performance of each above techniques in terms of recognition rate against increased levels of gradient-shadowing occlusion.

2   Literature Review

AFR which is based on PCA principles was first suggested by M. Kirby and L. Sirovish [6] in 1990, and it was utilized by And M. Turk and A.P. Pentland [7] later on. Aleix M. Martinez and Avinash C. Kak [8] have tested the performance of both PCA and LDA against subspace dimensionality. The simulation results showed a slight improvement of 20% on average in terms of FRR in favor of LDA over PCA. Similar results were obtained by Peng P. et al. [11]. The work of Chandolu P. and Jayesh G [12] illustrate almost 16% on average FRR improvement of LDA performance over PCA when the techniques run against increased number of face samples. Similar results have been obtained by Önsen T. and Adnan A. [13]. Tomesh Verma and Raj Kumar Sahu [14] tested LDA-PCA technique which incorporate PCA feature vectors versus PCA technique alone. Simulation results exhibit once again 16–20% improvement in FRR gained by LDA-PCA over PCA alone. However, the work does not include the effect of face image occlusion on the FRR for both techniques, in addition to the fact that the combined LDA-PCA can prove to be too costly in terms of the required processing speed for real-time application. The work of Zhao W. et al. [15] retains similar result but with different type of face image occlusions.

3   PCA Technique

3.1   Theoretical Background

PCA technique is based on Eigen vector decomposition of data covariance matrix. The information of data, say, that of face image, is represented by a set of most significant Eigen vectors of the covariance matrix. This basic concept of PCA will be explained in the following example. Let X and Y be data vectors such as

$$X = [2.3\ 0.5\ 2.2\ 1.9\ 3.1\ 2.3\ 2\ 1\ 1.5\ 1.1]$$

and

$$Y = [2.4\ 0.7\ 2.9\ 2.2\ 3\ 2.7\ 1.6\ 1.1\ 1.6\ 0.9]$$

The two data vectors are plotted in Fig. 1. The sample means are

$$\overline{X} = \frac{\sum_{i=1}^{10} x_i}{10} = 1.81 \tag{1}$$

and

$$\overline{Y} = \frac{\sum_{i=1}^{10} y_i}{10} = 1.91 \tag{2}$$

The normalized data vectors to the sample means are

$$A = X - \overline{X}$$
$$= [0.69 - 1.31\ 0.39\ 0.09\ 1.29\ 0.49\ 0.19 - 0.81 - 0.31\ 0.71]$$

and

$$B = Y - \overline{Y}$$
$$= [0.49 - 1.21\ 0.99\ 0.29\ 1.09\ 0.79 - 0.31 - 0.81 - 0.31\ 1.01]$$

