International Journal of Scientific Research in Computer Science, Engineering and Information Technology
 | Volume 3 | Issue 3 | ISSN : 2456-3307

Person Identification using Data Mining of Facial Features Extraction

Dr. (Mrs.) R. A. Ingolikar *1, Ms Shweta Shirpurkar2

*1 Head Computer Science, S.F.S. Nagpur, Maharashtra, India

2 Research Scholar, IICC, RTM university, Nagpur, Maharashtra, India

ABSTRACT

This paper deals with the person identification using data mining of facial features. As facial features play important role in identifying the person, they are extracted and used for data The geometric features such as distance between two eyes, distance between left eye and centre of nose, distance mining.between right eye and centre of nose, mouth length and lips portions are extracted. The random forest data mining algorithm is used for extracting face patterns. It reveals that there is a specific pattern of facial features for normal and abnormal persons. Which helps us to classify the person into two categories viz. normal and abnormal person.

Keywords: Person identification, facial feature extraction, random forest algorithm, data mining

I. INTRODUCTION

Automatic human face and facial feature extraction plays an important role in person identification in the areas of video surveillance, human computer interaction and access control. [1] To extract facial features we need first to detect human faces in images.

Face detection methods can be classified into four categories: knowledge-based, feature invariant, template matching and appearance- based. Gaussian clusters can be used to model the distribution of face and non-face patterns.

A complete facial image analysis system should be able to localize faces in a given image, identify and pin-point facial features, describe facial expressions, and recognize people.

Most facial expression analysis and face recognition systems work with the assumption that the location of the face within a frame is known. This assumption is suitable for scenes with a uniform background; however, in images with complex background, faces must be localized before any recognition can be performed. Though people are good at face identification, recognizing human face automatically by computer is very difficult. Face recognition has been widely applied in security system, credit-card verification, and criminal identifications, teleconference and so on. Face recognition is influenced by many complications, such as the differences of facial expression, the light directions of imaging, and the variety of posture, size and angle.

II. RELATED WORK

Even for the same person, the images taken in different surroundings may be unlike. The problem is so complicated that the achievement in the field of automatic face recognition by computer is not as satisfactory as the finger prints. Facial features extraction has become an important issue in automatic recognition of human faces. Detecting the basic features such as eyes, nose and mouth exactly is necessary for most face recognition methods.

Recently, techniques achieved in the research for detection of facial feature points can be broadly classified as:

(i) approaches based on luminance, chrominance, facial geometry and symmetry,

(ii) template matching based approaches , (iii) PCA- based approaches and the combination of the above approaches along with curvature analysis of the intensity surface of the face images [5] .

Also other facial feature detection approaches exist. Feris et al. presents a technique for facial feature localization using a two-level hierarchical wavelet network [6] .

In this paper, we propose an approach based on human visual characteristics, using the geometry and symmetry of faces, which can extract the features with properties of scale, translation and rotation invariance and locate the vital feature points on eyes, nose and mouth exactly.

This method needn't normalize the images to same size before processing. It can also help to improve the accuracy of face recognition.

III. FACE FEATURE EXTRACTION TECHNIQUES

In order to perceive and recognize human faces, we must extract the prominent characteristics on the faces. Usually those features like eyes, nose and mouth together with their geometry distribution and the shape of face is applied.

3.1 The feature points on human face

Applying human visual property in the recognition of faces, people can identify face from very far distance, even the details are vague. That means the symmetry characteristic is enough to be recognized. Human face is made up of eyes, nose, mouth and chin etc. There are differences in shape, size and structure of those organs, so the faces differ in thousands ways, and we can describe them with the shape and structure of the organs so as to recognize them. One common method is to extract the shape of the eyes, nose, mouth and chin, and then distinguish the faces by distance and scale of those organs (as shown in Fig.1). The other method is to use deformable model to describe the shape of the organs on face subsequently [7] .

In order to perceive and recognize human faces, we must extract the prominent characteristics on the faces. Usually those features like eyes, nose and mouth together with their geometry distribution and the shape of face is applied.

