ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 10, October 2015*

# Feature Extraction Using PCA Using Algorithm

**Nishu[1], Navdeep Kaur[2]**

Student, Department of CSE, (Rayat)Bahra Group of Institutes, Patiala, Punjab, India [1]

Assistant Professor, Department of CSE, (Rayat)Bahra Group of Institutes, Patiala, Punjab, India [2]

**Abstract**: In face recognition technology, facial marks identification method is one of the unique facial identification tasks using soft biometrics. Also facial marks information can enhance the face matching score to improve the face recognition performance. As numbers of folk apply their face with cosmetic items.Some of the facial marks are invisible from their faces. Iris recognition is the most accurate biometrics which received increasing attention in departments which require high security. In this paper, we make a comparative study of performance of image transforms using principle of component analysis (PCA).The main aim of this paper is to show that how can we get better overall accuracy than the existing methods of feature extraction of iris recognition system.

**Keywords**: Face recognition, facial marks, PCA, edge detection, Iris recognition, edge detection

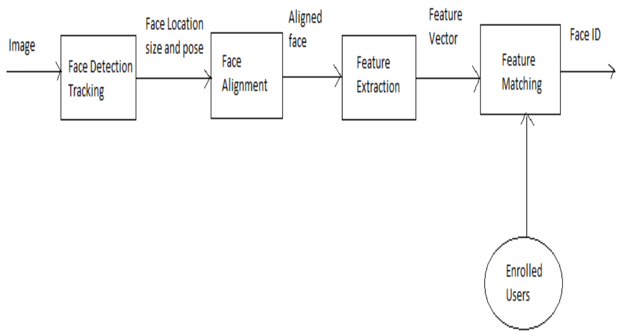
## I. INTRODUCTION

Face recognition is a biometric identification by scanning a person's face and matching it against a stored template. Face recognition can be used for both identification and verification. Shadows in images create lots of problems on image analysis. There is no need of shadow in an image i.e. shadow is unwanted part in images. Shadow affects the images because of shadow lots of data and information is lost from images. Security related an issue has become an important aspect in each and every organization. Every organization requires that there security methods should be as efficient as possible. Development are being made day by day to enhance and improve the security. One such efficient method that is brought forward by means of this paper is biometric. Biometric refers to the use of psychological or biological. Characteristics of human beings to determine the identity of the person. Identification of person is very important and is done is the most of the important sites such as banks, airports, companies and many more to identify the identity of the person. There are various methods to do the job such as assigning Id and password which is also called as knowledge based processing. But these approaches have limitation. Face recognition and speech recognition have also been widely studied over the last 30 year, whereas this recognition is a newly emergent approach to personal identification in the last decade among all biometrics (such as finger print, face, print, giant, voice, iris, dental radiographs etc.),iris recognition is the most consistent one. The iris is a thin circular diaphragm, which lies between the cornea and the lens of the human eye. The pattern the human iris differs from person to person there are not ever two iris alike not even for genetically identical twins. The iris is considered one of the most stable biometrics as it is believed to not change significantly during a person's life time and its physiological response to light which provides the detection of a dead or artificial iris, avoiding this kid of counterfeit? Other properties of the human iris that increase its suitability for use in automatic identification include its inherent isolation and protection from the external environment, being an internal organ of the eye, behind the cornea and the aqueous humans.

### 1.1 Primary Iris Recognition Process:

A typical iris recognition system is schematically. The whole iris recognition process is basically divided into steps:

1) Image acquisition;
2) Iris image preprocessing;
3) Iris feature extraction; and
4) Matching




A typical iris recognition system is schematically. The whole iris recognition process is basically divided into four steps 1) Image acquisition 2) Iris image preprocessing 3) Iris feature extraction 4) Matching. Now a day, various algorithms for iris recognition have been presented. For therefore, the preprocessing of iris image includes four aspects: localization, normalization, enhancement, delousing, and the selection of iris valid areas. Various algorithms have been applied for feature extraction and pattern matching processes. These methods use local and global features of the iris. A great deal of advancement in iris recognition has been made through these efforts; therefore, a detailed performance comparison of all these algorithms is necessary. The goal of this paper is to compare feature extraction algorithm based on PCA, Haar transform, Block sum algorithms and proposed algorithm.


