

## Study on Face Recognition Algorithm

Mr. Sarath P. S<sup>1</sup>, Ms. Kumary R Soumya<sup>2</sup>

<sup>1,2</sup>(Department of Computer Science and Engineering, Jyothi Engineering college/ Calicut University, India)

**Abstract:** Biometrics authentication is started to be used worldwide since the security it provides is of greater extends. Face recognition is an important method of such an authentication. Face recognition can be carried out by many algorithms such as PCA, LDA, ICA, BIC, AAM etc. In this paper, a detailed study about these algorithms is performed. The study deals with the definition part, methodology and the classification as well.

**Keywords:** Face Recognition, Biometrics, PCA, LDA, ICA, BIC, AAM.

### I. INTRODUCTION

Biometrics or biometric authentication is a way of identifying humans using their characteristics. It is a method of identification and access control. Biometric identifiers are categorized into physiological characteristics which are related to the shape of the body including fingerprints, face recognition ,iris recognition etc. and behavioral characteristics related to the pattern of behavior of person including typing rhythm, voice etc. Universality, uniqueness, measurability, permanence, performance, acceptability and circumvencce are the seven weighted factors for determining the biometrics for use in a particular application. Universality means that the person should posses the trait. Uniqueness means the trait should be unique. Permanence is the manner in which the trait varies over time. Measurability is the ease of acquisition. Acceptability is how others accept the technology. Performance refers to the accuracy and speed and robustness of the technology. Circumvention is the easiness in imitating the trait. A face recognition system compares the video or image of a person in database to the input captured. Comparison is done on the structure, shape and proportions of face. Distance between the eyes , jaw, mouth and nose. Several pictures of the subject at different angle with different facial expressions are enrolled to the database at the training phase. Identification and verification is performed by the subject standing in front of the camera for a few seconds. This image is compared to the one previously added. Different face recognition techniques are feature extraction, neural network, automatic face procession Eigen faces etc. Biometric techniques can be used for surveillance purpose. I.e. Face recognition can be done from a particular distance without having any contact with the subject so that they are unaware about it. Among several information in human faces, its efficient to choose the unique information while discarding other irrelevant information for making the system efficient. In this paper, a study on different face recognition algorithms is performed. These algorithms include PCA, LDA, BIC, ICA, AAM and Gabor Vectors.

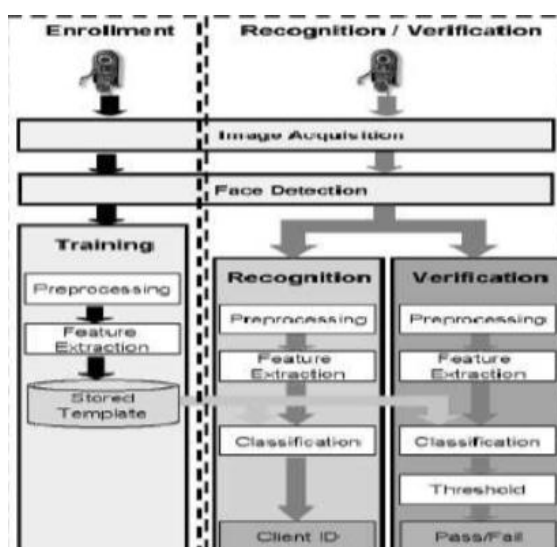


Fig 1: Flow chart of face recognition

## II. Pca (Principal Component Analysis)

PCA is a technique used for face recognition, image compression, and to find patterns in data of high dimension. It finds the pattern and highlights the similarities and dissimilarities in the data.. The main advantage of PCA is that after finding the data, it is able to compress the data by reducing the number of dimension without much loss of information. This section shows you the steps needed to perform Principal Component Analysis when a set of data is provided.

### 2.1 Method

Step 1: Input the data set

Step 2: Subtract the mean from the data set.

All the x values are subtracted by the mean of all x values and all y values are subtracted by the mean of all y values. This step is required for the PCA to work properly.

Step 3: The covariance matrix are calculated for the data set. Covariance matrix can be defined as

$$C^{m \times n} = (c_i, c_j = cov(Dim_i, Dim_j))$$

Step 4: Calculate the Eigen values and the Eigen vector of this covariance matrix.