Then the covariance matrix is

$$C = \begin{bmatrix} 0.61655 & 0.61544 \\ 0.67787 & 0.71655 \end{bmatrix}$$

And the corresponding Eigen vectors are

$$e = \begin{bmatrix} -0.73517 & -0.67787 \\ -0.67787 & 0.73517 \end{bmatrix}$$

The two Eigen vectors represent two PCA components—PCA1 and PCA2—as shown in Fig. 1. In general, we chose Eigen vectors so that data have maximum variance along these vectors to be the PCAs and neglect the rest. Data can be projected on these major PCAs and thus reducing data dimensionality. In our example, we have chosen the first Eigen vector $e^T = [-0.73517 - 0.67787]$ as single PCA and neglected the other.
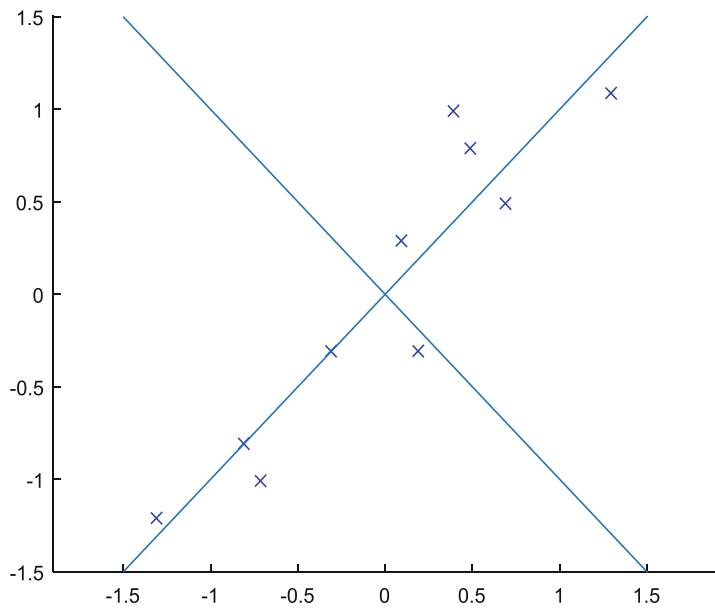
Fig. 1 Ten data points representing X and Y vectors
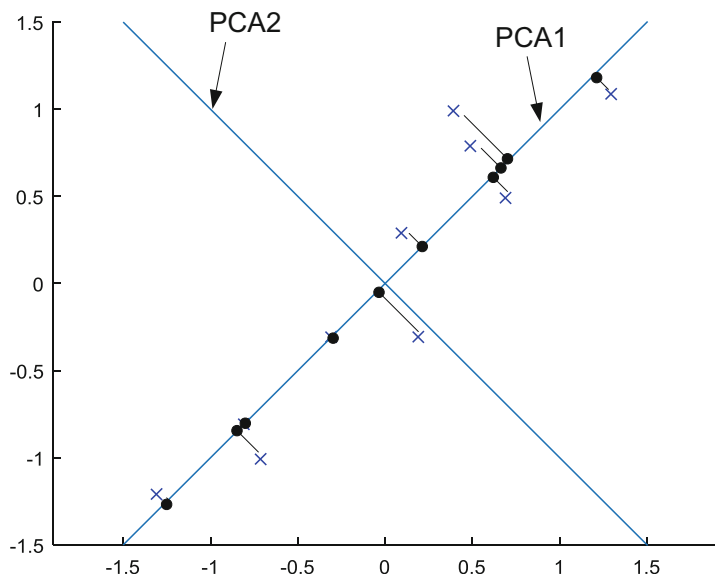
Fig. 2 Projection of data on main PCA vector

When data is projected on this PCA (black dots), the data dimensionality (size) is reduced (compressed) from two dimensions to one dimension as shown in Fig. 2. This is the first function of PCA analysis.

However, and most importantly, these two Eigen vectors represent what is called *feature vectors* of the data, and hence they can be used for recognition in general and for face recognition in particular. Which is considered to be the second benefit of PCA principle.

3.2 PCA Applied to Face Recognition

Regardless of the technique being used, AFR is usually accomplished by executing two phases. These are training phase and recognition phase.

3.2.1 Training Phase

Stage 1: The system starts by digitizing the camera snapshot into 2D face image of black/white intensity of size $n \times m$ pixels. Also, in this stage, pixels of the 2D image are concatenated to form 1D image vector X of size $(L \times 1)$ where $L = m \times n$.

Stage 2: Calculating the "average face" which is the averaging of M given faces vectors such that

$$\varphi = \frac{1}{M} \sum_{i=1}^{M} X_i \tag{3}$$

Stage 3: Normalizing all available M face images to zero average value by subtracting

$$\psi_i = X_i - \varphi \tag{4}$$

Stage 4: Constructing face covariance matrix C and calculating PCAs through the following:

Construct face image matrix

$$A = \begin{bmatrix} \psi_1^T & \psi_2^T \dots & \psi_M^T \end{bmatrix} \tag{5}$$

Calculate covariance matrix

$$C = \frac{1}{M} A A^T \tag{6}$$

where A^T is transpose of matrix A.

Calculate M Eigen vectors of C matrix as

$$e = Av \tag{7}$$

where v is the largest Eigen vectors of matrix:

$$C = A^T A \tag{8}$$

Vectors e of largest corresponding λ are the required PCA for this person image.

