Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Improvement principal component analysis model (PCA) in the process of face tracking using neural networks

Pantea Alipour[1,*], HamidReza Ahmadifar[2]

[1]Rasht Branch, Islamic Azad University, Rasht, Iran
[2]Faculty of Computer, Engineering Department, University of Guilan, Rasht, Iran

ARTICLE INFO

Article history:
Received 10 November 2015
Received in revised form
14 January 2016
Accepted 14 January 2016

Keywords:
Face-detection
Principal component analysis
Neural network
Equalization
Histogram
Logarithm conversion

ABSTRACT

Face-tracking using PCA (Principal Component Analysis) model is a technique which is used for lowering a dimension of an especial space. This technique can make a less dimension space through a series of information which describes better different information of all of them. So, in this research we try to improve the face-detection system through this technique by presenting an optimum model. It has been tried to promote accuracy in face-detection system in this method through presenting levels as pre-procession and also by the use of categorization within Feed Forward neural network and optimization of its parameters. Additionally, it has been presented a method for calculation of basic vectors of covariance matrix in order to approximate this program to the goal of being real time which promotes its calculation time and increase the accuracy of results on databases of Yale and AT&T databases by 15 percent. This research shows that the face-detection system in Yale database with the accuracy of 99.50 % and at AT%T with accuracy of 98% gets output.

1. Introduction

Several researchers act in different fields such as psychology, pattern recognition, neural networks, machine vision and computer graphics. Sometimes a system of face-detection includes a set of different techniques which have originated different rules. Using these different techniques in different parts of face-detection system make it hard to use from a simple division for different methods of face-detection based on techniques which have used for classification and feature extraction. The problem of face-detection by machine in an automatic way is one of the subjects of artificial intelligence field and almost in different branches of this field the goal is to approximate machine decision making to human intelligence. Psychologists believe that human mind may act in two ways for detection of person's face. Paying attention to the totality of person's faces which we want to detect or pay attention to positional features like mouth, nose and eyes. So we divide face-detection methods into three groups here: holistic methods, positional methods and combinational methods (Tan et al., 2006; Schölkopf, 1998).

Due to the importance of this issue in society and different functions of face-detection system, in this research we want to reach accurate results of system with the lowest process and the best available data in picture by the use of PCA model and improvement of its function in order to promote the real time function of system.

Assuming that the input image of image system is in different angles and lighting conditions and have database of images of different persons in opposite angel, we want to detect the identity of intended person.

2. Literature of research

Many studies have been conducted in this area and we point to some of important ones. Every face is defined in holistic methods by a vector with high dimensions that each component of it is indicator of brightness of related pixel. By this method of indication, it should be said that firstly, information of texture and shape of face has been saved for detection and secondly more general aspects towards the description of positional features will be considered. However this method has some disadvantages which are recognized in detection of every image of every man. Firstly the problem of high data and a few numbers of data has been more complex and secondly since there is just one vector for every class, extraction of intra-class features with general methods of pattern recognition is almost impossible. It has been presented two methods in

* Corresponding Author.
Email Address: Pantea.Alipour.Mail@gmail.com (P. Alipour)

order to solve this problem. It should be said that the most possible information is extracted from the only available vector with high dimensions or we can use method of vector lowering such principal component analysis. Second method is the combination of past information in order to make new display for available face and increase the training data set (Blanz and Vetter, 2003; Abdolali and Seyyedsalehi, 2011).

Linear analysis of space which considers the atmosphere of features as a linear combination of some bases has been used vastly in face-detection applications. This is more because of their effectiveness and efficiency of calculations in features extraction and data representation. Different scales produce different bases so that they will have different features under produced atmosphere (Belhumeur et al., 1997).

