Different Approaches to Face Recognition

Vanlalhruaia, R. Chawngsangpuii
Assistant Professor,
Department of Information Technology,
Mizoram University

Yumnam Kirani Singh
Senior Engineer,
C-DAC, Silchar

Abstract— **Biometrics recognition system aims at providing security to a level where a person's identity and authentication is indispensable. Face Recognition, among others, received more attention due to its simple and non-intrusive method of sample collection. Collection of sample data can be easily done even without the direct participation and voluntary action from the subject which is not the case of its counterpart like fingerprint and iris recognition system. Recognizing a face from a face database is a huge challenge as captured image from the image database has different modalities due to variation of its environment. This paper present literature review of popular face recognition technique and discuss the methodology and its functioning. An evaluation of certain recognition techniques is also given.**

Keywords— Face Recognition; Eigenfaces; Fisherfaces; Neural Network; Elastic Bunch Method; Graph Matching; Hidden Markov Model; Support Vector Machine.

I. INTRODUCTION

Unique identification of a person has become an important area of studies as it finds wide range of commercial and law enforcement applications. Unique identification is done by applying various biometric recognition systems using features like fingerprint, hand geometry, palm print, iris, face, voice, gaits, handwriting, signature and even odor. These features have their specific advantages and disadvantages in regarding to the applications and requirements. One recognition system that works well for a particular requirement may not be applicable to others. Among these feature recognition system, face recognition also finds an importance from the past few years spanning numerous fields. Face recognition is necessary not only because it finds several practical applications, but also as it is a fundamental human behaviour for expressing emotions [1]. The use of face as a feature for recognition system also has advantages in law enforcement purpose, as it does not require active participation from the person and as large data sample collection is achievable in a convenient manner.

Face recognition is carried out in the basic steps of face detection, feature extraction and classification. Face detection is carried out to find the face in the image and whether face is actually present in the image or not. If a face is detected, it returns the location of the image and extent the face region. Pre-processing is done to remove the noise present in the image captured. Then the input image known as probe is compared with the database called gallery. Classification is done to identify the population to which the new observed belong. Problems of face recognition are: changing facial expression, illumination variation, aging, pose variation,

scaling factor (image size), frontal vs. profile, presence and absence of spectacles, beards, moustaches etc., occlusion and obstacle while capturing. These problems needs to be addressed to have high rate of recognition as capturing of face images are more likely to be crucial in uncontrolled environment.
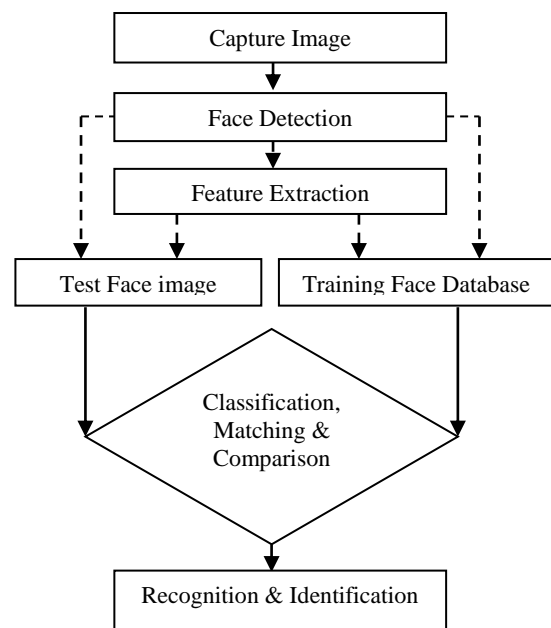

Fig 1: Block diagram of basic face recognition technique.

II. FACE RECOGNITION TECHNIQUES

Many techniques applying different methods have emerged for pattern recognition in general. Such methods are also implemented for face recognition. This section gives an overview on the major human face recognition techniques that apply mostly to frontal faces, giving advantages and disadvantages of each method. Apart from the methods mentioned below, hybrid approach by combining any of these techniques is also intensely research.

A. Principal Component Analysis

Principal Component Analysis is one of the most famous holistic approach i.e. appearance-based methods for face recognition. PCA is a dimensionality reduction technique used for compression and face recognition problems. Sirovich and Kirby [2] used principal component analysis (PCA) in their technique to represent a face images. The face images are decomposed into various collections of eigenfaces and are reconstructed by applying weights to each of these eigenfaces

known as the eigenvalues. Turk and Pentland [3] further carried out the findings of Sirovich and Kirby's and realized that projections along eigenpictures could be used as classification features to recognize faces. Problem arises when performing recognition in a high-dimensional space, which can be improved upon by mapping the data into a lower-dimensional sub-space. The goal of PCA is to reduce the dimensionality of the data while retaining the variation of the data set as much as possible.