Stage 5:

Project face images on PCAs (Eigen vectors) as

$$P_i = e_i^T \psi_i \tag{9}$$

and form person image feature vector

$$\Omega = [P_1 P_2 \ldots . P_M]^T \tag{10}$$

3.2.2 Recognition Phase

Stage 1: Enter query images, repeat stages 1–5, and obtain the feature vector of the query image Ω_x.

Stage 2: Calculate Euclidian distance between query image feature vector and each one of the stored images.

$$d_j = \left\| \Omega_x - \Omega_j \right\| \tag{11}$$

where $J=1,2\ldots,K$, where K is the number of persons in the data base and consequently identify the query image that has the index

$$j = \text{argmin}(d_j) \tag{12}$$

4 LDA Technique

Principal component analysis (PCA) is an effective method mainly for reducing data dimensionality, thus for data compression, and it has been used for face recognition systems effectively. However, PCA technique does not take into account class separation. In other words, there could be two well-separated classes of data having the same spreading direction as shown in Fig. 3, and the PCA technique considers them as single class with one Eigen vector.

Another statistical technique that is called Fisher's linear discernment analysis or LDA for short which was first introduced by Sir Ronald Fisher in 1936, and reformulated later on by S. Balakrishnama and A. Ganapathiraju [16] and Alaa Tharwat et al. [17], provides both Eigen vectors and class separation at the same time; hence, there could be a big chance for the LDA to outperform PCA technique in terms of ROC for a wide verities of ill-illuminated images.

The main principle of LDA technique is to find the Eigen vectors that maximize the separation between classes in the data so that a powerful class recognition system may be obtained.

With LDA technique it may well be possible to obtain both PCA projection and concise class separation at the same time (Fig. 3a, b).

In order to drive mathematical formula for obtaining these projection vectors, let us assume two classes X_1 and X_2. Each class contains M number of the same person's image vectors, of which each is of $N \times 1$ size such as
$X_1 = \left(X_1^1, X_1^2, \ldots, X_1^M \right)$. The mean value for class X_1 is

$$\mu_1 = \frac{1}{M} \sum_{i=1}^{M} X_{1,i} \tag{13}$$
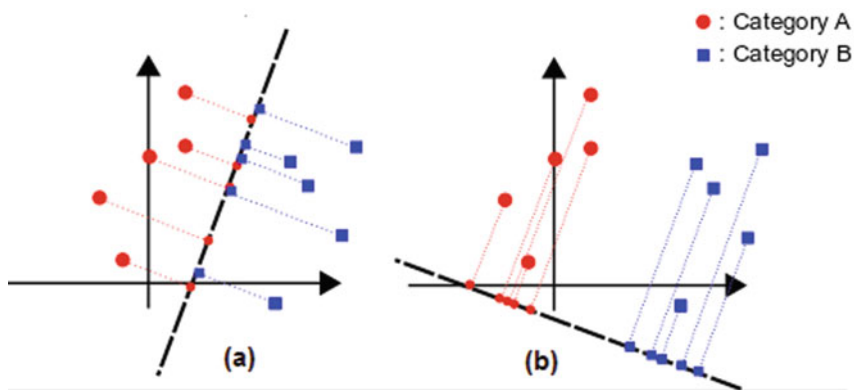
Fig. 3 (**a**) Projection of data on main PCA. (**b**) Projection of data on main LDA showing effective class (category) separation

$$\mu_2 = \frac{1}{M} \sum_{i=1}^{M} X_{1,i} \qquad (14)$$

And the mean value for class X_2 is

$$\mu_2 = \frac{1}{M} \sum_{i=1}^{M} X_{2,i} \qquad (15)$$

Note that both μ_1 and μ_2 are vectors on size $N \times 1$.