Linear discriminator analysis method/ Fisher are extension of Eigen faces. After presentation of principal component analysis, the Eigen face method is presented for recognition. Since the support vectors which have been used by PCA have mutual dimensions with the pictures of input image, we call them Eigen faces (Yu and Yang, 2001). Picture number one is an example of average picture and some Eigen faces with big especial amount. When a test image is kept on subspace then Euclidean distance between its coefficient vectors and advocators of every class is calculated (Fig. 1).

Fig. 1: Average image at left side following by seven most famous Eigen faces

When there is lighting and mood changes, most of transformations of data may be related to these changes. PCA selects a method basically subspace which holds the most transformations and necessarily this subspace is not the best space to separate people (Yu and Yang, 2001). The faces method of Fisher finds data in a way that classes in it has much linear potential separable and in other words it increases the space between classes and decreases the space between samples of each class.

In LPP by the use of mapping and through keeping the positional spaces of images, the face will convert into subspace face (Bartlett et al., 2002). The difference between this method with PCA and LDA is that unlike those methods, this method does not concentrate on just Euclidean structure of face space but through keeping positional information, it will find face subspaces in a way that manifold structures of face will be detected properly. In other word, manifold structures will be modeling through nearest neighborhood graph which hold the positional structures of face. This method is based on several databases which are better than principal component analysis or ICA. But this result will attain when we have several educational samples of each person (Scholkopf and Smola, 2002).

Analysis of main component bases on the core is one of the nonlinear methods that Cover hypothesis of nonlinear separable patterns at input subspace can likely be linear separable face if the input space became an especial space with more dimensions. For a nonlinear transformation of φ (Belhumeur et al., 1997), the input space of RN data can be written into feature space with more dimension of F.

$$\varphi: R^N \rightarrow F$$
$$Y \rightarrow \varphi(Y)$$

The first motivation of principal component analysis was based on a basis in which the actions of principal component analysis at feature space with more dimension can gain the statistical properties of higher rank of input variables. In order to solve the problem of calculation cost of dispersion matrix at feature space with high dimension, we can use from nuclear tricks which is done through nuclear function of inner product at input space with lower dimension (Xie and Lam, 2006; Rashidi et al., 2008).

$$K(Y_i, Y_j) = (\varphi(Y_i).\varphi(Y_j))$$

The use of this we can show that the actions of principal component analysis at transformation space are possible by the use of K function. In practical application of face detection (Rashidi et al., 2008), it has been used of three classes of nuclear functions in a vast range which includes core polynomial, Gaussin core and Sigmoid core (Blanz and Vetter, 2003; Lin, 2004).

The methods based on PCA didn't use of the information of total image, won't be strong in front of shape or lighting changes. On the other hand if the images be broken into smaller pieces and vector weighing calculated for each one of these parts, it cause more coefficients to be advocates of image positional information. When the shape or lighting changes is transformed, some of these parts will be changed and the coefficients of these parts will be different from the coefficients of normal image. Also it should be considered that if the images be fractioned into very small pieces, the general information of image will be eradicated and the accuracy of this method will decrease.

A conclusion of conducted accuracy at different databases which are gained at past chores has been shown in Table 1.

3. Suggested method

In this research, we have used from histogram integration in order to set the lighting of image.

Suppose that X is a gray image and its amount is defined in [0-K] interval in which o is indicator of black pixel and k-1 indicates the white color in image and n_i is the number of pixels that have illuminance of i, so the chance of presence of a pixel with i illuminance in image is calculated in the way below:

Table 1: A summary of the last methods accuracy

Paper	Year	Yale	AT&T (ORL)
KPCA Based on the Nonlinear mapping double (Gui et al., 2012)	2006	94%	81.8%
LDA (Shan et al., 2003)	2012	79.83%	82.42%
LPP (Shan et al., 2003)	2012	88.33%	74.28%
Analytic-to-holistic approach (Blanz and Vetter, 2003)	2006	84%	-
Face-specific subspace (Belhumeur et al., 1997)	2003	-	95.3