Eigenfaces are the principal component analysis of the distribution of faces, or eigenvector that is the covariance matrix of the set of face images. Eigenvector are used to quantify the variations between multiple faces. Linear combination of eigenfaces can represent each face by applying the exact eigenvalues for reconstruction. Face can be represented by using eigenvector having the largest eigenvalues. The best M eigenfaces define an M dimensional space, which is called the "face space".

First, training face images is obtained and *labeled I_1, I_2,…, I_M*. These images are represented as a vector Γ_i. Average face vector Ψ is computed

$$\Psi = \frac{1}{M} \sum_{i=1}^{M} \Gamma_i \qquad (1)$$

and the mean face is obtained by applying $\Phi_i = \Gamma_i - \Psi$. Covariance matrix C is obtained by getting C= AA^T where A=$[\Phi_1 \ \Phi_2 \ …\Phi_M]$. However, AA^T is very large and impractical. The eigenvectors v_i of A^TA is computed where $A^TAv_i = \mu_i v_i$. A^TA can have up to M eigenvalues and eigenvectors. K eigenvectors, which is the best eigenvector is kept such that $u_i = Av_i$. Each face in the training set can be represented as a linear combination of the best K eigenvectors. Distance e_r using Euclidean or Mahalanobis distance is used to find a match of face image.

$$\left\| \Omega - \Omega^k \right\| = \sum_{i=1}^{K} \frac{1}{\lambda_i} (w_i - w_i^k)^2 \qquad (2)$$

There is correct classification of 96 percent over lighting variation, 85 percent over orientation variation and 64 percent over size variations in a database containing 2,500 training images of 16 individuals using PCA algorithm. The technique is largely influence by the different lighting conditions. The correlation between the testing and training images of the whole face is not efficient under variation of illumination for good recognition performance [3].

A new method is proposed by L. Zhao and Y.H Yang [4] to compute the covariance matrix using three images taken under different lighting conditions. In this method, a modular eigenspace was composed instead of eigenfaces. This eigenspace composed consists of eigenfeatures such as eigeneyes, eigennose, and eigenmouth. This method is less insensitivity to appearance changes. It achieved better result and a recognition rate of 95 percent on the FERET database consisting of 7562 images of approximately 3,000 individuals.

B. *Linear Discriminant Analysis*

In 1930 R.A Fisher developed linear discriminant analysis (LDA) which contribute for the term 'fisherface'[5]. Fisherfaces is appearance-based method that is successfully used for face recognition. Belhumeur, Hespanha and Kriegman used both the PCA and LDA which produce a subspace projection matrix, similar as done in the eigenface technique. The goal of LDA is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible. In LDA, the scatter between-classes are considered along with the scatter within-classes. It has better recognition rate than PCA due to variations of illumination and expression. Within-class scatter matrix is obtained by applying:

$$S_w = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (y_j - \mu_i)(y_j - \mu_i)^T \qquad (3)$$

where C is the number of classes, μ_i the mean vector of class i, and M_i the number of samples within class. Between-class scatter matrix is obtained by computing:

$$S_b = \sum_{i=1}^{C} (\mu_i - \mu) - (\mu_i - \mu)^T \qquad (4)$$

LDA computes a transformation that maximizes the between-class scatter while minimizing the within-class scatter: maximize $\frac{\det(S_b)}{\det(S_w)}$ The linear transformation is given by a matrix U whose columns are the eigenvectors of $S_w^{-1}S_b$ called Fisherfaces.

$$\begin{bmatrix} b_1 \\ b_2 \\ … \\ b_K \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ … \\ u_K^T \end{bmatrix} (x - \mu) = U^T(x - \mu) \qquad (5)$$

Fisherface method is similar to eigenface method but with improvement of classification of different class images. Fisherface method is able to reduce variation within each class by means of within-class information. The first three principal components that are responsible for changes in light intensity are removed in fisherface. It is therefore more invariant to light intensity than eigenface. However, fisherface requires more complex computation for not only projection of face space but also for calculation of the ratio of between-class scatter to within class scatter. Fisherface therefore requires larger storage and computational time than eigenface [6]. PCA perform better when the training set is small considering its accuracy and computational time. However, LDA outperforms PCA when the number of samples is much larger.