Let w be the LDA projection vectors that will guarantee the separation of face image classes, and then the mean of the first face image projection on this vector is

$$\tilde{\mu}_1 = w^T \mu_1 \qquad (16)$$

And the mean of second face image projection on this vector is

$$\tilde{\mu}_2 = w^T \mu_2$$

The projection of class X_1 on vector w is the vector Y_1 such that

$$Y_1 = w^T X_1 \qquad (17)$$

And for class X_2, the projection vector is Y_2 such that

$$Y_2 = w^T X_2 \qquad (18)$$

The distance between these classes' means is

$$d(w) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |w^T \mu_1 - w^T \mu_2| = |w^T (\mu_1 - \mu_2)| \qquad (19)$$

The covariance matrix, otherwise known as scatter matrix for projection Y_1 as

$$\tilde{S}_1 = \sum_{N} (Y_1 - \tilde{\mu}_1)(Y_1 - \tilde{\mu}_1)^T \qquad (20)$$

And for class Y_2

$$\tilde{S}_2 = \sum_{N} (Y_2 - \tilde{\mu}_2)(Y_2 - \tilde{\mu}_2)^T \qquad (21)$$

Rewriting Eq. (20),

$$\tilde{S}_1 = \sum_{N} w^T (X_1 - \tilde{\mu}_1)(X_1 - \tilde{\mu}_1)^T w = w^T \left[\sum_{N} (X_1 - \tilde{\mu}_1)(X_1 - \tilde{\mu}_1)^T \right] w \qquad (22)$$

Or

$$\tilde{S}_1 = w^T S_{w1} w \tag{23}$$

where

$$S_{w1} = \sum\nolimits_N (X_1 - \tilde{\mu}_1)(X_1 - \tilde{\mu}_1)^T \tag{24}$$

is called within-class scatter matrix.

Then

$$\tilde{S}_1 + \tilde{S}_2 = w^T S_w w \tag{25}$$

where

$$S_w = S_{w1} + S_{w2} \tag{26}$$

is again called within-class scatter matrix.

Now in order to obtain the required LDA projection vector w which will separate the two classes, we seek the vector w that maximizes the ratio of mean distance to within-class matrices:

$$\rho(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{S}_1 + \tilde{S}_2} \tag{27}$$

On the other hand, rewriting

$$|\tilde{\mu}_1 - \tilde{\mu}_2|^2 = (w^T \mu_1 - w^T \mu_2)^2 = w^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w \tag{28}$$

$$|\tilde{\mu}_1 - \tilde{\mu}_2|^2 = w^T S_B w \tag{29}$$

where

$$S_{Bi} = (\mu_i - \mu_m)(\mu_i - \mu_m)^T \tag{30}$$

is called the between-class scatter matrix.

In general and for N face images, the between-class scatter matrix is given by

$$S_B = \sum\nolimits_{i=1}^N (\mu_i - \mu_m)(\mu_i - \mu_m)^T \tag{31}$$

Then the ratio to be maximized in order to get class separation becomes

$$\rho(w) = \frac{w^T S_B w}{w^T S_w w} \tag{32}$$

The LDA-required projection vector is obtain by

$$w = \text{argmax}\left(\frac{w^T S_B w}{w^T S_w w}\right) \tag{33}$$

And this can be carried out by doing

$$\frac{d\rho(w)}{w} = 0 \tag{34}$$

And this leads to

$$S_w^{-1} S_B w = \rho w \tag{35}$$

The interpretation of Eq. (35) is that the required LDA projection w vector is the Eigen vector of matrix

$$A = S_w^{-1} S_B \tag{36}$$

which can be obtained in similar manner to PCA technique.

Example:

Let two images be represented by the two vectors:

$$X_1 = \{(4, 1), (2, 4), (2, 3), (3, 6), (4, 4)\}$$
$$X_2 = \{(9, 10), (6, 8), (9, 5), (8, 7), (10, 8)\}$$

Then we get

$$S_1 = \begin{bmatrix} 0.8 & -0.4 \\ -4.0 & 2.6 \end{bmatrix}, S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64. \end{bmatrix}$$

