Comparison of face recognition algorithms in terms of the learning set selection

Simon Gangl*
Domen Mongus†
Supervised by: Borut Žalik‡

Laboratory for **Ge**ometric **M**odelling and **M**ultimedia **A**lgorithms
Faculty of Electrical Engineering and Computer Science
University of Maribor / Slovenia

Abstract

A suitable selection of facial features is of key importance for the successfulness of face recognition algorithms. Because a straightforward selection of them does usually not ensure sufficient reliability, statistical tools are often used for feature extraction. In this paper the influence of the selected set of learning samples on the efficiency of face recognition algorithms is observed. For this purpose, three of the most often used algorithms are presented in detail. The feature description based on the Gabor wavelet transformation is presented first. In this approach features are selected based on human physiognomy basis and formed to feature graphs, where the actual recognition is performed by graph matching. On the other hand, principal component analysis (PCA) is a statistical tool for identifying patterns in data by reducing its dimensionality. That way, key features for face recognition are extracted to a comparable form. Meanwhile, linear discriminant analysis (LDA) allows for face recognition by establishing the borders between classes in multidimensional data. To ensure equal conditions for those algorithms, a method for image normalization is presented also. By the results it is shown, that the statistical approaches are significantly more reliable yet at the same time strongly dependant on the learning set selection. Even if no significant influence of the learning set on the Gabor wavelets based method can be observed, its successfulness is clearly below those of PCA and LDA.

Keywords: Face recognition, PCA, LDA, Gabor wavelets, Learning set selection

1 Introduction

Although various methods for face recognition have been developed, it remains an important field of research. One of the main reasons certainly lies in high market demands for secure systems based on biometrical identification. Face recognition is recognized as one of the more elegant approaches, since it is user-friendly as well as cost-efficient. At the same time, findings in face recognition research are often applied to industrial projects for the purpose of pattern recognition in general [1]. Whatever the purpose may be, the efficiency is strongly dependent on the detected features and the quality of their representation in the model base [2]. Although many approaches are known for this task [2, 3], features are usually assembled as components of a feature vector [2, 5, 6, 7]. In such cases, each component of a vector caries important information, which is the basis for distinguishing between faces. In a most simple case, features can describe the colour of the human eye, the colour of the skin or the shape of the face, but unfortunately such simple features are usually not sufficient enough. Therefore, statistical techniques are often used to determine adequate features. When the feature extraction is based on statistical attributes of the selected face population, then the final set of features may be very dependent on the subset of faces that were used in the learning process. Because of that it makes sense to study the possible influence on the accuracy of the face recognition algorithms.

In this paper we present a study of the influence of the training set selection on the face recognition accuracy. For this purpose the training set dependency of three algorithms was analyzed: feature graphs based on a wavelet transform, principle component analysis (PCA), and linear discriminant analysis (LDA). The efficiency of recognition techniques was compared in terms of dependency on the learning set of faces.

A detailed review of the image normalization procedure, which ensures robust detection of features and provides equal conditions for testing the efficiency of methods, is given in Section 2. This is followed by a detailed presentation of the used face detection techniques (Section 3). Results, obtained using these procedures are presented in Section 4. The most important conclusions are emphasized in Section 5.

*simon.gangl@uni-mb.si
†domen.mongus@uni-mb.si
‡zalik@uni-mb.si

2 Input image normalization

Main issues that need to be considered when dealing with computer based detection and recognition systems are related to unequal light distribution, camera position, image quality, and image resolution. The capabilities of such systems can be reduced dramatically by these parameters. Therefore, the elimination of these factors needs to be accomplished before the recognition is performed. This process is called image normalization [5, 6] and is, in our case, achieved in four steps:

Step 1 (face detection): The detection of the face is usually the first step of the image normalization process. A neural network is used for this purpose. In the training process the neural network was trained to detect the presence of a face in an image with resolution of 128x128 pixels. Such a neural network is capable of detecting faces only in images with the same resolution. Because the input images are of arbitrary size, face detection cannot be performed directly. Therefore a sliding window is defined. By testing the sliding window region for the presence of a face at each position, faces can be detected. That way all faces in the input image, located at one of the regions sized 128x128 pixels, can be found. Nevertheless, the faces in input images are usually much larger and are not detected at this step. Therefore the input image size has to be reduced multiple times and scanning for faces repeated at each iteration. For this task a sufficient scaling factor has to be chosen, which is a trade-off between execution speed and reliability of detection. In our case a scale factor of 0.9 is used that assures us with 100% face detection ratio on the FERET database [4].