$$p_x(i) = p(x = i) = \frac{n_i}{n} \qquad 0 \le i \le K-1 \qquad (1)$$

In connection (Eq. 1) the total number of pixels is n and px(i) image histogram for I pixel in a normalized way is at [0-1] interval and so it can be supposed as a random variable. So cumulative distribution function can be written in this way:

$$cdf_x(i) = \sum_{j=0}^{i} p_x(j) \qquad (2)$$

In order to have a uniform histogram, and due to above connections, the following formula is gained for histogram integration for each pixel of V:

$$h(v) = \text{round}\left(\frac{cdf(v) - cdf_{min}}{n - cdf_{min}} \times (l-2)\right) + 1 \qquad (3)$$

Fig. 2 shows that histogram integration.

(A) (B)

Fig. 2: Histogram integration, A. the main image B. after histogram integration

After histogram preprocessing and integration, we extract the features in image. PCA is a linear conversion which is gained by the variance of input data. Coordinate system that data are drawn on it will turn by PCA in a way which becomes the coordinate parallel axis with the direction of most variance at data (Abdi and Williams, 2010). Other coordinate axis should be in line with the parallelism to the greatest variance of data and also all coordinate axis should be perpendicular to the post axis. We can say that the firs axis includes the most amounts of variances and the second axis includes the greatest variance of data and etc. For better understanding, they have been given an example in the components of Fig. 3 that compare new and old axis and also data at two dimensions.

There are different methods for calculating the support vectors of PCA and in this thesis we use from eigenvalue and eigenvectors of covariance matrix.

Suppose that the data are like the connection number (Eq. 4) in which n is the number of data and Γi is the i-th number data with d dimensions.

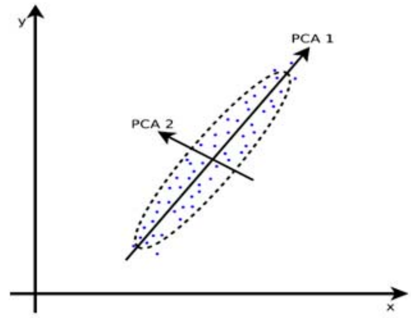

Fig. 3: Principal component analysis (PCA) for two-dimensional data

$$X = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_n\} \qquad (4)$$

The first action we should do for data analysis is to normalize data and show reduction the X average from every data.

$$\Psi = \frac{1}{n} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \Gamma_{\mathbf{i}} \qquad (5)$$

$$\Phi_i = \Gamma_i - \Psi = \{\Gamma_1 - \Psi, \Gamma_2 - \Psi, \ldots, \Gamma_n - \Psi\} \qquad (6)$$

In connection (Eq. 5), we calculated the amount of average of input data and the decrease of average from data in connection (Eq. 6) has been shown in mathematic form.

You can see the average input images at first data base in the Fig. 4.

Fig. 4: Shows the average of first database images

Then principal component analysis will be done on this series of vectors with great size in which we should find the n of orthogonal vector and normalized um that these vectors describe the best mood of data distribution. The k-th vector of uk is chosen in a way in that connection 10 is maximized with connection (Eq. 1).

$$\lambda_k = \frac{1}{n} \sum_{m=1}^{n} (U_k^T \Phi_m)^2 \qquad (7)$$

$$u_l^T u_k = \begin{cases} 1 & \text{if } l = k \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

Covariance matrix also is calculated due to connection (Eq. 9).

$$\Sigma = \frac{1}{n}\sum_{m=1}^{n}\Phi_m \Phi_m^T = \frac{1}{n}\Phi\Phi^T \tag{9}$$

In other words the main axes are equal to especial axis of U which is gained from covariance matrix based on the following connection in matrix method.

$$\Sigma U = U\Lambda \tag{10}$$

In connection (Eq. 10), Λ is a diagonal matrix and U is eigenvectors which is shown below.

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_d \end{bmatrix} \tag{11}$$

$$U = [u_1, u_2, \ldots, u_d] \tag{12}$$

So the especial corresponding vector with the greatest eigenvalue includes the greatest variance of data that is u1.

