An overview of face recognition methods

Mohammed Jawad AL_dujaili1, Haidar Zaeer Dhaam2 and Ahmed Al-Fatlawi3*

1Department of Electronic and Communication, Faculty of Engineering, University of Kufa, Iraq
2Department of Laser and Optoelectronics Technical Engineering, Technical Engineering College/Najaf, Al-Furat Al-Awsat Technical University (ATU), Najaf, Iraq
3Computer Technical Engineering Dept., Technical Engineering College, University of Alkafeel, Najaf, Iraq

Abstract. Face recognition is one of the most important parts of artificial intelligence and machine vision systems. In a face recognition system, the face is identified in a search database, and the face closest to it is identified. Hence, the identity of the person can be determined. There are many applications for face recognition systems, the most important of which is authentication. In this article, we select the feature extraction method and classification from different face recognition methods and review its different methods.

1 Introduction

Face recognition is an important requirement for many intelligent systems. Face recognition is one of the basic areas in the field of biometrics and security in the last few decades, which has high accuracy of identification and unlike other validation methods, which requires the user to remember at least his username and password, it can be used in the validation process. Research on how humans perceive faces has resulted in many interesting findings that can be a useful help in designing practical and practical systems [1]. One of the topics that has attracted the attention of researchers and scientists today is the use of image processing in various sciences. The traces of image processing can be seen in many engineering sciences and industries, and some of these applications are so dependent on image processing that they fail to fulfill their goals without it. The application of image processing in the field has become very widespread, among the applications that have attracted the attention of many researchers in recent years is the use of image processing techniques in face recognition. Research in the field of facial recognition provides diverse and extensive study opportunities that will draw scientists and engineers to serious challenges for the next few years. Face recognition is widely used in biometric systems [2,3]. For example, the creation of a powerful facial recognition system can be used in related projects for national security, human-computer interaction, and many other cases. The operation of a face recognition system by an automatic machine is as follows. With the entry of given still or moving images from an environment, the person or persons whose images are available in a specific database must be reviewed or identified. By doing the pre-processing process on the input image, we separate the face images from the crowded and complex or simple scene. Extracting the features of the face image from the range of the face will reduce searches and, as a result, increase the efficiency of review and recognition; Face recognition techniques include two main methods, which are appearance-based methods and feature-based methods. In the feature-based method, they directly use human information about the face structure[4]. The appearance-based method considers the face recognition problem as a two-class pattern recognition problem, which is based on statistical learning. In recent years, different algorithms have been proposed for face recognition with high accuracy. Some algorithms have strengths and weaknesses as well. Considering these things, in this article, while studying the face recognition system, some proposed methods for face recognition in digital images are evaluated. There are several methods for face recognition systems. One of these methods is face recognition based on feature extraction and classification. In such a system, facial features are first extracted using one of the available methods. Then, using a trained classifier, the person's face is classified and identified. In this article, we will review these methods. Figure 1 shows the outline of such a system [5].

* Corresponding author : ahmed.fatlawi@alkafeel.edu.iq

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
2 Related works

In recent years, many studies have been conducted in the field of face recognition using different techniques. Facial pattern recognition is considered as one of the intelligent computer analysis modes. There are always improvements in this field, which lead to high accuracy and precision in face recognition.

The first work on automatic face recognition was presented by Wilkes et al. [6] in 1992 to recognize the image taken from a mobile phone camera with low quality [7]. Kim et al. [8] have designed an algorithm in which feature extraction is done using local and global features. These features are coded using a quantization method based on the identity of persons and finally used to improve the results of a communication feedback scheme that uses several reference images to retrieve and sort the images.

In [9], a method called chain LDA is introduced, which consists of two LDA stages. In the first step, several sub-images of the selected face image are displayed using LDA. LDA coefficients are also calculated for the whole image. Then the coefficients obtained from the sub-images are applied to this set of features with other LDA coefficients to obtain a compact description of the face.

Milki et al. [10] in 2020 introduced a method based on deep learning for face recognition in an uncontrolled environment. The proposed method involves creating a deep neural network architecture by combining global and local features at different scales. The architecture consists of two main networks: The first one is a region proposal network that creates a list of regions of interest, and the second is related to a network that uses these points of interest for classification. The obtained results have proved the efficiency and also the feasibility of the proposed method to deal with multi-scale face recognition problems.