Step 2 (histogram equalization): According to the detected face region, the image is then cropped and thus the background is removed. However, noise, caused by illumination, may still present a disturbing influence. To increase robustness of the face recognition process against that, histogram equalization is performed on the cropped image.

Step 3 (eye detection and rotation of the image): Eye detection is performed during image normalization to increase robustness of the following steps against camera rotation (or rotation of the head) and thus ensures that all faces appear in horizontal position. Similar to the face detection, the detection of eyes is performed with a neural network. The middle points of the eyes are then used to calculate the sufficient angle of rotation θ. The rotation of the image is formally defined by the following equation:

$$
\begin{aligned}
x_1' &= cos(\theta) \cdot (x_1 - x_0) - sin(\theta) \cdot (y_1 - y_0) + x_0\,, \qquad (1)\\
y_1' &= sin(\theta) \cdot (x_1 - x_0) - cos(\theta) \cdot (y_1 - y_0) + y_0\,,
\end{aligned}
$$

where $(x_0,\, y_0)$ is the centre point of the rotation; in our case this is the middle point between the eyes, $(x_1,\, y_1)$ is the pixel that is transformed at the given step, and $(x_1',\, y_1')$ are the transformed coordinates of the pixel.

Step 4 (scaling of the image): To achieve the best possible matching among the normalized images, they are scaled so that the centre points of the eyes are located at the same positions on the normalized images. The scaling factor is defined as the ratio between the desired and observed between eye distance.

When scaling the image is completed, an additional mask is applied to it, so the remaining background factors, like hair for example, are eliminated. Examples of normalized images can be seen in Figure 1.

Figure 1: Normalized images, which are the input to face recognition techniques.

3 Face recognition techniques

The process of image normalization leads us intuitively to the possibility of face recognition by feature graph matching. This is the first presented approach, where features for face representation are selected on human physiognomy basis and represented using a wavelet transformation with Gabor wavelets. In the continuation, two more often used techniques for face recognition are presented also. Both are based on a linear transformation of the image to a feature subspace. These techniques are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

3.1 Feature graph matching based on Gabor wavelet transformation

Because of the image normalization process, the introduction of face recognition according to feature graph matching is relatively straightforward. In our case a modified approach presented in [7] is used for this purpose. The presented approach introduces a vector of wavelet coefficients (jet), which carries the facial features at a given key point. The components of such a jet describe the response to a Gabor wavelet transformation at a given key point. Since the key points are at fixed positions, the structures of the graphs are equal and thus the graph matching can actually be performed only by comparing the jets, using the given metric.

The Gabor wavelet transform is employed here because it is robust against variations in illumination and small changes in phase [7]. In our case 40 different wavelets (5 different frequencies at 8 different orientations) are used. Figure 2 shows the construction of such jets, and their formation to a feature graph.

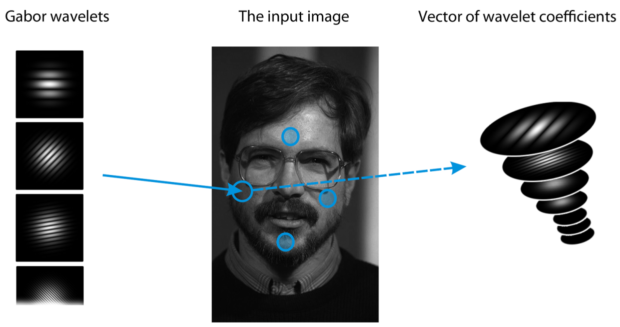

Figure 2: Construction of a vector of wavelet coefficients, where the convolution with the Gabor wavelets is performed at a given point.