After calculating the vectors and eigenvalue, i-th input data on the PCA space is written through connection (Eq. 13).

$$\omega_i = U^{-1}(\Gamma_i - \Psi) = U^T(\Gamma_i - \Psi) \tag{13}$$

Since U matrix is orthogonal in connection 16 then$U^{-1}=U^T$.

Finally it is necessary to say that data in new space of PCA of connection (Eq. 14) can be returned to the main space of primary image.

$$\Gamma_i = U\omega_i + \Psi \tag{14}$$

One of the most important functions of PCA is dimension reduction, in a way that if a subset of U choses instead of U, we can decrease data redundancy greatly through reflecting the data on this subset. So we should just use special vectors having bigger eigenvalue. It can be proved in an experimental way that around 90 percent of data is recoverable through few number of greater eigenvector.

Now suppose that past hypothesis include n input data in a d dimension space. Almost all the items in this problem are $n \ll d$. So in PCA the calculation of vectors and eigenvalue of covariance matrix into the dimensions of $d \times d$ is so time-consuming and in some items calculation regarding current systems is impossible. Required time for calculating the eigenvectors will increase through increasing the size of matrix. On the other hand we can calculate simply the corresponding eigenvectors with eigenvalue of matrix with an amount very fewer than $n \times n$ through the connections below.

$$\Sigma = \frac{1}{n}\Phi\Phi^T \tag{15}$$

We can calculate Φ from connection (Eq. 6) in above connection.

Due to connection (Eq. 10) and connection (Eq. 15) we can say that:

$$\Phi\Phi^T U = \lambda U \tag{16}$$

If the parties of connection (Eq. 16) multiply by Φ, the outcome will be the sentence below:

$$\Phi\Phi^T \Phi U = \lambda \Phi U \tag{17}$$

Due to the above relation we can say that in order to calculate the eigenvectors of ΦΦT, at first we should calculate the eigenvectors of ΦTΦ and then multiply them by Φ. For example suppose that there are 100 images at educational database and each size is 300 * 200 pixel and also the dimensions of covariance matrix is 60000 * 60000, so instead of direct calculation of eigenvectors of covariance matrix at first we can easily extract the eigenvectors of a matrix with the size of 100 * 100 and then we can multiply the obtained vectors by images matrix. This method can sharply decline the calculation time and the use of it leads to approximating one of main goals namely right time of system.

In the conducted tests, the average of face detection rate is calculated due to several numbers of features and parameters. Additionally this process in some tests will be repeated in lieu of several numbers of educational data for each class.

At the final level of image detection, after extraction of features, the matrix of educational data will be given to neural network input as educational data. The applied neural network in this test is a two-layer neural network with sigmoid activation function. These two layers include a hidden layer and an output layer. An output layer due to the available persons has 15 neurons at second database and 40 neurons at third database. The selection of the number of neurons of hidden layer due to the mentioned method in article (Liu et al., 2010) is conducted. It is recommended in this article that the primary number of neurons be selected due to connection (Eq. 18). Then we can start the education process by this number of neurons and consider the time of education. After that, we can increase the neurons of hidden layer until the education time is kept constant.

$$n = \log_2 N \tag{18}$$

For considered education data in this thesis, the primary number of hidden neuron equals to 9.81 that output diagram of number of neuron due to the education time has been pointed. This diagram suggests that education time for 41 neurons are about 4 seconds and through increasing the number of neurons, the education time will be around 5 seconds. So the neuron numbers of hidden layer at all tests is considered 41 neurons.

Likewise, the each parameters of neural network will affect the categories and one of the most important of them is gradient amount. This amount of gradient is dependent on algorithm output and the amount of goal and tests results indicate that the lowest amount of gradient will increase the accuracy of categories and also the low amount of gradient leads to the wrong choose in database (Fig. 5).