and

$$\mu_1 = [3.00 \ 3.60],$$
$$\mu_2 = [8.40 \ 7.60]$$

Substituting in Eqs. (35) and (36), we get

$$S_w = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix},$$

$$S_B = \begin{bmatrix} 29.16 & 21.60 \\ 21.60 & 16.00 \end{bmatrix}$$

$$S_w^{-1} S_B = \begin{bmatrix} 11.98 & 8.81 \\ 5.08 & 3.76 \end{bmatrix}$$

Solving Eq. (35) we get

$$S_w^{-1} S_B w = \lambda w \longrightarrow \left| S_w^{-1} S_B - \lambda I \right| =$$

$$\begin{vmatrix} 11.98 - \lambda & 8.81 \\ 5.08 & 3.76 - \lambda \end{vmatrix} = 0 \longrightarrow \lambda = 15.65$$

Then

$$\begin{bmatrix} 11.98 & 8.81 \\ 5.08 & 3.76 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 15.65 \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

Solving for w_1, w_2

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix}$$

Figure 4 illustrates the two image classes and the LDA projecting vector on which the projections of two images can be easily separated.

5 Computer Simulation

A computer simulation for evaluating the performance of both PCA and LDA techniques using the same images have been implemented through writing MatLab program which was run on laptop computer core i5 machine.

The simulation consists of processing face images of 100 different person samples. Each face image is represented by 200×200 pixels of monochromatic gray intensity, a sample of which is shown in Fig. 5.

These face images have been borrowed with courtesy from free image library of the University of Sheffield [18].

The simulation consists of both training and recognition phases for each of PAC and LDA techniques. In training phase, the data base which contains feature vectors are obtained using ideal images regarding the lighting conditions. Only in
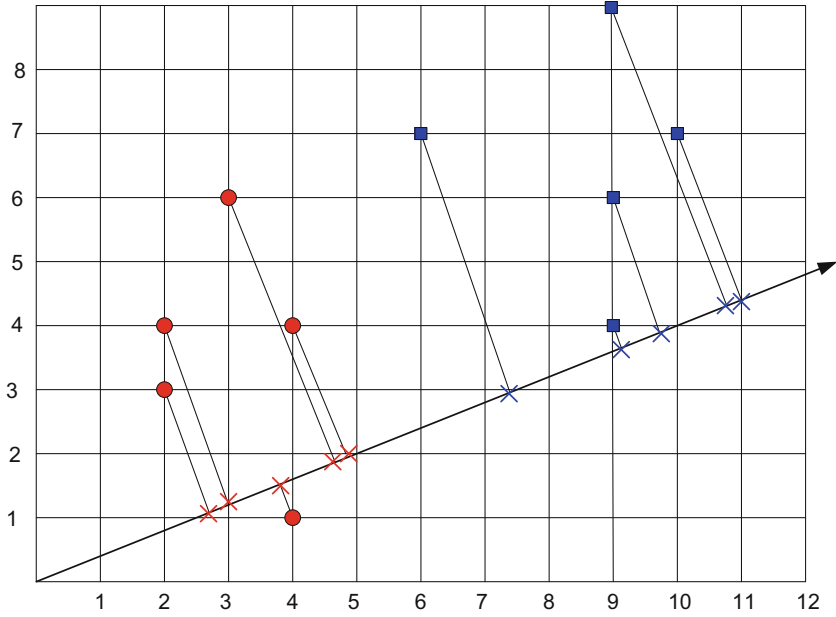
Fig. 4 Projection of two simple images on the main LDA vector showing clear class separation
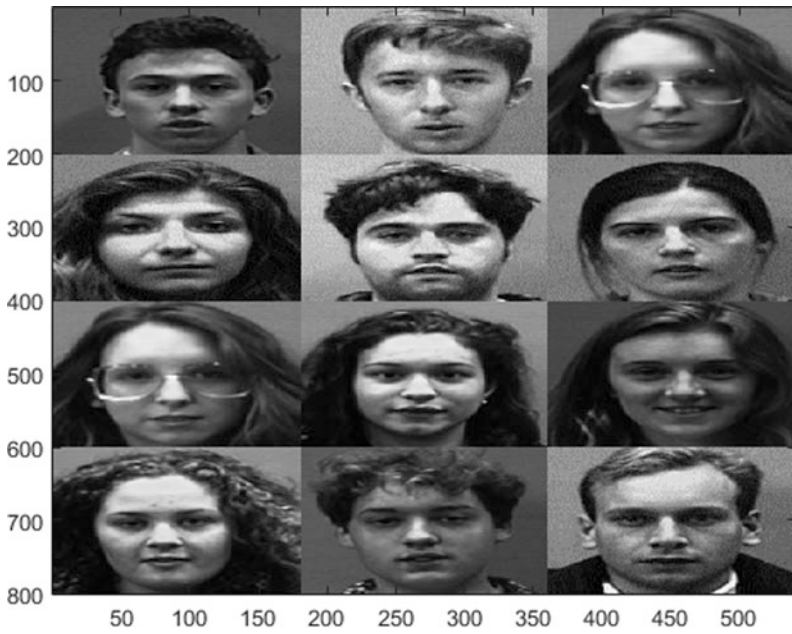