DOI 10.17148/IJARCCE.2015.41061

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 10, October 2015*

## II. RELATED WORK

There are limited works an identical twin recognition using 2d face biometrics, below is a brief survey of existed work in face biometrics, twin identification and image retrieval systems.

Sun et al. [10] were the first to evaluate the performance of appearance based face recognition to distinguish between twins. They compared with performance of iris, finger print and a fusion of them. Their database was collected in 2007 at the fourth annual festival of Beijing twin day. The face subset used in the experiments contained 134 subjects, each having around 20 images. All images were collected during a single session over a short interval. Experiments were conducted using the commercial mat char and showed that identical twins are a challenge to current face recognition systems.

Phillips et al. [15] thoroughly extended the analysis of the performance of the face recognition system in distinguishing between identical twins on the another database collected at the twins days festival in ohio in 2010. It consisted of image of identical twins collected on the same day and 24 pairs with images collected on the same day and 24 pairs with images collected one year apart. Facial recognition performance was tested using three of the top submissions to the still. Face track at multiple biometric evaluation 2010. Based on their experimental results, but under more realistic conditions, distinguishing between identical twins was very challenging.

Klare et al. [11] analysed the features of each facial component to distinguish identical twins from the same database in [16]. They also analysed the possibility of using facial marks to distinguish identical twins. They also confirmed the challenge of recognizing identical twins merely based on appearance. All these works showed the need for new approaches to help improve performance when recognizing identical twins. The accuracy of distinguishing between identical twin pairs is measured using the entire face, as well as each facial component (eyes, eye- brows, nose, and mouth). The impact of discriminant learning methods on twin face recognition is investigated. Experimental results indicate that features that perform well in distinguishing identical twins are not always consistent with the features that best distinguish two non-twin faces.

Paone et al. [18] used an experimental data set comprised of 17486 images from 126 pairs of identical twins (252 subjects) collected on the same day and 6864 images from 120 pairs of identical twins (240 subjects) with images taken a year later to measure the performance on seven different face recognition algorithms. According to it, Facial recognition algorithms should be able to operate even when similar looking individuals are encountered, or even in the extreme case of identical twins. Performance was reported for variations in illumination, expression, gender, and age for both the same day and cross-year image sets. Regardless of the conditions of image acquisition, distinguishing identical twins are significantly harder than distinguishing subjects who are not identical twins for all algorithms.

Lim et al. [21] proposed A Face Recognition System Using Fuzzy Logic and Artificial Neural Network. They developed a method that extracts a feature vector that is very important to recognize the facial image. They used the eye blinking method to get the location of eye roughly. They have gotten feature vector using locations and distances between feature points, that is, eyes, nose, mouth and the outline of the face.

## III.PROPOSED METHODOLOGY

In this work, a method has been proposed to differentiate between identical twins using facial marks alone. Facial marks are considered to be unique and inherent characteristics of an individual. Although they are similar in appearance, they can be distinguished using facial marks. High-resolution images enable us to capture these finer details on the face. Facial marks are defined as visible changes in the skin and they differ in texture, shape and colour from the surrounding skin. Facial marks appear at random positions of the face. By extracting different facial mark features we aim to differentiate between identical twins. Difficulties in manual annotations noted in [10] include:

Manual annotation is a difficult task because it involves training and familiarizing an individual with the definitions and characteristics of the different types of facial marks (apart from learning how to use the facial annotation tool).Observers experienced difficulty in differentiating between categories of facial marks, especially in the case of moles and freckles. Although all observers appear to have annotated the prominent facial marks, some observers failed to annotate the less prominent but visible marks. It is a time consuming and an expensive task. Hence there is a need for an automatic facial mark detector to overcome these problems.

To overcome these difficulties we have proposed a new model which can automatically locate markings. In next section, the proposed work has been explained step by step but below is a brief description of proposed model.

We have divided the whole work into two sections or algorithms as explained below.

**Algorithm:** Pre-processing of face image and Retrieval of twin images from database using PCA (Principal component analysis)

Below is a brief explanation of all the minor steps which are involved in running each algorithm.