Dong et al. [11] in 2020, based on the theory of deep learning and neural network, have proposed a method for face recognition. The proposed method creates a deep neural network by creating a table of edge features, contour features, local features, and semantic features. According to the latest semantic features, input data information is obtained to accurately realize face recognition in non-ideal conditions. The simulation results show that this algorithm performs well for recognizing one and several faces and has high robustness for face rotation. At the same time, the algorithm is fast and can basically meet the requirements of real-time face recognition.

Lu et al. [12] in 2019, used the combined features based on the Viola-Jones algorithm to improve the accuracy of face recognition. In the proposed method, first the features are extracted by the Viola Jones algorithm, then they enter the classification phase for face recognition. The implementation results indicate a high rate of recognition accuracy compared to other face recognition algorithms.

In 2019, Kirana et al. [13] proposed a facial recognition system based on the Ant Colony Algorithm. In the proposed method, the ant colony algorithm is used in order to avoid exploring non-face areas in the image. The proposed method has less computing time than other methods and also shows higher detection accuracy than other methods.

In 2018, Yi et al. [14] proposed a new method based on Multi-stage Decision based on skin color and geometry features for face recognition. In addition, face recognition algorithms in natural images have been deeply analyzed in this study. The results of the proposed method showed a detection accuracy rate of 94% and also the proposed method has a lower false positive rate than traditional methods.

In 2017, Jun et al. [15] proposed a method for face recognition combination based on local binary model. In the proposed method, when the number of skin pixels is greater than the determined threshold, it is determined as a candidate face area, otherwise it is a non-face area. For the candidate region, local binary model detection is performed to obtain the real face. The experimental results show that through the local binary model stage, the rate of false diagnosis is significantly reduced.
3 Algorithms of face recognition

3.1 Extract face recognition features

3.1.1 Particle Swarm Optimization (PSO)

The PSO (Particle Swarm optimization) algorithm is a new method of optimizing random search. This algorithm was first proposed by Kennedy and Eberhart in 1995 [16]. This algorithm originates from observing the social behavior of birds, fish and bees. In the PSO algorithm, several particles are generated as the initial answers in the search space. The motion of these particles in the search space is a random motion influenced by their own experience and that of other particles. In other words, the motion vector of each particle is determined randomly from the result of the motion in the direction of the best place it has been before (LBest) and the best place previously observed in all the particles (gBest). To formulate the PSO algorithm, the two variables x and v are defined as particle position and velocity, respectively. The best particle position (in terms of suitability in the objective function) is known as pBest, and the best particle position in the whole group is known as gBest [17,18]. The modified relations of velocity and position of the particle with respect to the coefficients of contraction can be written as follows:

\[v_{d+1} = \alpha \cdot (w \cdot v_d + \varphi_1 \cdot r_{and} \cdot (p_{best} - x_d) + \varphi_2 \cdot r_{and} \cdot (g_{best} - x_d)) \]  

(1)

\[x_{d+1} = x_d + v_{d+1} \]  

(2)

d Repetition counter, \(x_d\) Particle position in iteration \(d\), \(x_{d+1}\) Particle position in iteration \(d + 1\), \(v_d\) Particle velocity in iteration \(d\), weight of inertia and \(\varphi_2, \varphi_1\) are called acceleration coefficients [68]. In relation (1) \(r_{and}\) is a function of generating a random number with a uniform distribution in the range \([0,1]\). \(\alpha\) is a function of, \(\varphi_2, \varphi_1\) proposed by Clerc to enhance convergence and stability of PSO [19]. In each iteration of the PSO algorithm, using the system decision variables is performed by the MODSIMP model and by determining the optimal values of the variables, which include the normal digit, the minimum operating number and installation capacity in each dam, and according to the value of the objective function and new values. Variables are determined using Equations (1), (2). This process continues until the maximum number of iterations or the minimum value obtained for the objective function is repeated in a number of consecutive iterations.