Fig. 5 Face samples taken from the University of Sheffield

Fig. 6 Side-shadowing occlusion of deferent qualitative grades: (**a**) multi $= 0.3$ and (**b**) multi $= 0.6$
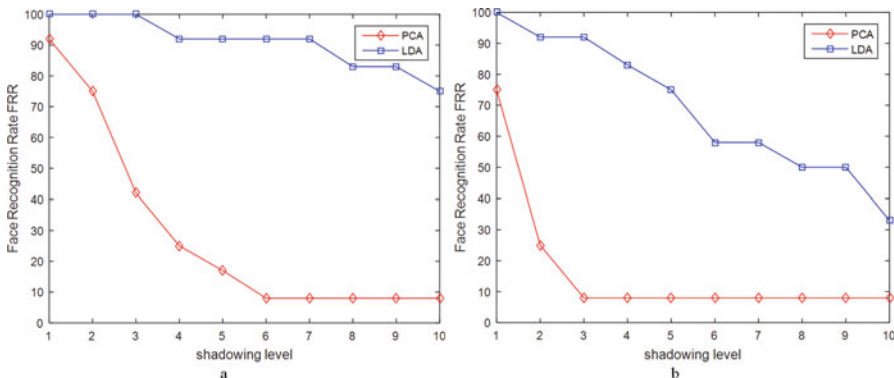

Fig. 7 Face recognition rate of PCA versus LDA. (**a**) multi $= 0.3$ and (**b**) multi $= 0.6$

recognition phase that occlusion are introduced over images for testing. Since this work is mainly concerned with relative comparison of PCA and LDA, then qualitative image occlusion might be a sensible choice. Having said that, within the recognition phase, the nonideal (*shadowing level*), which is acting as image occlusion, was introduced by adding a black hue of linear gradient to each image. Figure 6a, b illustrates the increased black hue on two separate experiments with multiplying factor multi $= 0.3$ and multi $= 0.6$ respectively.

Each of PCA and LDA algorithms is performed on the same corrupted images set after adding "salt-and-pepper" noise so that we get random *ensample* which is necessary to calculate the FFR which is in turn is a probability. The performance of both PCA and LDA was evaluated in terms of FRR against increased hue intensity (multiplying factor) which is denoted by *shadowing level* as it is depicted in Fig. 7a, b. Figure 7a depicts the simulation outcome for average black hue of multi $= 0.3$, while Fig. 7b depicts simulation outcomes for multi $= 0.6$. By studying these figures, we notice the following: first, for ideal illumination, i.e., shadowing $= 0$.

We may notice that the improvement of FRR gained by LDA over PCA is ranging between 10% and 25%, which is in complete agreement with the works of [19–21]. Secondly, as the severity of occlusion increase, the performance in terms of FRR of PCA deteriorates rapidly, while that of LDA deteriorates gracefully. Even we get fairly constant and good FFR of 80% for mild shadowing (Fig. 7b). Thirdly,

still LDA technique maintains higher FRR compared with PCA for all levels of shadowing occlusions. From FRR curves in Fig. 6a, b, we notice that LDA technique outperforms PCA techniques in terms of higher FRR for the same degree of nonideal image lighting.