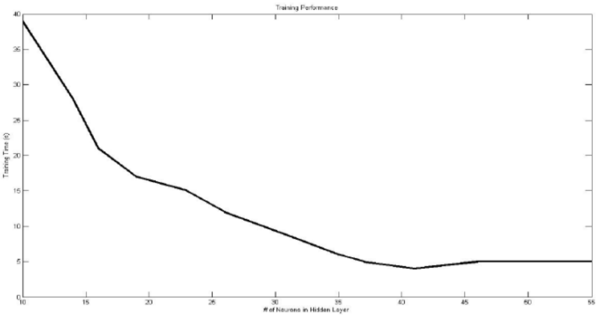

Fig. 5: Comparison diagram of neuron numbers of hidden layer based on education time

The comparison of gradient amounts for errors 1e-6 and 1e-17 for an image of education data at second database has been shown at Table 2. Due to mentioned results in Table 2, in order to increase the accuracy classification we should consider the lowest amount of gradient.

Table 2: The effective amounts of gradient in classification

1e-6	1e-17
0.0000	0.0000
0.0000	0.0000
0.0000	0.0000
0.6032	0.0049
98.0832	99.9925
0.0040	0.0000
0.0982	0.0000
0.1248	0.1210
0.0723	0.0000
0.0000	0.0000
6.9192	0.0004
0.0015	0.0000
0.0001	0.0000
0.0021	0.0000
0.0223	0.0000

4. Analysis

One of the most important challenges in selecting the number of eigenvectors is related to feature extraction. The important issue is how much the number of these eigenvectors should be in order to recover the image with proper accuracy. In the following, you can see the recovered images that have been shown for one of the data images of AT&T database related to 10 to 310 eigenvector. As you can see in Fig. 6, the number of 10 vectors is certainly improper and also 50 numbers of eigenvectors are also enough for extraction of main features. Likewise, there will be a very good accuracy in recovering the images with 300 eigenvectors. Due to these images, we can almost consider the amount of selective eigenvector but this amount is strongly dependent on input data.

At Yale database there are 11 different face images in different situations for every person that due to random selection of data, there can be considered different numbers of images for education and test data each time.

In the following, you can see the accuracy diagram of face detection due to number of eigenvector for feature extraction for Yale database and third database.

In this test 80 % of face images have been considered for education data and 20 % for test data. So, around 8 face images has been considered in random way for education and 2 ones for test of each person. Due to Fig. 3 and 4, we can consider that the amount of eigenvector of feature extraction for every input data is different, for example at second database, the best results have been calculated for 40 eigenvector, while at AT&T database, the best amount of eigenvector equals to 60 (Figs. 7 and 8).