3.1 The feature points on human face

Applying human visual property in the recognition of faces, people can identify face from very far distance, even the details are vague. That means the symmetry characteristic is enough to be recognized. Human face is made up of eyes, nose, mouth and chin etc. There are differences in shape, size and structure of those organs, so the faces differ in thousands ways, and we can describe them with the shape and structure of the organs so as to recognize them. One common method is to extract the shape of the eyes, nose, mouth and chin, and then distinguish the faces by distance and scale of those organs (as shown in Fig.1). The other method is to use deformable model to describe the shape of the organs on face subsequently [7] .

Figure 1. Features on human face

We can tell the characteristics of the organs easily by locating the feature points from a face image. If we normalized the characteristics which have the properties of scale, translation and rotation invariance, we can normalize the faces in the database through geometric transformations.

Additionally, the selection of face feature points is crucial to the face recognition. We should pick up the feature points which represent the most important characteristics on the face and can be extracted easily.

The number of the feature points should take enough information and not should be too many. If the database has different postures of each person to be recognized, the property of angle invariance of the geometry characteristic is very important. This paper has presented a method to locate the vital feature points of face, which select 9 feature points that have the

property of angle invariance, including 2 eyeballs, 4 near and far corners of eyes, the midpoint of nostrils and 2 mouth corners, as shown in Fig.1. According to these, we can get other feature points extended by them and the characteristics of face organs which are related and useful to face recognition.

$$c(\ddot{r}, \ddot{r}_0) = \begin{cases} 1 & if \; |I(\ddot{r}) - I(\ddot{r}_0)| \le t \\ 0 & if \; |I(\ddot{r}) - I(\ddot{r}_0)| > t \end{cases} \quad (1)$$

$$n(\ddot{r}_0) = \sum_r c(\ddot{r}, \ddot{r}_0) \quad (2)$$

$$R(\ddot{r}_0) = \begin{cases} g - n(\ddot{r}_0) \; if & n(\ddot{r}_0) < g \\ 0 & otherwise \end{cases} \quad (3)$$

In this paper, we choose the operator SUSAN (Smallest Univalue Segment Assimilating Nucleus) to extract the edge and corner points of local feature area. The principle of operator SUSAN is to make a mask on the circle area of one point with the radius of r (we set r = 4 in this paper) and then observe every point in the whole image on the consistency of this point with all points contained in the mask area.

3.2 Principal Component Analysis

In the result of application of the PCA algorithm an original data of image is projected into a new coordinate space. Each coordinate axis in the new coordinate space will represent a principal component vector. The first principal component vector is the direction along which the variance is maximum; the second principal component vector is defined by the direction which maximizes the variance among all directions orthogonal to the first vector and so on. PCA algorithm includes the following steps

The first step is the reading of the face images from the database and converting them into gray-scale values. After these operations obtained 2D face images are converted into 1D image vector. The images are converted to represent each face image of dimensions NxN to single beam of dimensions N×N to single beam of dimensions N 2 x1. The data are stored in the T = [T α]

In the second step the mean of images of T vector is calculated Equation 1:

They are defined as :

$$S_B = \sum_c N_C (m_c - \bar{X})(m_c - \bar{X})^T \quad (4)$$

$$S_w = \sum_c \sum_{i \in C} (X_i - m_c)(X_i - m_c)^T \quad (5)$$

Where

- S_B is the “between classes scatter matrix” and
- S_w is the “within classes scatter matrix”

and

$$m_c = \frac{1}{N_C} \sum_{i \in C} X_i \quad (6)$$

$$\bar{X} = \frac{1}{N_C} \sum_i X_i = \frac{1}{N_C} \sum_C N_C m_C \quad (7)$$

Where, Nc is a number of cases in the class C

The total scatter S_T :

$$S_T = \sum_i (X_i - \bar{X})(X_i - \bar{X})^T \quad (8)$$

Is given by Equation (9)

$$S_T = S_W + S_b \quad (9)$$

General view of a frontal face image containing a mouth and two eyes is shown
in Figure 2.