Step 5: Chose components and obtain the feature vector. The principal component of data set is chosen such as it holds Eigen vectors with highest value. After the eigenvectors are obtained from the covariance matrix, the Eigen values are sorted according to ascending order. This yields the components in order of significance and helps to ignore the one that have lesser significance. Ignoring small Eigen values doesn't result in losing much information. After ignoring the m Eigen vectors from a original data set of n dimension, resulting matrix have a dimension of p. These eigenvectors are arranged in a column to yield the feature Vector.

$$\text{Feature Vector} = (eig_1 eig_2 \dots eig_p)$$

In our data set, the feature vectors are obtained by leaving out the smaller, less significant values and arranging the rest in a single column.

Step 6: The new data set is derived. It is done by taking the transpose of the feature vector and multiplied by the left of original data set.

$$\text{Final Data} = \text{RowFeatureVector} \times \text{RowDataAdjust}$$

## III. Lda (Linear Discriminant Analysis)

Linear Discriminant Analysis is a commonly used technique for data classification and dimensionality reduction. LDA deals with the data whose within class variance is unequal. If the ratio of between class variance to the within class variance is maximized, maximal seperability is guaranteed. The major difference between the LDA and PCA is that PCA looks into the feature classification while LDA does for data classification. LDA tries to draw a decision region between the given set of data.

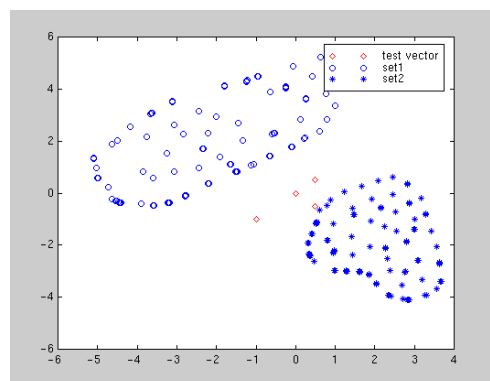


Fig 2: Decision region of a given data set

LDA defines a linear combination of features that characterizes two or more classes of events, and hence, separates the two classes explicitly. The PCA is also used for finding the linear combination of features separating the classes of events. The goals of the LDA are

1. Selection of separator variables.
2. Selection of Discriminant functions.
3. Classification

The steps for the Discriminant analysis are as follows.

1. The best Discriminant variable is identified so that it separates the classes effectively.
2. Development of function for the computation of new index that describe dissimilarity.
3. The observations are classified based on the Discriminant function.

Classification of test vectors in the transformed space can be done by two different approaches namely class-dependent transformation and class-independent transformation.

**Class-dependent transformation:** In this approach, the ratio of the between class variance to the within class variance is maximized so that accurate class separability is obtained.

**Class-Independent transformation:** The ratio of the overall variance to the within class variance is maximized in this approach. All the data points are transformed using only one optimizing criteria. Each class is considered as separate classes in this approach.

### 3.1 Method

1. The data set and the test sets are formulated.
2. The mean of each data set and entire data set is calculated. If  $\mu_1$  and  $\mu_2$  are the mean of each data sets and  $\mu_3$  is the mean of entire data, the formula is given by the equation

$$\mu_3 = p_1 \mu_1 + p_2 \mu_2$$

3. Within class and between classes scatters are calculated using the equations

$$S_w = \sum_j p_j \times (cov_j)$$

$$S_w = 0.5 \times cov_1 + 0.5 \times cov_2$$

Where, the second equation is used for two class problems. Expected covariance of each of the classes is the within class scatter and the covariance matrices are symmetric and computed by,

$$cov_j = (x_j - \mu_j)(x_j - \mu_j)^T$$

Where  $S_w$  is the within class scatter.

The between class scatter is given by the equation

$$S_b = \sum_j (\mu_j - \mu_3) \times (\mu_j - \mu_3)^T$$

4. Set of Eigen vectors whose Eigen values are nonzero is linearly independent and invariant under the transformation. Zero Eigen values represent the linear dependency between the features. The Eigen vectors with nonzero Eigen values are considered and that of zero Eigen values are discarded to obtain a non redundant set of features.

5. Obtain the transformation matrices. Using single LDA transform, the data sets are transformed.

For a class dependent LDA, Transformed set is given by,

$$\text{transformed\_set} = \text{transform\_specT} \times \text{data\_setT}$$

and for a class independent LDA, transformed set is

$$\text{transformed\_set} = \text{transform\_specT} \times \text{data\_setT}$$

The test vectors are also transformed and classified similarly.

6. Euclidean Distance or RMS distance is used to classify point  $s$  after the completion of transformation. Euclidean distance is computed using the formula

$$\text{dist}_n(\text{transform}_n\text{-spec})^T = \sqrt{\|x - \mu_{\text{trans}}\|^2}$$

Where, the  $\mu_{\text{trans}}$  refers to the mean of transformed data set.