6   Conclusion

In principle, theoretical formula which may enable us to quantify the difference in performance between PCA and LDA in terms of FRR is needed. However, this task can be proved to be highly difficult. Monte Carlo simulation as it was used in this work may provide a valid alternative. And as such, we succeeded in getting tangible indication that LDA can outperform PCA technique in terms of FRR versus ill-lighting condition. Both figures show a graceful degradation in performance in the case of LDA techniques with respect to worsening lighting conditions, while PCA fails fast and dramatically. However, the price is a much increased computation burden in the case of LDA, and this might not be a prohibiting factor given the fact that computation power of modern computers might fulfill the requirement for implementing LDA-based face recognition system for real-time applications.

References

1. C. Liu, H. Welchsler, Gabor feature classifier for face recognition. Proc. ICCV **2**(5), 270–275 (2001)
2. B. Kepenekci, F.B. Tek, G.B. Akar, Occluded face recognition by using Gabor features, in *3rd COST 276, Workshop on Information and Knowledge Management for Integrated Media Communication, Budapest*, (2002)
3. T.M. Abhishree, J. Latha, K. Manikantan, S. Ramachandran, Face recognition using Gabor filter based feature extraction with anisotropic diffusion as a pre-processing technique. Procedia Comp. Sci. **45**(C), 312–321 (2015). https://doi.org/10.1016/j.procs.2015.03.149
4. K. Pearson, On lines and planes of closest fit to systems of points in space. Philos. Mag. **2**, 559–572. University College London (1901)
5. H. Hotelling, Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. **24**, 417–441 and 498–520 (1933)
6. M. Kirby, L. Sirovish, Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 103–108 (1990)
7. M. Turk, A.P. Pentland, Eigen faces for recognition. J. Cogn. Neurosci., 71–86 (1991)
8. A.M. Martinez, A.C. Kak, PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 228–233 (2001)
9. G. Guo, S.Z. Li, K. Chan, Face recognition by support vector machines, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), 28–30 March 2000*,
10. S.A. Nazeer, N. Omar, Marzuki Khalid Face Recognition System using Artificial Neural Networks Approach, 2007 International Conference on Signal Processing, Communications and Networking

11. P. Peng, I. Portugal, P. Alencar, D. Cown, *A Face Recognition Software Framework Based on Principal Component Analysis* (University of Waterloo, Waterloo, 2021)
12. C. Prasanthi, J. Gangrade, Face recognition for different facial expressions based on PCA, LD analysis. Int. J. Dev. Res. **07**(07), 14109–14112 (2017)
13. Ö. Toygar, A. Acan, Face recognition using PCA, LDA and ICA approaches on colored images. J. Electr. Electron. Eng. **3**(1), 735–743 (2003)
14. T. Vermaand, R.K. Sahu, PCA-LDA based face recognition system & results comparison by various classification techniques, in *Proceeding of International conference on Green High Performance Computing, India 2013*,
15. W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face recognition: A literature survey, in *Technical Report CAR-TR-948, CS-TR-4167, N00014-95-1-0521*, (October 2000)
16. S. Balakrishnama and A. Ganapathiraju. Linear discriminant analysis-a brief tutorial. Institute for Signal and information Processing, 1998
17. A. Tharwat, T. Gaber, A. Ibrahim, A.E. Hassanien, *Linear Discriminant Analysis: A Detailed Tutorial* (Ai Communications, 2017)
18. https://www.sheffield.ac.uk/eee/research/iel/research/face
19. F. Mahmud, M.T. Khatun, S.T. Zuhori, S. Afroge, M. Aktar, B. Pal, Face recognition using principle component analysis and linear discernment analysis, in *International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, (2015)
20. J. Yang, H. Yu, W. Kunz, "Linear Discernment Analysis" school of computer science interactive systems laboratories Carnegie Mellon University Pittsburgh, pa 15213
21. P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)