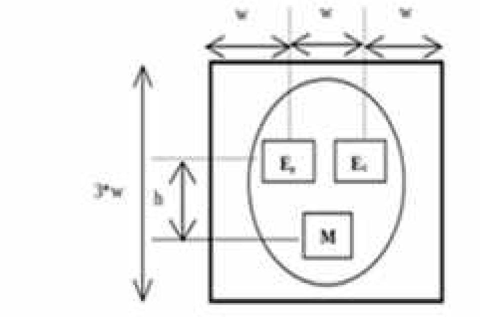

Figure 2. Geometric face features

E_l and E_R represent left and right eyes respectively, while M represents the mouth feature. The distance between the two eyes is w, and the distance from the mouth to the eyes is *h*. In frontal face images, structural relationships such as the Euclidean distance between the mouth, and the left and right eye, the angle between the eyes and the mouth, provide useful information about the appearance of a face. These structural relationships of the facial features are generally useful to constrain the facial feature detection process. A search area represented by the square of size (3w x 3w) is also an important consideration in order to search for faces based on the detected eye feature positions in the image.

3.3 Random Forests

Random forests is a idea of the general technique of random decision forests that are an ensemble learning technique for classification, regression and other tasks, that is controlled by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forests decision are known for its accuracy.

The first algorithm for random decision forests was created by Tin Kam Ho using the random subspace method which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg. An extension of the algorithm was developed by Leo Breiman[7] and Adele Cutler,[8] and "Random Forests" is their trademark[9].The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[10] in order to construct a collection of decision trees with controlled variance.

3.4. Random Tree

Random Tree is a supervised Classifier; it is an ensemble learning algorithm that generates lots of individual learners. It employs a bagging idea to construct a random set of data for constructing a decision tree. In standard tree every node is split using the best split among all variables. In a random forest, every node is split using the best among the subset of predicators randomly chosen at that node. Random trees have been introduced by Leo Breiman and Adele Cutler.The algorithm can deal with both classification
and regression problems. A Random tree is a group (ensemble) of tree predictors that is called forest. The classification mechanism is as follows: the random trees classifier gets the input feature vector, classifies it with every tree in the forest, and outputs the class label.

IV. PROPOSED WORK

In this section, the system analysis and design of the implementation of biometric using data mining techniques for identification of person is discussed. The research work begins with the rigorous literature review of the research work conducted by various researchers in this field. In system implementation point of view, there are four modules viz. Pre-processing, Facial features extraction, classification of images using data mining algorithm and person identification.

4.1 Image dataset :

The image dataset is consisting of 650 face images of children. This dataset will be used for the research work.

4.2 Pre-processing:

The images in the image dataset are having varied resolutions and they are all colored images. The images are resized into a uniform resolution of 400 x 300 and are converted into gray level format.

Figure 3. a)Color image b) Gray image c)Binary image

This process is called pre-processing. This is the important requirement of this image processing based research work.

4.3 Facial Features Extraction:

Facial features extraction deals with extraction of various facial features viz. distance between two eyes, length of nose, width of mouth etc. These features are extracted from the binary image by performing the following operations:

- ✓ Histogram equalization
- ✓ Morphological operations

4.4 Histogram Equalization :

Histogram equalization distributes the pixel intensity uniformly across the given range of pixel intensity values. Figure 2 depicts the histogram of face image.

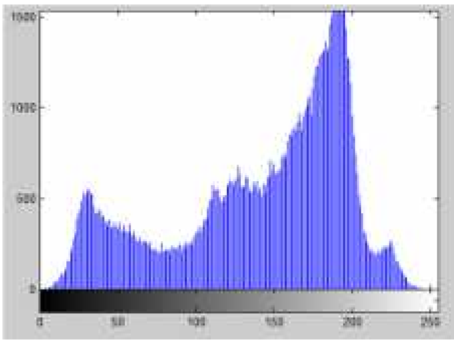

Figure 4. Histogram

For detecting the objects in the image we need to find its image histogram. Which is used to determine in which row and column the pixels containing the object are belonging to.

4.5 Morphological Operations:

Morphological operations such as edge detection is used for finding edges/boundaries of various objects of an image. We have used edge detection algorithm with 'sobel' operator for finding the edges.

4.6 The algorithm

The random forests algorithm (for both classification and regression) is as follows:

1. Draw n_{tree} bootstrap samples from the original data.
2. For each of the boot strap samples, grow an un-pruned classification or regression tree.
3. At each node, randomly sample mtry of the predictors and choose the best split from among those variables. (Bagging can be thought of as the special case of random forest algorithm for bests obtained when $m_{try}=p$, the number of predictors.)
4. Predict new data by aggregating the predictions of the n_{tree} trees (i.e., majority votes for classification, average for regression).