For each test point,  $n$  Euclidean distances are obtained for  $n$  classes.

7. The smallest Euclidean distance perform the classification of data belongs to the class.

#### IV. BIC (BAYESIAN INTRAPERSONAL/EXTRAPERSONAL CLASSIFIER)

BIC is a probabilistic classifier for face images presented by Moghaddam and Pentland. This classifier is distinct in the sense that the difference image in two photos is examined for the determination of whether the photo belongs to same subject or not. Difference images originated from different subjects are called extrapersonal whereas that of the same kind is called intrapersonal. This classification is based on the key that each difference image is viewed as a point in the high dimensional space. This space is sparsely populated since each difference image occupies a point in the space. The difference images form clusters. Major assumptions Moghaddam and Pentland's work is that the difference images are originated from distinct and localized Gaussian distribution within a space of all possible images. The determination of actual parameter for such distributions are unknown and hence performed by a couple of stages where the training stage, known as density estimation is performed using PCA and estimates statistical properties for the difference images belong extrapersonal and intrapersonal separately denoted by  $\Omega_I$  and  $\Omega_E$  respectively. The second stage is the testing phase in which an arbitrary difference image  $\Delta$  and estimate the probability distribution  $\Delta/\Omega_I$  and  $\Delta/\Omega_E$  as a means of identification. The posteriori probability of  $\Delta$  belonging to the intrapersonal is estimated by the maximum a posteriori (MAP) classifier using Baye's rule and a similarity measure SMAP is equated with this probability.

$$S_{MAP} = P(\Omega_I/\Delta) = \frac{\hat{P}(\Delta/\Omega_I) \cdot \hat{P}(\Omega_I)}{\hat{P}(\Delta/\Omega_I) \cdot \hat{P}(\Omega_I) + \hat{P}(\Delta/\Omega_E) \cdot \hat{P}(\Omega_E)}$$

Consider a novel probe image is comparing with a total of  $n$  known images in the gallery, a  $n$  difference image is created by subtracting the probe image from the  $n$  known images. These difference images are then ranked and the image yielding the highest similarity score is chosen. A simpler Maximum Likelihood (ML) classifier formulation only consider the intrapersonal while ignoring the extrapersonal class information. Moghaddam and Pentland propose this ML classifier as a substitute for the MAP classifier. The similarity measure for ML classifier is given by

$$S_{ML} = \hat{P}(\Delta/\Omega_I)$$

The MAP classifier is trained by running the PCA twice ie, one for a set of intrapersonal and other for a set of extra personal. The ML classifier requires PCA to run only one time. The training parameters are  $N$ ,  $M$  and  $T$  in the both cases

Where,

$N$  is the dimensionality of original data

$T$  is the number of difference images

$M$  is the number of dimensions kept

The number of pixels in the difference image depends on the image data used. It is 150 by 130 pixels for a FERET image preprocessed in standard manner. ie,  $N= 19,500$ . The upper bound on the intrinsic dimensionality of the difference image subspace is limited to  $T$  which typically ranges from 100 to 1,000. For the computational easiness, the PCA subspace is further truncated by only retaining the  $M$  most significant dimensions. After each subspace training, we have

1. A projection matrix  $\Phi_M$  and vector of eigen values  $\lambda$ .
2. Parameter  $\rho$  which is the average of eigen values from  $M$  to  $T$ .
3. Projection of each difference image into PCA subspace.

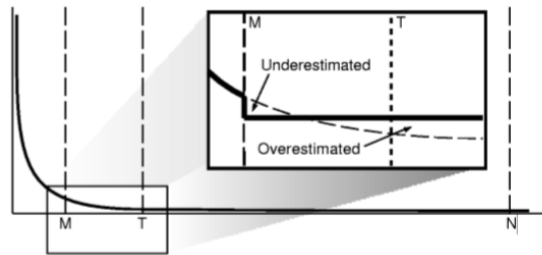


Fig 3: A representative distribution of  $\lambda_i$  and the effects of  $\rho$ .

### V. ICA (INDEPENDENT COMPONENT ANALYSIS)

ICA deals with higher order statistics and independent source components are identified from their linear mixture. Therefore, ICA is more powerful than PCA in representing a data as its main goal is to provide an independent representation rather than uncorrelated one. Linear transformation is searched by the ICA of random vector for minimizing the statistical dependence between its components.

Let  $X \rightarrow R^N$  is a vector of images

Where, N is the dimensionality of image space. Its covariance matrix is given by the equation

$$E\{X - E(X)[X - E(X)]^T\}$$

Where E() is expectation operator.