**Algorithm level design**

Algorithm 1: Pre-processing of image and Identical twin retrieval using PCA

Below are the steps that are involved in this algorithm

**Step 1: Evaluating Face region from whole image**

This step is the one that processes the input image. This step feeds the succeeding step with an image which is

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 4, Issue 10, October 2015*

derived from the original image. The algorithm used for this part is the Viola-Jones Algorithm for Face Detection. This algorithm works by looking for Haar features. These features are basically black and white rectangles that the algorithm hunts for in an image. The algorithm adds the number of pixels of the rectangles to a box depending on whether it is contained in a threshold or not. This step processes the image in such a way wherein it first tries to determine whether a human face is present in the image or not. Once a face is detected, the system crops the image, removing the background and leaving only the face region.

**Step 2: Apply clustering for cluster the face region into number of segments**

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 [44] and improved by Bezdek in 1981 [45]) is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \qquad \text{.......................... 2}$$

$$1 \le m < \infty$$

where $m$ is any real number greater than 1, $u_{ij}$ is the degree of membership of $x_i$ in the cluster $j$, $x_i$ is the $i$th of d-dimensional measured data, $c_j$ is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership $u_{ij}$ and the cluster centers $c_j$ by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}\right)^{\frac{2}{m-1}}} \qquad \text{....................... 3}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m} \qquad \text{...................................... 4}$$

This iteration will stop when

$$\max_{ij}\left\{ \left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right| \right\} < \varepsilon \qquad \text{.......................... 5}$$

where $\varepsilon$ is a termination criterion between 0 and 1, whereas $k$ are the iteration steps. This procedure converges to a local minimum or a saddle point of $J_m$.

The algorithm is composed of the following steps:

Initialize U= [$u_{ij}$] matrix, $U^{(0)}$

At k-step: calculate the centres vectors $C^{(k)} = [c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

Update $U^{(k)}$ , $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}\right)^{\frac{2}{m-1}}}$$

If $\| U^{(k+1)} - U^{(k)} \| < \varepsilon$ then STOP; otherwise return to step 2.

**Step 3: Replacing clusters with mean average value**

In this step, an average value is applied to all pixels in individual clusters by taking mean value of intensity of all pixels in that cluster taken from original image. It helps in improvement of PCA results.

**Step 5: To store the output image into database into PGM format**

**Step 6: Evaluate eigenvectors and Eigen faces of database**

Principal component analysis (PCA) is one of the most popular methods for reducing the number of variables in face recognition [38]. In PCA, faces are represented as a linear combination of weighted eigenvectors called as Eigen faces [35][36][37]; These eigenvectors are obtained from covariance matrix of a training image set called as basis function. The number of Eigen faces that obtained would be equal to the number of images in the training set. Eigen faces takes advantage of the similarity between the pixels among images in a dataset by means of their covariance matrix. These eigenvectors defined a new face space where the images are represented. To fix the required notation, let us introduce the following symbols.

Let training image set I consist of N images each having size a×b pixels. Using conventional row appending method converts each of the images into a×b dimensional column vector.

$$I = \{i_1, i_2, \ldots\ldots\ldots i_N\} \qquad \text{......................................6}$$

Covariance matrix c of training image set are calculated by using equation (7)

$$c = \frac{1}{N}\sum_{n=1}^{N} (i_n - \bar{i})^T (i_n - \bar{i}) \qquad \text{.........................7}$$

where 'i' is the mean vector of all images in the training set. Eigenvalue and eigenvectors of covariance matrix is calculated using equation (8)

$$cv = \lambda v \qquad \text{.............................................. 8}$$

where 'λ' denotes the eigenvalues of c, and v stands for the corresponding eigenvectors. Note that the rank of the covariance matrix is N, hence at most N number of eigenvectors can be computed.

$$U = (i_n - \bar{i}) \times v \qquad \text{..........................................9}$$

Where n = 1, 2,..., N.

The Eigenvectors found, U have a face like appearance, they are called Eigen faces. Sometimes, they are also called as Ghost Images because of their weird appearance After the face space has been constructed, the feature vectors are formed as a linear combination of the eigenvectors of the covariance matrix. Project an image n i into the face space with the help of following equation.

$$P_n = U^T \times (i_n - \bar{i}) \quad \ldots\ldots 10$$

Where Pn, n = 1, 2,…….N are the vector of weights associated with the eigenvectors in c. One can experiment with the number of eigenvectors to compute the weights, generally only a few amount provide sufficient information for adequately representing the images in the face space. For recognition of unknown face or test image, normalize it by subtracting from mean vector of all images in the training set. Then using equation (9) project the normalized test image as shown in the following equation

$$T = U^{T\,\prime} \; D \quad \ldots\ldots 11$$

where D is normalize the test image. After the feature vector (weight vector) for the test image have been found out, next step is to classify it. For the classification task we could simply use Euclidean distance classifier. If the distance is small, we say the images are similar and we can decide which the most similar image in the database.