The basic difference between our approach and the approach, presented in [7], is that in our case no complete adaptation of the feature graph is needed, since the input images have already been normalized. This way the procedure is much more time efficient, but it also becomes much less flexible. Some important information regarding distances between features is lost, making the recognition less reliable (Section 6).

The facial features are chosen with regard to facial physiognomy [8], where four points, which are important for human facial recognition yet not subject to quick evolution, are chosen. These points are selected at the left and right cheek, on the forehead, and above the chin. At each of the selected points the vectors of wavelet coefficients can now be obtained by calculating the responses to all of the 40 Gabor wavelets. Because even small changes in position can cause a phase shift in the response [7], the wavelet transformation is calculated actually in a 7x7 adjacency of the selected key points. Thus a single face is presented in the model base with 49 feature graphs.

In the process of recognition the jets can now be computed only at the selected key points of the test image. The model base is then searched for the best match to the resulting graph, where the distance between jets is measured with the L^1 or Manhattan metric, defined with the

$$d(\boldsymbol{p},\boldsymbol{q}) = \|\boldsymbol{p}-\boldsymbol{q}\|_1 = \sum_{i=1}^{n} p_i - q_i\,, \tag{2}$$

where $d(\boldsymbol{p}, \boldsymbol{q})$ is the distance between two vectors $\boldsymbol{p}$ and $\boldsymbol{q}$.

3.2 PCA

PCA is a statistical tool for identifying patterns in data. It allows a representation of various sets of data in a way, where similarities between samples are emphasized. Because it is difficult to search for patterns in multidimensional data, PCA is an important tool for data analysis. Nowadays, PCA is present also as one of the most popular approaches for recognition of faces [9, 10, 11] and patterns in general [12, 13]. The implementation of PCA for face recognition can be described in six steps:

Step 1: The inputs of the process are normalized facial images, from which a model database is built. The images are transformed to vectors by dividing them to rows (or columns) which are placed one after another (in our case the images are of dimension 256 x 256, thus each vector has 65.536 components). Each image now represents a base vector of a vector space, with as many dimensions as there are input images. These vectors are formed in a matrix, where each vector represents a column, for a clearer representation.

Step 2: The origin of the vector space is then translated to the point (0, 0, ... , 0) by subtracting the average value of each base vector from its components (the average image intensity is subtracted from its pixels).

Step 3: The dimensionality of the vector space is then decreased by expressing the mutual dependency of the base vectors with a covariance matrix:

$$\boldsymbol{C}_{i,j} = \frac{\sum_{i=1}^{N}(\boldsymbol{x}_i-\bar{x}_i)(\boldsymbol{x}_j-\bar{x}_j)}{(N_{PCA}-1)}\,, \tag{3}$$

where $\boldsymbol{C}_{i,j}$ is the *(i,j)*-th element of the covariance matrix, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are the vectors for which the covariance in the given step is calculated, $\bar{x}_i$ and $\bar{x}_j$ are their average values, which are because of step 2 in our case always 0 and N_{PCA} is the dimensionality of the vectors.

Step 4: The eigenvectors and according eigenvalues of the covariance matrix can then be obtained. Because the eigenvectors represent the interdependency of data, they can be interpreted as facial features in which the patterns from the learning set resemble or differ (Figure 3). Although the obtained vector space allows for face recognition, its efficiency can be increased by discarding eigenvectors corresponding to the highest eigenvalues. These vectors are namely under the influence of illumination distribution and do not resemble valid facial information [11]. In our case, the eigenvectors are sorted descending in terms of their eigenvalues and the first two vectors are discarded. The remaining eigenvectors form the vector subspace $\boldsymbol{E}$, and are presented in matrix form, where each vector represents one column.

Figure 3: The eigenvectors, where the influence of illumination in the input images on the vectors with the biggest eigenvalues can be seen (the upper row), and the vectors with smaller eigenvalues, which represent features (the two bottom rows).

Step 5: In vector space $\boldsymbol{E}$ face recognition can be performed. The base of known faces is created by projecting the input images to the vector space $\boldsymbol{E}$, thus expressing them as a linear combination of the eigenvectors, what can be defined by the following equation:

$$\boldsymbol{y}_i = \boldsymbol{E}^T \cdot \boldsymbol{x}_i \,, \tag{4}$$

where $\boldsymbol{y}_i$ is the projection of the input image $\boldsymbol{x}_i$ to the PCA vector subspace $\boldsymbol{E}$ defined by the reduced matrix of eigenvectors of the covariance matrix $\boldsymbol{C}$.