3.1.2 Independent Component Analysis (ICA)

There are different types of evolutionary algorithms. One of them is ICA. So The Colonial Competition Algorithm is an evolutionary algorithm from the group of meta-discovery algorithms. Transcendental algorithms are a type of approximate optimization algorithms that can escape local ambushes and can be used in many problems. The theory of this algorithm is based on socio-political policies. This algorithm starts with a number of random populations. Each of these populations is named (country). In fact, these countries are the same chromosomes in the genetic algorithm. Each country consists of a set of variables. These variables can be in the real world of culture, language, economy, etc. These countries are divided into two categories based on the cost function of the problem: colony and colonizer. First, the number of colonizing countries is determined. To determine the number of colonies of each colonial country, their power must be calculated because this number depends on their power. A key step of this algorithm is the policy of assimilation. In the policy of assimilation of the colonists by pursuing a policy of absorption, it causes the colonies to approach them in different socio-political dimensions. The absorption policy in this algorithm is modeled in the form of a colony movement towards the colonizers. Sometimes it is possible for the colony to move towards the colonizer to a position that in a new position has a lower cost function than the colonizer. In this case, the colonizer and the colony change places with each other. Now the algorithm continues to move with the new colonial country, and the new colonizer begins to engage in policy-making. After creating an empire at the expense of each empire, the total power of each is gained. The cost of the whole empire is in fact the cost of the colonizer plus a percentage of the cost of all its colonies. The second major step is the colonial competition algorithm. In colonial rivalry, every empire tries to increase or at least prevent its decline. In this competition, strong empires will become stronger and weaker empires will gradually weaken, until they are eventually eliminated. To model this fact, strong emperors compete for the weakest colony of the weakest empire. Given the power of empires, the empire that has more power to gain is more likely to seize a weak colony. In this way, by repeating the algorithm, the weak empire will lose its weak colonies. When the empire loses all its colonies, it no longer has the power to compete, it falls and is eliminated from the empires. The algorithm continues in this way until the condition is terminated or the total number of iterations expires. Gradually, with the implementation of the algorithm, weak emperors will fall until there will be an empire. In this new world, colonies and colonizers have the same costs and opportunities, and in fact there is no difference between them [20,21].
3.1.3 Discrete cosine Transform (DCT)

Converting an image to a signal is a good strategy for image processing. Discrete cosine transform (DCT) is a very popular conversion that converts an image from a spatial domain to a signal domain. This conversion has been widely used in recent studies as a feature extractor. DCT reduces the torque required to display information in an image. This is done by removing the redundancy between the values of the neighboring pixels. DCT An image basically consists of low, medium, and high n frequencies, and each of these three bands contains detailed information about an image. Low frequency is generally the average brightness of an image and is commonly used for feature extraction in research. In general, the mathematical equivalent.

\[
F(u, v) = \frac{r^2}{\sqrt{MN}} \left[ \frac{m}{2N} (2l + 1) \right] \left[ \frac{n}{2M} (2l + 1) \right]
\]

A(i), A(j) = \frac{1}{\sqrt{2}} \text{ for } u, v = 0, 2 \text{ otherwise}

In Equation 3 (M * N) is the image size, P (i, j) is the pixel brightness intensity with coordinates j, i and F (u, v) DCT coefficient with coordinates (u, v) and as a result of vectorization with dimensions MN An image will be created with M * N resolution. In cosine conversion, the original image is divided into small blocks of pixels, such as 8 * 8, 6 * 6, or 10 * 10. Then the cosine conversion is performed independently on each of the blocks. It is created. The next step in the quantization process is the DCT n * n coefficients. Finally, it sorts the quantized coefficients using the zigzag scan function. The results of the DCT conversion with a number of K pixels are a DC coefficient and (K-1) an AC coefficient. The low-frequency DC coefficient is located in the upper left corner, indicating the average energy (brightness) of each pixel, which has the most information and the highest light intensity [22-24].

3.1.4 Histogram of Oriented Gradients (HOG)

Histograms of directional gradients have been widely used as powerful visual descriptors in many machine vision applications related to object recognition. And it has been widely used to extract local features of the face. The horizontal and vertical gradients of the input image are calculated and the image area is divided into CxCy grid cells. HOG feature extraction includes from Cutting a part of the image that is supposed to obtain the characteristics of that part of the image and then it's done the dividing the pixels of the image by 255 and placing the brightness values in the range (0,1) finally, the calculation is done the gradient by applying the filters [25,26].