**Step 9: Retrieving identical twins using PCA**

In this step, identical twins have been retrieved from whole database.

## IV. CONCLUSION

The use of face recognition in forensic applications is becoming more and more common, especially because when other biometric modalities may not be available. Law enforcement and security agencies around the world are using face recognition to detect fraud and to identify unknown individuals depicted in the act of committing crimes, even when fingerprints or DNA may not be left behind. Similarly civil programs, such as driving licensing and passport issuance, use face recognition to detect duplicate applicants because the face has long had social acceptance in identity credentials and because capture equipment is so widely available. When utilizing such biometric tools, however, it is important that miss-identifications be avoided to minimize - or eliminate - the chance of inadvertently implicating an innocent person. Face recognition system is expected to identify faces presented in images and videos automatically. Face verification (or authentication) and face identification (or recognition) are two important aspects of face recognition systems. Face verification is concerned with validating the claimed identity of a person based on the face image available, in other words it is considered as one to one match while face identification involved in identifying a person based on the face image by comparing it with all the registered persons or equivalently one to many matching.. So in this work, we proposed a method which uses scars, facial marks, moles etc. as features to differentiate identical twins in the image. As these features vary from person to person, it is easy to recognize them according to the location of these features in identical twin images. In this work, first of all, we automatically find the location of facial marks of each individual and they are matched with all identical twins which are in the database.

## ACKNOWLEDGMENT

Thanks to my family, my guide **Er. Navdeep Kaur** (AP, CSE Department), **Er. Harpreet Kaur** (AP & HOD, CSE Department), Bahra Group of Institute, Bhedpura, Patiala for having faith in me and allowing me to work with all terms and conditions and advising me time to time about my work.

## REFERENCES

[1] Shinde Anagha A. Ruikar Sachin D., ―Face Recognition using PCA and Eigen Face Approach," International Conference on Recent Trends in Engineering & Technology (ICRTET'2013), pp.: 7-12, 2013.

[2] Ho Jan-Ming,Lin Shu-Yu,Fann Chi-Wen,Wang Yu-Chun,Chang Ray-I, ―A Novel Content Based Image Retrieval System Using K-Means With Feature Extraction,‖ International Conference on Systems and Informatics , Yantai, pp.:785-790, May 2012.

[3] Iván González-Díaz, Member, IEEE, Carlos E. Baz-Hormigos, and Fernando Díaz-de-María, Member, IEEE, "A Generative Model for Concurrent Image Retrieval and ROI Segmentation,‖ IEEE Transactions on Multimedia, ISSN: 1520-9210, Issue No. 1, Vol. 16, pp.: 169-183, January 2014

[4] B.Poon , M. Ashraful Amin , Hong Yan "Performance evaluation and comparison of PCA Based human face recognition methods for distorted images" International Journal of Machine Learning and Cybernetics, Vol 2, No. 4,pp. 245–259, July 2011. Springer-Verlag 2011.

[5] R Gross, V Brajovic, "An image pre-processing algorithm for illumination invariant face recognition", 4th International Conference, AVBPA 2003.

[6] Lin, S.H., "An Introduction to Face Recognition Technology", Multimedia Informing Technologies, Vol. 3, No. 1, 2000, pp. 1-7

[7] Ramesha, K., Raja, K.B.,Venugopal, K.R. and Patnaik, L.M. "Feature Extraction Based Face Recognition, Gender and Age Classification," *International Journal on Computer Science and Engineering (IJCSE),* Vol. 02, Issue No. 01S, 2010, pp.14-23.

[8] Xu, Z., and Wu, H.R., "Shape Feature Based Extraction For Face Recognition", *IEEE*, 2009.

[9] A.Jain, "Fundamentals of Digital Image Processing" Englewood Cliffs, NJ:Prentice-Hall,1989