Step 6: In the process of recognition each input image is projected to the vector subspace $\boldsymbol{E}$ and then compared to the vectors in the model base of known faces using the normalised Euclidean, or the Mahalanobis metric, defined by:

$$d(\boldsymbol{y}_i,\boldsymbol{y}_j) = \sqrt{\sum_{n=1}^{N_{PCA}} \frac{(y_{i_n} - y_{j_n})^2}{\sigma_n^2}} \,, \tag{5}$$

where $d(\boldsymbol{y}_i,\boldsymbol{y}_j)$ is the distance between vectors $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$, σ_i is the standard deviation, which is in our case replaced by the eigenvalue corresponding to the i-th eigenvector.

3.3 LDA

Similar to PCA, also LDA can be used for data classification. LDA is based on maximizing the between-class variance to within-class variance ratio. The most important difference between PCA and LDA is that PCA minimizes the projection error by emphasising similarities between samples; meanwhile LDA defines the classification boarders. Both methods include a projection of data to a subspace, where classification can be performed more accurately. PCA changes the form and location of the input data, while LDA leaves the input data unchanged [14]. In our case LDA is performed globally on the PCA output vectors and this can be described in five steps:

Step 1: The inputs to the LDA process are vectors already projected to the PCA subspace. Because LDA permits many samples belonging to a single class (each person can be presented by multiple images), an additional component is added that defines the class of the vector.

Step 2: The average of each class separately $(\mu_1, \mu_2, ..., \mu_{NR})$ and the average of all classes μ are then computed. The average of all classes is obtained using the following equation:

$$\mu = \sum_{i=1}^{N_R} p_i \mu_i \,, \tag{6}$$

where p_i is the probability of occurrence of a specific class, and can be computed as straightforward as $p_i = 1/N_R$ for all classes, where N_R is the number of all classes.

Step 3: From the data collection two scatter matrices can now be obtained. The scatter matrix $\boldsymbol{S}_w$ describes the expected covariance within each class $R_j; 1 \le j \le N_R$, while the scatter matrix $\boldsymbol{S}_b$ describes the scattering between classes. When many samples of a class $\mathbf{y_i^j}; 1 \le i \le M_j$, exist, the matrix $\boldsymbol{S}_b$ can be understood as a description of covariance between the average vectors μ_j of each class. The equations for calculating the two matrices can be written as:

$$\boldsymbol{S}_w = \sum_{j=1}^{N_R} \sum_{i=1}^{M_j} (y_i^j - \mu_j)(y_i^j - \mu_j)^T \,, \tag{7}$$

$$\boldsymbol{S}_b = \sum_{j=1}^{N_R} (\mu_j - \mu)(\mu_j - \mu)^T \,,$$

Step 4: As already mentioned the LDA optimization criterion is defined as the ratio between $\boldsymbol{S}_w$ and $\boldsymbol{S}_b$. Since Fisher LDA is used, the optimization criterion can be written as:

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T \boldsymbol{S}_b \boldsymbol{W}}{\boldsymbol{W}^T \boldsymbol{S}_w \boldsymbol{W}} \,, \tag{8}$$

where the matrix $\boldsymbol{W}$ is obtained by maximizing the value J($\boldsymbol{W}$). Although this can be achieved by several methods, in our case the ratio $det|\boldsymbol{S}_b|/det|\boldsymbol{S}_w|$ is maximized [15]. It has already been shown that, if $\boldsymbol{S}_w$ is a nonsingular matrix, the ratio is maximized, by forming the columns of $\boldsymbol{W}$ from the eigenvectors of $\boldsymbol{S}_w^{-1}\boldsymbol{S}_b$ [16]. Although in real cases $\boldsymbol{S}_w$ is usually nonsingular, its non-singularity can be ensured by using at least two samples of each class [15]. After that, the matrix $\boldsymbol{W}$ is normalized before it is used in the following procedures.