T is transpose operator.

Comom developed an algorithm consisting of three operations for deriving the ICA transformation F. whitening, rotation and normalization is the three operations where whitening perform the transformation of random vector X into another vector U which has a unit covariance matrix. Source separation is performed by the rotation operator by minimizing mutual information approximated by higher order cumulants. Derivation of unique independent components in terms of orientation, order of projection and unit term is done by normalization. For enhanced performance, ICA is required to be carried out in a compressed and whitened space where information regarding original data is preserved and discarding small trailing Eigen values.

### VI. Aam (Active Appearance Model) And Gabor Feature Vectors

Face recognition using AAM and Gabor feature vector uses three stages named modeling, registration and recognition. At first, Construction of Model Bunch Graph of about M normalized face model is carried out. This stage is termed as modeling. After which, in the registration process, facial feature points are localized from the normalized face images and this localization is refined based on Gabor jet similarity. For each facial feature point, Gabor feature vectors are extracted and registered into the database. In the recognition stage, the location of facial feature points are performed as in the registration process i.e. using AAM and Gabor jet similarity. Then these features are extracted. This extracted Gabor vector features are then compared with the registered Gabor feature vectors for the measurement of similarity. The result whether match or mismatched is decided based on this similarity.

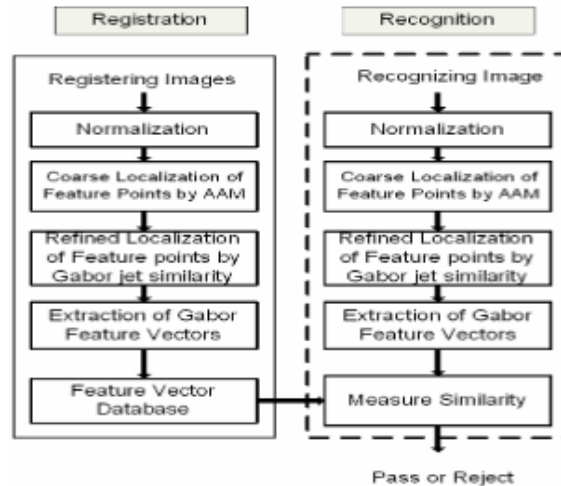


Fig 4: Active Appearance Model

## VII. Conclusion

All the algorithms explained in this paper is used for the face recognition or broadly we can say pattern recognition. PCA and LDA are sharing both the dissimilarities and similarities. The major difference is that the PCA consider the feature classification while LDA uses the data classification. Whereas both of them are used for finding the linear combination of features separating the classes of events. BIC is distinct since difference image between two photos are used for the recognition purpose. Each image is viewed as a point in the high dimension space. ICA works with higher order statistics and independent components are calculated. Hence ICA is a more powerful way of data representation than PCA. ICA mainly aims on providing an independent representation. A Face recognition using AAM and Gabor Vectors proceed through three stages named modeling, Registration and recognition. Facial feature points are localized after modeling the normalized face model. Recognition is performed by comparing registered feature and extracted feature.

## References

### Journal paper

- [1] S. Gong et al., "Dynamic Vision: From Images to Face Recognition", Imperial College Press, 2001.
- [2] Lindsay I Smith , " A tutorial on Principal Components Analysis" , February 26, 2002
- [3] Maricio Luis Teixeira, "The Bayesian intrapersonal/extrapersonal classifier" colorado state university, 2003
- [4] Marcio L. Teixeira and J. Ross Beveridge , " An Implementation and Study of the Moghaddam and Pentland Intrapersonal/ Extrapersonal Image Difference Face Recognition Algorithm" , may 2003.

### Chapters in Books:

- [5] R. Duda et al. ,John Wiley, "Pattern Classification", 2001 (pp. 117-124, soft copy).
- [6] A. Martinez and A. Kak, "PCA versus LDA", IEEE Transactions on Pattern Analy sis and Machine Intelligenvol. 23, no. 2, pp. 228-233, 2001, (on-line)
- [7] A Webb Statistical Pattern Recognition, Arnold, 1999 (pp. 112-116, soft copy).

### Theses:

- [8] DOC493: Intelligent Data Analysis and Probabilistic Inference Lecture 15
- [9] Special Issue of International Journal of Computer Science & Informatics (IJCSI), ISSN (PRINT) : 2231-5292, Vol.- II, Issue-1, 2.
- [10] CSCE 666 Pattern Analysis | Ricardo Gutierrez-Osuna | CSE@TAMU