C. *Neural Networks*

Neural networks are a non-linear network that is suitable for representation of non-linear face. The neural network is used in many applications like pattern recognition problems, character recognition, object recognition etc. Neural networks main objective is to capture the complex class of face pattern. Multi-Layer Perceptron (MPL) with a feed forward learning algorithm was applied for face detection. T.J Stonham first attempt to recognise face using a single layer adaptive network called WISARD which contains a separate network for each stored individuals [7]. S. Lawrence, Giles, Tsoi and

A.D. Back proposed a hybrid neural network which combines local image sampling, a self-organizing map (SOM) neural network, and a convolutional neural network [8]. SOM provides a quantization of the sample images, thereby providing reduced dimension and invariance to minor changes. The SOM contains N nodes ordered in a two-dimensional lattice structure and each node has 2 or 4 neighboring nodes respectively. Typically, a SOM has a life cycle of three phases: the learning phase, the training phase and the testing phase. During the learning phase, the neuron with weights closest to the input data vector is declared as the winner. Learning algorithm may be summarized as follows [9]:

1. Initialization: Choose random values for the initial vector weights $w_j(0)$, for $j=1,2,...,l$ where l is the total number of neurons. $w_i=[\ w_{i1},\ w_{i2},...,\ w_{il}]^T \in \Re^n$

2. Sampling: Draw a sample x from the input space with a certain probability. $x=[x_1,x_2,...,x_l]^T \in\ \Re^n$

3. Similarity Matching: Find the best matching (winning) neuron $i(x)$ at time t, $0< t \leq n$ by using the minimum distance Euclidean criterion:

$$i(x) = \arg\min\min_j \left\| x(n) - w_j \right\|, j = 1,2,...,l \tag{6}$$

4. Updating: Adjust the synaptic weight vector of all neurons by using the update formula:

$$w_j(n + 1) = w_j(n) + \eta(n)h_{j,i(x)}(n) \left(x(n) - w_j(n) \right) \tag{7}$$

where $\eta(n)$ is the learning rate parameter, and $h_{j,i(x)}$ is the neighborhood function centered around the winning neuron $i(x)$.

5. Repeat from *step 2* until no changes in the feature map are observed.

Self-organizing map record two values: the total number of winning times for both subject present and not present in image database. During training phase, the number of wins is recorded along with the label of the input sample for each node. During testing phase, each input vector is compared with all nodes of SOM, and the best match is found based on some minimum distance calculator.

Face recognition that uses features derived from discrete cosine transform (DCT) coefficients along with a SOM-based classifier achieved a recognition rate of 81.36 percent for 10 successive trials after training for approximately 850 epochs using an image database of 25 face images of 5 subjects each having 5 images with different facial expressions [10].

D. Graph Matching

Lades, Vorbruggen, Buhmann, Lange, Malsburg, Wurtz, and Konen presented a dynamic link structure for distortion invariant object recognition [11]. Elastic bunch graph matching is employed to find the closest stored graph by estimating a set of features. In these, fiducial points on the face are described by sets of wavelet components (jets). Elastic Bunch Graph Matching algorithm recognizes novel faces by first localizing a set of landmark features and then measuring similarity between these features. Both localization and comparison uses Gabor jets extracted at landmark positions. In localization, jets are extracted from novel images and matched to jets extracted from a set of training/model jets. Similarity between novel images is expressed as function of similarity between localized Gabor jets corresponding to facial landmarks.

Graph is an un-directed graph and object-adaptive which have nodes at the fiducial point of face image. The structure of the graph is the same for each face. However, some nodes may be undefined due to occlusion. Face bunch graph is a stack-like structure consisting of a jet bunch rather than a jet. Each node is labeled with a bunch of jets and each edge labeled with average distance between corresponding nodes in face samples. Assume for a particular pose that there are M model graphs G^{Bm} $(m = 1,2,...,M)$ of identical structure. The corresponding FBG B is then given the same structure, its nodes are labeled with bunches of jets J_n^{Bm} and its edges are labeled with the averaged distances $\Delta \vec{x}_e^B = \sum_m \Delta \vec{x}_e^B / M$. Graph similarity between an image graph and FBG is evaluated by using similarity function. For an image graph with nodes $n = 1,2,...,N$ and edges $e = 1,2,...,E$ and an FBG B with model graphs $m = 1,2,...M$ the similarity is defined as