Step 5: $\boldsymbol{W}$ now represents the vector subspace in which, according to the given set of learning patterns, optimal classification can be performed. Formally the process of recognition in the LDA space can be defined with the following equation:

$$\boldsymbol{y}_i = \boldsymbol{E}^T \cdot \boldsymbol{x}_i\,, \qquad (9)$$
$$z_i = \boldsymbol{W}^T \cdot \boldsymbol{y}_i\,.$$

where $\boldsymbol{E}$ is the vector space of PCA, $\boldsymbol{W}$ is the projection matrix of LDA, $\boldsymbol{y}_i$ the image $\boldsymbol{x}_i$ projected to PCA subspace and z_i the projection of $\boldsymbol{y}_i$ to subspace $\boldsymbol{W}$. The patterns are compared using the Euclidean or L^2 metric, defined by the following equation:

$$d(z_i, z_j) = \sqrt{(z_{i_1} - z_{j_1})^2 + \ldots + (z_{i_{N_{LDA}}} - z_{j_{N_{LDA}}})^2}\,, \quad (10)$$

where $d(z_i, z_j)$ is the distance between N_{LDA}-dimensional vectors z_i and z_j.

4 Measurements and results

Tests were performed on the FERET image database [4]. The FERET image database consists of images of more than one thousand people, taken at different time intervals, with different poses and facial expressions. The presented approaches were tested on three model bases, selected from the FERET base. The first base contains 10 individuals, the second 20 and the third 40 individuals, varying in gender, age, pose, and race. For each individual five images were used for testing, while one images was employed as the learning sample. Because the main interest of our work is the influence of the learning set selection, and not the efficiency of the algorithms, only limited numbers of individuals and only one training image per individual were employed. That way this effect can clearly be studied. Table 1 shows the number of correctly identified samples $\bar{x}$ in percents, while the standard deviation σ describes the class variance from the average efficiency.

Table 1: Successfulness of face recognition with Gabor wavelets, PCA, and LDA.

	Gabor wavelets	PCA	LDA
Base 1: $\bar{x}$	68%	80%	77%
Base 1: σ	0.9940	0.5164	0.4830
Base 2: $\bar{x}$	56%	88%	85%
Base 2: σ	1.0940	0.9987	0.9679
Base 3: $\bar{x}$	56%	82%	79%
Base 3: σ	1.3940	1.1873	1.4118

The first, perhaps a bit surprising, result is that PCA as well as LDA produced better results on base 2 than on base 1, where less testing samples were used. The main reason for this is that both methods can become over-determinate [11, 14], when the learning base contains a smaller number of samples. LDA is especially prone to this effect, since PCA is part of it, and thus produces worse results than PCA alone. Because it is evident that statistical methods of recognition need more learning samples to extract important features, it could be expected that both methods would work even more efficient on base 3. But that is not the case, even an unexpected high decrease in effectiveness can be observed.

To study this effect, several testing sets were employed, created by replacing the learning samples of individuals. By doing that, a high increase of efficiency was noticed when a specific image (see figures 4a and 4b) was not included in the training process. The efficiency of PCA increased to 90% (with standard deviation σ = 1,0671) and that of LDA increased to 87% (with standard deviation σ = 0,9901). In figures 4c and 4d the significant influence of an individual on the entire projection space can be seen clearly. When the specific image of an individual is included in the formation of the projection space, features are accented weakly and are under the influence of illumination distribution (figure 4c). This is evident even on eigenvectors with smaller eigenvalues, although this effect would usually be expected only on some of the eigenvectors with higher eigenvalues (figure 4d).

For the verification of the obtained results, a base of known faces with a higher number of training samples was created. In test base 4 images of 236 persons were included (again only one training image per person was used). The results of our experiments have shown that by applying the same methods as mentioned above, the recognition ratio could be improved by 4%. The relative improvement is this time smaller, which is not entirely unexpected, but the number of additionally recognized images is still a good motivation to observe the influence of the learning set selection also on bases including a higher number of samples.