An estimate of the error rate can be obtained, based on the training data, by the
following:

1. At each bootstrap iteration predict the data not in the bootstrap sample using the tree grown with the bootstrap sample.

2. Aggregate the Out of Bag (OOB) predictions. (On the average, each data point would be out-of-bag around 36% of the times, so aggregate these predictions.)
3. Calculate the error rate, and call it the Out of Bag estimate of error rate.

Algorithm of Proposed Work

Input :

TrImg : Training Images

TsImg : Testing Images

N_1 : No. of training Images

N_2 : No. Of testing Images

Output :

ClImg : Classified Images

Method :

1) Perform Preprocessing of image data sets
2) Perform Facial Feature Extraction with following techniques :
   - Histogram Equalization
   - Morphological Operations
3) Create Training Feature vector matrix
4) Classify Test Images using Random Forest Algorithm
5) Display result of classification

Figure 5. Edge detection using 'Sobel'

4.7 Feature Extraction :

Features of face such as f1= distance between the eyes, f2= distance between left eye and centre of nose, f3=distance between right eye and centre of nose and f4 = width of mouth are determined by using geometric transformations.

Figure 6. Facial Features

Some of the extracted features of abnormal objects are shown below

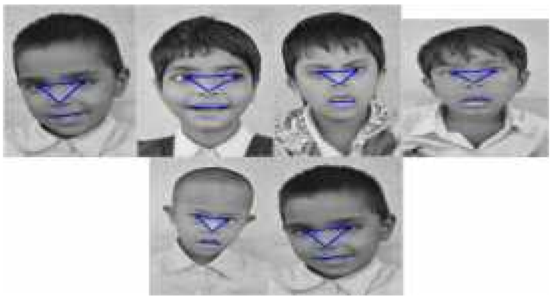

Figure 7. Features of abnormal children

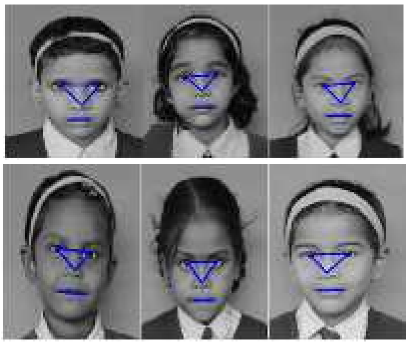

Figure 8. Features of Normal children

V. RESULTS AND DISCUSSION

Experimentation is carried out on different test images. Table I shows the result of human identification. We have displayed the values of various geometric features of face images. Symbol A represents Abnormal face and N represents Normal face.

Table 1. Result of Human Identification

Face Image	Distance between the eyes	Distance between left eye and centre of nose	Distance between left eye and centre of nose	Width of mouth	Result
1	75	46	50	45	A
2	80	47	63	40	A
3	90	50	41	42	A
4	84	43	39	48	A
5	61	45	47	49	N
6	85	38	45	40	A
7	36	35	50	36	A
8	62	42	60	40	N
9	63	45	51	45	N
10	56	49	49	39	A
11	60	44	40	48	N
12	32	41	48	40	A
13	27	44	36	42	A
14	60	45	45	40	N
15	60	42	50	36	N
16	63	42	49	42	N
17	40	30	50	40	A
18	55	40	44	33	A
19	64	40	48	40	N
20	64	48	48	36	A
21	87	39	39	39	A
22	63	41	45	42	N
23	60	40	50	40	N
24	65	42	42	42	A
25	60	45	45	41	A
26	64	64	48	48	A
27	71	51	37	33	A
28	61	44	44	33	N
29	60	40	48	36	N
30	61	42	39	39	N
31	68	62	42	38	A
32	57	57	39	39	A
33	61	42	42	42	N
34	80	40	48	40	A
35	60	45	45	41	N
36	64	40	48	48	N
37	61	41	37	33	N
38	62	42	42	38	N
39	63	41	39	39	N
40	60	40	48	40	N

Results are evaluated on the basis of various parameters such as TP, FP,TN and FN. Then Precision and accuracy of person identification techniques is determined and it is depicted in Figure 9.