3.1.5 Gabor filters

The next plan is Gabor filters. One of the most successful face recognition methods is based on matching graphs of coefficients obtained from Gabor filter responses [27]. The Gabor filter is directly connected to the Gabor wavelet. The wavelets combine with an unknown wave to obtain information from that unknown wave. Gabor filter is available in two types of one-dimensional and two-dimensional, the one-dimensional mode can be used as an excellent intermediate filter to process one-dimensional signals such as speech. Two-dimensional Gabor filter is widely used in detection and detection of facial area [28]. The Gabor filters shown in Equation (4) are obtained by combining an exponential function with a Gaussian function. It is proved that the Gabor wavelet has the highest correlation between the time domain display and the signal frequency domain display. In addition, these filters have the same display in terms of time and frequency. Gabor filters used in image processing are two-dimensional and each filter has a specific frequency and direction [29].

Images of older people or men with beautiful faces or low quality images are usually not smooth and have very delicate texture features. However, extracting these features is very difficult due to their delicacy, so the most widely used wave analysis techniques, the Gabor filter, are used to extract the feature from the area of the face in question. The general form of the Gabor filter is equation 4.

\[
G(x, y, \theta, u, \sigma) = \frac{1}{2\mu^2} \exp \left( -\frac{(x^2 + y^2)}{2\mu^2} \right) \ast \exp \left( 2\pi i (u x \cos \theta + u y \sin \theta) \right)
\]

In the above equation u, 1 = \sqrt{-1} the sine wave frequency, \( \theta \) is the controller of the function and \( \sigma \) is the Gaussian coverage standard deviation, in fact the Gabor function G (x, y, \theta, u, \sigma) with a The set of specific parameters for \( \theta, u, \sigma \) is converted to the discrete Gabor filter \( \hat{G}^* \) [x, y, \theta, u, \sigma] shown in Equation 5.

\[
\hat{G}[x, y, \theta, u, \sigma] = \frac{\sum_{j=-n}^{n} \sum_{i=-n}^{n} G[x, y, \theta, u, \sigma]}{(2n + 1)^2}
\]

In this regard \( (2n + 1)^2 \) indicates the size of the filter. The Gabor filter is canonized with face images and the results are stored in feature format. The results obtained from the Gabor filter are useful when the three parameters \( \theta, u, \sigma \) are selected correctly. To get the right parameters, the results must be tested on the images several times.
3.1.6 Principal Component Analysis (PCA)

The issue of dimensional decrease bears high importance in many applications (including categorization), because there are many features in these applications, many of which are either unused or they do not have much information load and are difficult to transfer storage or processing [30]. The PCA method (analysis of the main components) is the best way to reduce the data dimension linearly; i.e. the loss of information is less than other methods by removing the insignificant coefficients obtained from this conversion. The analysis of the main component is a linear transformation including \( y = w \cdot x \) in which, \( x \in \mathbb{R}^n \), \( y \in \mathbb{R}^m \), and \( w \) is the matrix of the \( m \times n \) order. Feature of this conversion is that it is optimal in terms of maintaining the maximum amount of information in the data, and its direction is in the way that we have the maximum scatter in the data. The PCA algorithm is as follows [31].

Assume that the data sum of \( D = \{ x_i | x_i \in \mathbb{R}^n \} \) exists as a matrix of \( X \in \mathbb{R}^{N \times n} \). In this matrix, \( N \) is the number of data and \( n \) is the dimension of each data.

1. First it is essential to normalize the data. This is because the domain of one feature may be much larger than the other feature and affect the final result. To do this, the data becomes normalized as follows:

\[
X_{\text{new}} = \frac{X_{\text{new}} - \min}{\max - \min}
\]  

(6)

2. Then, you need to compute the covariance matrix to see if the mean of each dimension of the data is as follows:

\[
\mu(j) = \frac{1}{N} \sum_{i=1}^{N} x_{ij}
\]

(7)

The result is \( j = 1, 2, \ldots, n \) and the average vector is defined as follows:

\[
\mu = \mu_1, \mu_2, k, \mu_n
\]

(8)

The variance of each dimension is also computed as follows:

\[
\sigma_j^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_{ij} - \mu_j)^2
\]

(9)

Ultimately, the covariance matrix is an \( n \times n \) matrix obtained as follows:

\[
\text{cov} = \begin{bmatrix} \sigma_{ij} \end{bmatrix}_{n \times n}
\]

(10)

\[
\sigma_{ij} = \frac{1}{N-1} \sum_{k=1}^{N} (x_{ik} - \mu_j)(x_{kj} - \mu_j)
\]

(11)

3. Compute the specific values and specific vectors corresponding to the covariance matrix of the previous step. Assuming that the corresponding specific values and vectors are as follows (specific values must be sorted in descending form)

\[
W = \begin{bmatrix} V_1 \n V_2 \n \vdots \n V_m \end{bmatrix} \begin{bmatrix} \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \end{bmatrix}, \quad V_1, V_2, \ldots, V_n
\]

(12)

4. We form the matrix was follows.

5. The decreased dimensional data is computed as follows.

\[
Y = X \cdot WT
\]

Have the highest dispersion. The figure (2) is shown for two-dimensional data.