$$S_B(G^I, B) = \frac{1}{N}\sum_n \max_m (S_\varphi(J_n^I, J_n^{Bm})) - \frac{\lambda}{E}\sum_e \frac{(\Delta \vec{x}_e^I - \Delta \vec{x}_e^B)^2}{(\Delta \vec{x}_e^B)^2}, \tag{8}$$

where λ determines the relative importance of jets and metric structure. J_n are the jets at nodes n, and $\Delta \vec{x}_e$ are the distance vectors used as labels at edges e [12]. EBGM uses the fiducial structure of the face and connect these points so that the bunch graph tends to translate, scale, rotate, and deform in the image plane. EBGM is less susceptible to feature missing and will still be successful in recognition of images that have a change or missing a few features unlike Eigenface and Fisherface which are both prone to missing feature. The recognition rate for the matching test of 111 faces of 15 degree and 30 degree rotation is 86.5 percent and 66.4 percent. EBGM in general perform better to other recognition technique in terms of rotation; however, the matching process is computationally expensive [1].

E. Hidden Markov Model

Samaria and Fallside applied hidden markov model to human face recognition [13]. Face was divided into regions that can be associated with the states of hidden markov model like the eyes, nose, mouth, etc. Two-dimensional images are converted to 1D temporal sequence or 1D spatial sequence as HMM deals with one dimensional sequence. A band sampling technique is used to extract a spatial observation sequence. Each observation vector is a block of L lines and M lines that overlap between successive observations. Hidden markov models consists of: (1) an underlying, unobservable Markov chain with a finite number of states, a state transition probability matrix and an initial state probability distribution and (2) a set of probability density function associated with each state. The elements of HMM are [14]:

1. N number of states in the model and set of states S where $S=\{S_1,S_2,...,S_N\}$. The state of the model at a given time t is given by $q_t \in S, 1 \leq t \leq T$, where T is the length of the observation sequence.

2. Π, the initial distribution, i.e. $\Pi = \{\pi_i\}$ where: $\pi_i = P[q_1 = S_i], 1 \leq i \leq N$

3. A, The state transition probability matrix, i.e. $A = \{a_{ij}\}$ where $a_{ij} = P[q_t = S_j | q_{t-1} = S_i]$ $1 \leq i, j \leq N$ with the constraint, $0 \leq a_{ij} \leq 1$, and $\sum_{j=1}^{N} a_{ij} = 1, 1 \leq i \leq N$

4. B, The state probability matrix, i.e. $B = \{b_j(O_t)\}$. The states are characterized by probability density function of the form:

$$b_i(O_t) = \sum_{k=1}^{M} c_{ik} N(O_t, \mu_{ik}, U_{ik}), 1 \leq i \leq N \qquad (9)$$

where c_{ik} is the mixture coefficient for the k th mixture in state I and O_t the observation symbol at time t. Without loss of generality $N(O_t, \mu_{ik}, U_{ik})$ is assumed to be a Gaussian pdf with mean vector U_{ik}.

HMM is defined as

$$\lambda = (A, B, \Pi) \qquad (10)$$

A test image that identifies the highest likelihood of similarity is considered the best match. HMM approach has a recognition rate of 87 percent using ORL database consisting of 400 images of 40 individuals. 95 percent successful recognition rate is achieved using a pseudo 2D (2 dimensional) HMM [13].

F. Support Vector Machine

Support vector machine has a high generalization performance and is recognized as an effective pattern recognition method. Given a set of points belonging to two classes, a SVM finds the hyperplane that separates the largest possible fraction of points of the same class on the same side, while maximizing the distance from either class to hyperplane [15]. The main characteristic of SVMs are that: (1) they minimize a formally proven upper bound on the generalization error; (2) they work on high-dimensional feature spaces; (3) prediction is based on hyperplane in these feature space; (4) outliers in a training set can be handled by means of soft margins.

SVMs belong to the class of maximum margin classifiers. They classify between two classes by finding a decision surface that has maximum distance to the closest points in the training set which are term support vector. SVM start with a training set of points $x_i \in \Re^n, i = 1, 2, \dots, N$ where each point x_i belongs to one of the two classes identified by the label $y_i \epsilon \{-1, 1\}$. The goal is to separate the two hyperplane such that the distance to the support vectors is maximized. The optimal separating hyperplane (OSH) has the form:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i x_i . x + b \qquad (11)$$

where α_i and b are coefficients. To perform multi-class classification the optimal separating hyperplane has the form:

$$d(x) = \frac{\sum_{i=1}^{l} \alpha_i y_i x_i . x + b}{\left\| \sum_{i=1}^{l} \alpha_i y_i x_i \right\|} \qquad (12)$$

where d is the classification result for x, and $|d|$ is the distance from x to the hyperplane. The larger the $|d|$, the more reliable the classification result [16].