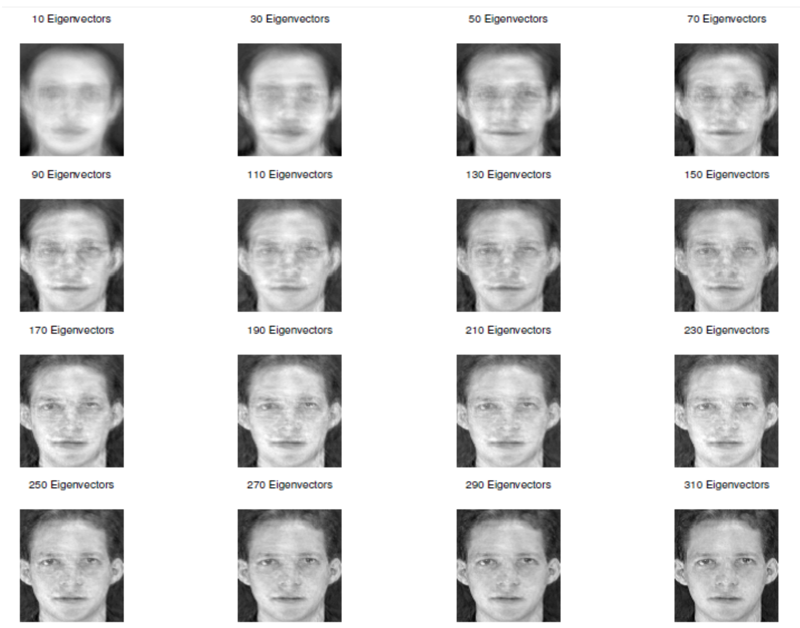

Fig. 6: Recovered images with the use of different amounts of eigenvector

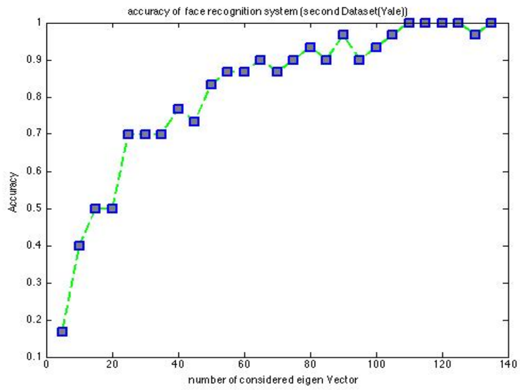

Fig. 7: Comparison of system accuracy of face detection due to the amount of eigenvector for Yale database

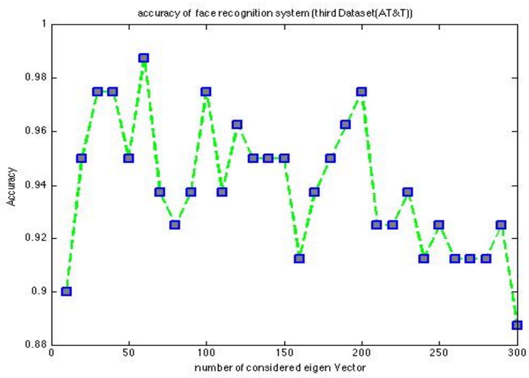

Fig. 8: Comparison of system accuracy of face detection due to the amount of eigenvector for AT&T database

Table 3 comprises the results of face detection of suggestive method with other suggestive methods at second database. This comparison is due to the education data in a way that in each class for each person, there has been chosen different amount of face images for education.

Based on the results of this table, we can see that the results of suggested method are better than other methods among database and as we expect with the increase of education data amounts for each person, algorithm, and accuracy, will increase. In order to consider the generality of suggestive method, we have calculated these results on third database and its results have been shown in Table 4. Also in this table suggestive method is better than other methods. Although one of the goals that mentioned in face detection system in chapter one was robustness of algorithm towards face turning, due to the lack of database and proper data, we are incapable of doing the test of this suggestive method about this subject that can be considered as chores of future. We guess that since we have not considered the rotation criterion in our connections, it is so likely that other methods, which act in this field, gain better results.

Following you can see the output of system implementation of face detection for some sample images from third database (Fig. 9).

Table 3: Comparison of system results of face detection at second database (%)

Amount of training data for each person	3	5	7
KPCA	74.38	76.44	78.67
LPP	80.17	84.11	85.83
LDA	81.02	83.17	85.83
Suggested Method	89.34	91.27	99.58

Table 4: Comparison of system results of face detection at third database (%)

Amount of training data for each person	2	3	4
Eigen face	66.98	77.02	80.92
MDA (Yan et al., 2007)	78.50	88.66	92.44
LDA	70.09	85.70	91.98
Suggested Method	79.61	91.25	94.80