Fig. 2. PCA method
3.1.7 Scale-invariant feature transform (SIFT)

The SIFT algorithm was invented in 1999 by David Lowe [32]. SIFT is a method for detecting and extracting key feature points from images that can be used for applications such as matching between images, object identification, 3D reconstruction of scenes, etc [33].

3.2 Classification

After extracting the features of each face, it is time to categorize the data. In this section we will mention some of the available data classifications. And the strengths and weaknesses of each will be pointed out.

3.2.1 Support Vector Machines (SVM)

Due to its simplicity, the support vector machine model has been considered in various tasks in the last decade. This algorithm was first presented by a researcher named Fisher in 1963 with the aim of reducing the classification error of training data [34]. Its evolution was also presented in 1995 by Vladimir Vepniak Bar, based on the statistical theory of learning the support vector machine problem. This algorithm is widely used in topics such as regression, classification, clustering and in general approximation of functions. The support vector machine algorithm initially included only two-class classification. This approach was then extended to multi-class classification by using various combination techniques [35,36]. Figure (3) shows The SVM classifier.

3.2.2 Genetic algorithm

Inspired by nature to design an algorithm is an interesting idea. Genetic algorithms, which are one of the types of evolutionary algorithms, with possible solutions to the problem in the form of chromosomes, form a population that is constantly replaced by another population and moves towards evolution. A chromosome in the population is assigned a value to indicate the success rate of that chromosome in solving a problem called the degree of compliance [37]. Chromosomes with more potential will have a better chance of producing offspring and therefore will be produced with higher adaptation rates after the emergence of several offspring. Generation in the discussion of genetic algorithms is done by the operators of selection, hembrriage and mutation. The selection operator uses the matching function to select the chromosomes that have the highest matching rate. Selected chromosomes for childbearing are provided to the Hambrian operator and, if successful, reproduce, creating two new chromosomes. The mutation vector is then applied to a small number of the baby's chromosomes, replacing some of their genes with random values. Eventually a new generation of chromosomes is created and replaces the current generation. This evolution continues until the maximum rate of adaptation of the current population is sufficiently close to the optimal state of the problem [38,39].

3.2.3 Classification method of neural networks Artificial neural network (ANN)

In this section, we introduce the types of neural networks used in the research. The studied networks are Multi Layer Perceptron (MLP), Radial Basis Function (RBF).

3.2.4 Multi Layer Perceptron (MLP)

The multi-layer perceptron is a powerful classification tool and is one of the first attempts to build intelligent systems similar to the human brain system. This network includes multiple layers, hidden input and output, and the number of
cells in the first layer depends on the number of selected features. And the number of hidden layer cells can also be selected by the designer. The number of output cells also depends on the number of people in the face bank or the number of classes. The initial value of the weights is considered randomly. This network is trained based on the error backpropagation algorithm, so that the outputs of the network are compared with the desired outputs and the weights are adjusted by the error backpropagation algorithm with the observer to create a suitable pattern for the input pattern [40,41].

3.2.5 Radial Basis Function (RBF)

Neural networks have emerged as an applied technology that is successfully used in a variety of fields [42]. In the neural network algorithm with the radial basis function, one neuron was added to the middle layer of the network in each iteration. If the network error was high, another neuron was added to the network and the error related to the new network was checked again. If the error was small enough, the algorithm would terminate. Otherwise another new neuron was added and this procedure was repeated until the algorithm reached the specified error value or the number of network neurons reached its maximum specified value. The scattering coefficient determines the width of the range to which the neuron responds. The scattering coefficient in the training algorithm had to be selected large enough so that the neuron with the radial base function was able to create a suitable reaction on the overlapping areas of the input space. Of course, if this factor was chosen too large, all the neurons in the network would behave in the same way, which had an adverse effect on the training process. Finally, after dividing the data into two sets of training and experimental data, the network was trained using the optimal scattering coefficient. For neural network training with radial basis function, the data were classified into two sets of training and experiments [43].