P.J. Phillips compared a SVM-based algorithm with principal component analysis (PCA) based algorithm on different set of images from the FERET database by measuring performance for both verification and identification condition. For identification, SVM performs better than PCA having success rate of 77-78 percent against 54 percent success rate for PCA. For verification, SVM performs better of having only 7 percent of error rate against error rate of 13 percent for PCA based algorithm [17].

III. CONCLUSION

Face recognition is widely and actively being research for the past few years and continues to evolve with the efforts of dedicated researchers all over the world. It has a huge number of commercial applications; however, the real application is in enforcing law and order and for identity measures of nationality. The discussed techniques are being research over to find a better algorithm that would perform at high accuracy and lesser computational time. Active research being conducted in the areas shows that face recognition technique available is far from being perfected. Most of these algorithms have high recognition rate for a small database and unsuitable in a large database as large computational time may not be practically feasible.

All approaches to face recognition mentioned are based on color and gray-scale images. Experiments are conducted after color images are converted to gray-scale. This requires large storage and large computational time. To reduce the space and time complexity, new approach based on binary images is suggested which will be faster in computation and will require less storage. A binary image has the advantages of simplicity in processing. New approach to face recondition based on binary images may found great advantages over the existing techniques based on color and gray-scale images.

REFERENCES

[1] A.S. Tolba, A.H. El-Baz, and A.A. El-Harby, "Face Recognition: A Literature Review", World Academy of Science, Engineering and Technology", Vol:2 2008-07-21.

[2] M. Kirby and M. Sirovich, "Application of the Kahunen- Loeve procedure for the characterisation of human faces," IEEE Trans. Pattern Recognition and Machine Intelligence, vol. 12, pp. 831-835, Dec. 1990.

[3] M. Turk and A. Pentland, "Face Recognition using Eigenfaces," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1991.

[4] L. Zhao and Y.H. Yang, "Theoretical analysis of illumination in PCA based vision systems," Pattern Recognition, vol. 32, pp. 547-564,1999.

[5] R.A. Fisher, "The use of multiple measurements in Taxonomic Problems", 1936.

[6] A. Pentland, B. Moghaddam and T. Starner, "View-Based and modular eigenspaces for face recognition," Proc. IEEE CS Conf Computer Vision and Pattern Recognition, pp. 84-91, 1991.

[7] T.J. Stonham, "Practical face recognition and verification with WISARD, "Aspects of Face Processing, pp. 426-441, 1984.

[8] S. Lawrence, C.L. Giles, A.C. Tsoi, and A.D. Back, "Face recognition: A convolutional neural-network approach," IEEE Trans. Neural Networks, vol. 8, pp. 98-113, 1997.

[9] D. Kumar, C.S. Rai, and S. Kumar, "Face Recognition using Self-Organizing Map and Principal Component Analysis" in Proc. on Neural Networks and Brain, ICNNB 2005, Vol. 3, Oct 2005, pp. 1469-1473.

[10] J. Nagi, S.K. Ahmed, and F. Nagi, " A MATLAB based Face Recognition System using Image Processing and Neural Networks", 4th International Colloquium on Signal Processing and its Applications, March 7-9, 2008.

[11] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. Von Der Malsburg, R.P. Wurtz, and M. Konen, "Distortion invariant object recognition in the dynamic link architecture," IEEE Trans. Computers, vol. 42, pp. 300-311, 1993.

[12] L. Wiskott, J-M. Fellous, N Krüger, and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching",Intelligent Biometric Techniques in Fingerprint and Face Recognition, Chapter 11, pp. 355-396, (1999).

[13] F. Samaria and F. Fallside, "Face identification and feature extraction using hidden markov models," Image processing: Theory and Application, G. Vernazza, ed.,Elsevier, 1993.

[14] A.V. Nefian and M.H. Hayes, "Hidden Markov Models for face recognition," in ICASSP98, pp. 2721-2724, 1998.

[15] C.J. Lin, "On the convergence of computer method for support vector machines," IEEE Trans. On Neural Network, 2001.

[16] B. Heisele, P. Ho, T. Poggio, "Face Recognition with Support Vector Machines: Global versus Component-based Approach", Computer Vision and Image Understanding 91 (2003), pp. 6-21.

[17] P.J. Phillips, "Support vector machines applied to face recognition," Processing system 11, 1999.