Fig. 9: The output of face detection system for some of images of AT&T database

5. Discussion

In mentioned challenges, the problem we tried to solve in suggestive method is about the face detection of real time system with decreasing the calculation amounts and using the features of holistic methods and simplicity of suggestive method and we almost reach our goal. The other challenge also was low quality of images and improper shooting condition in a way that in some situations there was not enough light for shooting and in addition to blurriness, this led to darkness of image or a part of it that we reached it with the use of effective preprocessing. The other problem is to promote the accuracy classification that we almost reach it by selecting the neural network for classification towards other methods such as SVM etc., and also optimizing each parameters and analyzing them. The results of this research applied on Yale database that its detection accuracy was 99.50 %. Also proposed solution reach 98.75% of detection accuracy at AT&T database that its images are more challenging towards Yale database that this amount of accuracy of introduced and known algorithms are significantly higher. Certainly, system designing and implementation that could solve all challenges and difficulties of this problem is a little far-fetched. In this article, we tried to consider the publicity of problem beside parts of challenges. It seems that in future we can consider more challenges of this issue. For example we can concentrate on learning methods and consider head rotation in that in a way that the shooting angle and can be calculated in addition to 3-dimentional simulation of human and concerned person be found.

6. Results

In this article, the results of suggestive methods in comparison with other methods on two databases of AT&T and Yale were presented and as you saw, the suggestive method in these two databases had higher accuracy. In the following, you can read a conclusion of important points of face detection system:

1. Use of histogram integration and conversion of logarithm as the preprocessing level of suggestive method
2. Use of optimized PCA for feature extraction
3. Use of two-layer Feed Forward neural network for classifying the extracted features
4. Education of neural network by the use of variable amounts of education data on both databases.
5. Calculating the highest accuracy in both databases due to 80 % of education data and 20 % of test data.
6. Taking into account the 41 neurons hidden layer in both databases.
7. Calculating error of 0.5 % at Yale database and 1.25 % of error percent at AT&T database.

References

Abdi H and Williams LJ (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4): 433-459.

Abdolali F and Seyyedsalehi S (2011). Improving pose manifold and virtual images using bidirectional neural networks in face recognition using single image per person. In Artificial Intelligence and Signal Processing (AISP), International Symposium on: 37-42.

Bartlett MS (2001). Independent component representations for face recognition. In Face Image Analysis by Unsupervised, Learning Springer US: 39-67

Belhumeur PN, Hespanha JP and Kriegman DJ (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7): 711-720.

Blanz V (2006). Face recognition based on a 3D morphable model. In Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on: 617-624.

Gui J, Sun Z, Jia W, Hu R, Lei Y and Ji S (2012). Discriminant sparse neighborhood preserving embedding for face recognition. Pattern Recognition, 45(8): 2884-2893.

Liu C (2004). Gabor-based kernel PCA with fractional power polynomial models for face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 26(5): 572-581.

Liu X, Geng G and Wang X (2010). Automatically face detection based on BP neural network and Bayesian decision. In Natural Computation (ICNC), 2010 Sixth International Conference on 3: 1590-1594.

Rashidi KH, Faez K and Gao Y (2008). Face recognition using adaptively weighted patch PZM array from a single exemplar image per person. Pattern Recognition, 41(12): 3799-3812.

Schölkopf B and Smola A J (2002). Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

Schölkopf B, Smola A and Müller KR (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 10(5): 1299-1319.

Shan S, Gao W and Zhao D (2003). Face recognition based on face-specific subspace. International Journal of Imaging Systems and Technology, 13(1): 23-32.

Tan X, Chen S, Zhou ZH and Zhang F (2006). Face recognition from a single image per person: A survey. Pattern Recognition, 39(9): 1725-1745.

Xie X and Lam KM (2006). Gabor-based kernel PCA with doubly nonlinear mapping for face recognition with a single face image. Image Processing, IEEE Transactions on, 15(9): 2481-2492.

Yan S, Xu D, Yang Q, Zhang L, Tang X and Zhang HJ (2007). Multilinear discriminant analysis for face recognition. Image Processing, IEEE Transactions on, 16(1): 212-220.

Yu H and Yang J (2001). A direct LDA algorithm for high-dimensional data—with application to face recognition. Pattern recognition, 34(10): 2067-2070.