3.2.6 Deep Neural Networks (DNN)

The deep neural network is inspired by the human brain [44]. To train deep neural networks, researchers put several layers of neural networks in a row. Although research on deep architecture has been around for a long time, no successful methods were reported until 2006 due to the difficulty of teaching deep networks. In 2006, Hinton and colleagues introduced deep belief networks and, using greedy learning algorithms, taught one layer at a time, using an unattended learning algorithm for each layer [45]. These networks are data-driven and feature extraction is done automatically. This has led to their very good accuracy and excellent performance in various fields. As networks deepen, their capabilities increase, but the lack of sufficient data, the lack of suitable and strong hardware, as well as severe over-fitting, plus the problem of signal transmission from the end to the beginning of a deep network, all make it impossible to use. These problems have been resolved in recent years, and this has led to a resurgence of the debate over deep architecture [46].

3.2.7 Random Forest Algorithm

The stochastic forest algorithm is a bagging method consisting of decision tree classifications. The accuracy of the random forest classification method has been significantly improved by constructing a set of trees and voting between them to obtain a category with a higher number of votes. Each classifier for each input instance is \( h(x, \emptyset K) \), where \( x \) is an input instance and \( \emptyset K \) is the instruction set for the km tree. \( \emptyset K \) are independent of each other but with the same distribution. For each instance \( x \) each the tree provides a prediction for the sample category \( x \), and finally the category with the highest number of votes on the input \( x \) is selected as the sample category, a process called random forest [47].

The stochastic forest algorithm has only one learning technique for the samples. With different sampling, along with the replacement of the new training set with the same initial size, a separate decision tree can be trained from each. Random warfare has shown in practice that very good results are obtained, especially when the size of the data set is small. In this technique, the main data set uses the sampling technique with localization and is divided into a number of datasets [48].

3.2.8 K_Nearest Neighbor (K_NN)

One of the most popular classification algorithms; the algorithm is the nearest neighbor. The knn algorithm became popular as the computing power of computers increased, and one of its common uses is pattern recognition. The K-means clustering method, while simple, is a very practical method and the basis of several other clustering methods such as fuzzy clustering. The method is that first the desired number of points is considered as the center of the cluster, then by examining each data, that data is attributed to the nearest center of the cluster. After completing this task, by taking the average in each cluster, cluster centers can be created and then new clusters. This process continues until there are no more changes in the clusters. Figure (4) shows The K_NN classifier.
3.2.9 Decision Tree Method (DT)

Classification of decision trees is a set of sequential rules that ultimately lead to decision making. Unlike numerical techniques, they have the ability to interpret and use the strategy of division and domination. Classification using the decision tree is such that the classification starts from the root, which asks the specific value of a characteristic from the sample. This root (attribute) is connected to its different values. Depending on how much the sample has for this attribute, it enters another node. Only one mode must be selected. At this stage, the data characteristic of the input node of the input node corresponding to the second level of the tree must be decided (in return, the root under the tree can also be found). These steps are repeated until the data reach the leaves of the tree [51].

3.2.10 Linear Discriminant Analysis (LDA)

The LDA Linear Discriminant Analysis or LDA method is one of the best methods for simultaneous dimension reduction and categorization due to the simplicity of the algorithm; In practice, however, it has an important weakness. In the LDA algorithm, scatter matrices are very important and are key elements in the calculations. In general, three scattering matrices are defined in this method, which are considered invertible by default in the definition of the algorithm, but it is observed that in some cases these matrices have zero determinants, in other words, the singularity of these matrices causes This method cannot be used in many cases. One of the suggested ways to solve this problem is to use LDA + PCA. Other methods include the use of false inverse matrices as an alternative to the standard inverse matrix definition [52,53].

4 Conclusion

Face recognition has many applications in machine vision and artificial intelligence. There are several ways to make an accurate and fast face diagnosis. In this paper, we review the methods of classification and classification. In this paper, after a brief introduction, we reviewed the feature extraction methods. Then, in the next section, various classification methods were reviewed. Examining this approach in face recognition can be a good incentive for future work in this field.

Reference