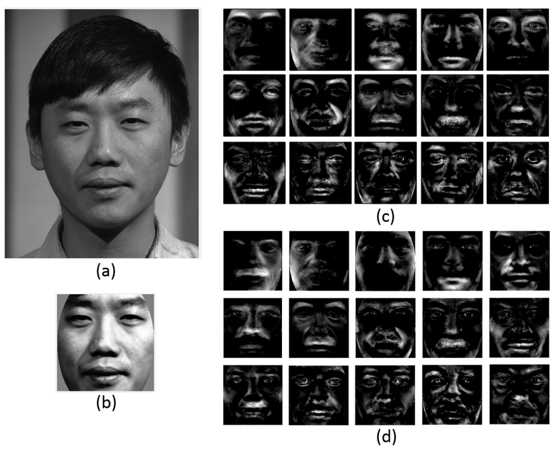

Figure 4: The influence of a learning sample on the eigenvectors, where (a) shows the image of the specific individual, (b) is the normalized form of that image, (c) displays the first 10 eigenvectors when this image was used as a learning sample and (d) are the corresponding eigenvectors when another image of the individual was used.

The selection of the model base has no such particular influence on the recognition method based on Gabor wavelets. The decrease of effectiveness of recognition with increasing number of models in the base is evident, but it is not unexpected. Anyhow, this method produces a relatively unsatisfying result compared to PCA or LDA. The main drawback here is the method for measuring distances between graphs. When using this method, a higher degree of recognition cannot be achieved with simple metrics (like L^1). Because of that, various authors have proposed a use of Gabor wavelet based methods, where the actual comparison is performed with statistically based tools, such as PCA or LDA [17, 18].

Situations, where the tested person is not present in the model base, are often encountered in real-world applications. In such cases the person must be classified as an unknown individual. Because of that, an additional threshold needs to be introduced. If the calculated distance between a sample and its nearest class is greater than the threshold, the given sample is identified as unknown. In our case this threshold is defined as the mean value between the average distance of correctly identified samples and the average distance of a set of unknown samples. For this purpose, additional 100 negative testing samples were included iqnto the previously described test base 3, which contains 200 positive testing samples. The results of this test are shown in Table 2, where the number of correctly identified positive samples is presented as TP (true positives), FN (false negatives) is the number of errors, where a positive sample is recognized as a negative one, FP (false positives) is the number of errors, where a negative sample is recognized as a person from the base and TN (true negatives) represents the number of correctly identified negative samples. The numbers TP and FN sum up to the percentage of correctly identified persons from the test base 3 as shown in Table 1 (thus they represent how many of the previous correctly recognized images are still correctly recognized - TP, and how many are recognized as unknown because of the introduced threshold - FN).

Table 2: Efficiency of face recognition with Gabor wavelets, PCA and LDA, tested on positive and negative samples

	Gabor wavelets	PCA	LDA
TP	50%	80%	76%
TN	81%	96%	94%
FP	19%	4%	6%
FN	6%	2%	3%

From the results shown in Table 2, it can be observed that the introduced threshold does not reduce significantly the efficiency of the presented methods. At the same time, a relatively high percent of negative images is identified. This is most obvious for the PCA and LDA techniques, which again confirms the mentioned fact about the influence of training samples on the efficiency. The distance between a negative sample and its closest class is in most cases significantly greater than the distance from a positive sample to the classification classes. Because of that, there is a higher error rate in recognition of known samples, than of those which are not. Based on the mentioned facts, it can be concluded that the selection of training samples for the creation of the projection subspace is of high importance for the efficiency of PCA and LDA. Consequently, this applies also for methods based on Gabor wavelets, if the actual comparison between features is performed with one of these two techniques.