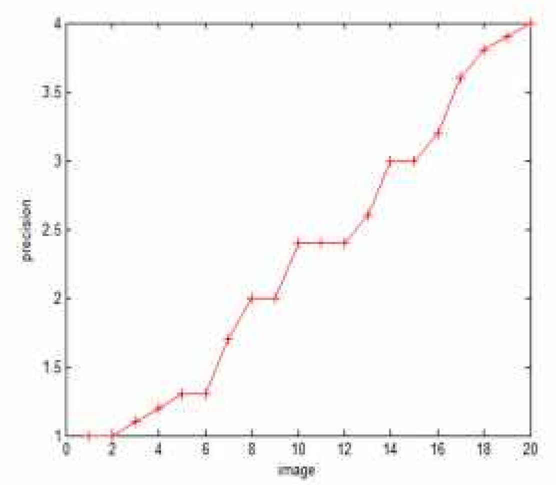

Figure 9. Precision

VI. CONCLUSION

In this work we have provided an approach towards person identification using biometric and data mining algorithm such as random forest. We have carried out the experimentation of 650 images and classified the persons into normal and abnormal categories.

VII. REFERENCES

[1]. Faisal Rehman,M. Usman Akram,Kunwar Faraz and Naveed Riaz,"Human identification using dental biometric analysis,

[2]. Anitha L.,Arunvinodh C. & Dhanya K.K. "Biometric system using graph matching",IEEE March 2015

[3]. Namrata Srivastava,Utkarsh Agrawal,Soumava Kumar Roy and U.S. Tiwary,"Human identification using Linear Multiclass SVM and Eye Movement Biometrics", IEEE paper,August 2015

[4]. Gil Santos,Paulo T. Fiadeiro & Hugo Proenca,"BioHDD: a dataset for studying biometric identification on heavily degraded data IEEE paper,March 2015

[5]. Daniel A. Reid,Mark S. Nixon,Sarah V. Stevenage, "Soft Biometrics; Human Identification Using Comparative Descriptions IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 36,Issue: 6,June 2014)

[6]. Han,J. & Kamber,M. (2002). Data mining Concepts and Techniques,Morgan Kaufman Publishers,ISBN 1-55860-489-8,CA,USA.

[7]. Data Mining Techniques: Michael J. Berry,Gordon Linoff International Journal of Computer Technology and Electronics Engineering (IJCTEE)

[8]. Ming-Syan Chen,Jiawei Han,Philip S yu. Data Mining: An Overview from a Database PerspectiveJ]. IEEE Transactions on Knowledge and Data Engineering,l996,8(6):866-883.

[9]. R Agrawal ,T 1 mielinski,A Swami. Database Mining: A Performance PerspectiveJ]• IEEE Transactions on Knowledge and Data Engineering,1993,12:914-925.

[10]. Fayyad,Usama; Gregory Piatetsky-Shapiro,and Padhraic Smyth

[11]. "Data mining tools",by Ralf Mikut,Markus Reischl,Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,2011

[12]. Burge,M. and Burger,W. Ear Biometrics. IA. Jain R. Bolle and S. Pankanti,editors,BIOMETRICS: Personal Identification in a Networked Society,pp. 273-286. Kluwer Academic,1998.

[13]. Burge,M. and Burger,W. Ear Biometrics in Computer Vision. In the 15th International Conference of Pattern Recognition,ICPR 2000,pp. 826-830. Carreira-Perpinan,M.A. Abstract from MSc thesis Compression neural networks for feature

[14]. Hui Chen,Bir Bhanu,"Human Ear Recognition in 3D," IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 29,no. 4,pp. 718-737,Apr. 2007,doi:10.1109/TPAMI.2007.1005

[15]. Data Mining Approaches for Intrusion Detection Wenke Lee Salvatore J. Stolf

Computer Science DepartmentColumbia University500 West 120th Street,New York,NY 10027

[16]. Biometric Data Mining Applied to On-line Recognition Systems Jose Alberto Hernandez-Aguilar[1],Crispin Zavala[1], Ocotlan Díaz[1],Gennadiy Burlak[2], Alberto Ochoa[3] and Julio Cesar Ponce4 [1]FCAeI-UAEM.

[17]. EAR BIOMETRICS Hanna-Kaisa Lammi Lappeenranta University of Technology, Department of Information Technology, Laboratory of Information Processing,P.O. BOX 20,53851