5 Conclusion

Three approaches to face recognition were presented in this paper; an approach based on the Gabor wavelet transform, PCA, and LDA. Additionally, a method for image normalization, which ensures sufficient conditions for face recognition, was demonstrated. The presented methods were tested on different testing sets with special emphasis on analyzing the influence of learning samples on their efficiency. The first conclusion is that the efficiency of the PCA and LDA techniques improves with an increasing learning set. Because both methods are based on statistical laws, they require a larger set of learning samples that provide high representability. Even further, using PCA or LDA the selection of eigenvectors for the projection subspace formation is of great importance. Some of the eigenvectors associated with the highest eigenvalues namely represent illumination distribution in the learning samples, and it thus makes sense to exclude them. It also makes sense to observe the influence of each learning sample on the efficiency of recognition. It was shown how a single learning sample can noticeably change the projection space and decrease the efficiency of the PCA and LDA techniques. At the same time, it is possible to identify unknown samples reliably by projecting them to the eigenvector subspace. Because these samples were not included in the formation of projection subspace, their features are not emphasized and the distance to the defined classes is noticeably greater. Most of the undesired effects can be omitted using the technique based on the Gabor wavelet transform.

References

[1] C. Liu, H. Wechsler *Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition, pp. 211-216.* Second International Conference on Audio- and Video-based Biometric Person Authentication, Washington, 1999.

[2] R. Jain, R. Kasturi, B. G. Schinck *Machine Vision.* New York: McGraw-Hill, 1995.

[3] H. F. Liau, K. P. Seng, Y. W. Wong, L.-M. Ang *New Parallel Models for Face Recognition, pp. 306-309*. International Conference on Computational Intelligence and Security, Washington: IEEE Computer Society, 2003.

[4] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss *The FERET Evaluation Methodology for face-recognition algorithms, pp. 1090-1103*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22(10), 2000.

[5] Y. Zana, R. M. Cesar *Face recognition based on polar frequency features, pp. 62-82*. ACM Transactions on Applied Perception, vol. 3(1), 2006.

[6] M. Lee, C. H. Park *An efficient image normalization method for face recognition under varying illuminations, pp. 128-133*. Proceeding of the 1st ACM international conference on Multimedia information retrieval, New York: ACM Press, 2008.

[7] L. Wiskott, J.-M. Fellous, N. Krueuger, C. von der Malsburg *Face Recognition by Elastic Bunch Graph Matching, pp. 775-779*. Transactions on Pattern Analysis and Machine Intelligence, vol. 19(7), 1997.

[8] W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips *Face Recognition: A Literature Survey, pp. 399-458*. ACM Computing Surveys, vol. 35(4), 2003.

[9] M. Turk, A. Pentland *Eigenfaces for Recognition, pp. 71-86*. of Cognitive Neurosicence, vol. 3(1), 1991.

[10] A. Pentland, B. Moghaddam, T. Starner *View-Based and Modular Eigenspaces for Face Recognition, pp. 84-91*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle: IEEE Computer Society, 1994.

[11] H. Moon, P.J. Phillips *Computational and Performance aspects of PCA-based Face Recognition Algorithms, pp. 303-321*. Perception, vol. 30, 2001.

[12] A. Ferraz, E. Esposito, R.E. Bruns and N. Durn *The use of principal component analysis (PCA) for pattern recognition in Eucalyptus grandis wood biodegradation experiments, pp. 487-490*. Journal of Microbiology and Biotechnology, vol 14(4), 1998.

[13] R. Saegusa and S. Hashimoto *Pattern Recognition Using a Nonlinear PCA, pp. 73-80*. Proceedings of GVIP 05 Conference, Cairo: CICC, 2005.

[14] W. Zhao, A. Krishnaswamy, R. Chellappa, D.L. Swets, J. Weng *Discriminant Analysis of Principal Components for Face Recognition, pp. 73-85*. Face Recognition: From Theory to Applications, H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie, T.S. Huang (eds.), Berlin: Springer-Verlag, 1998.

[15] A. M. Martinez, A.C. Kak *PCA versus LDA, pp. 228-233*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23(2), 2001.

[16] R.A. Fisher *The Statistical Utilization of Multiple Measurements, pp. 376-386*. Annals of Eugenics, vol. 8, 1938.

[17] C. Liu *Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition, pp. 572-581*. IEEE Transactions on Pattern Analysis and Machine Intelligenc, vol. 26(5), 2004.

[18] W. Li, W. Cheng *Face Recognition Based on Adaptively Weighted Gabor-LDA, pp. 130-134*. Proceedings of the 2008 Fourth International Conference on Natural Computation, Washington: IEEE Computer Society, vol. 4, 2008.