Data Mining of Facial Features using Random Forest Algorithm for Person Classification

[1]Ms. Shweta Shirpurkar, [2]Dr.Mrs. R.A. Ingolikar

[1]ResearchScholor, [2]Head of the department
[1]Computer Science,
[1]IICC, RTM university, Nagpur, India

Abstract: **This paper deals with the person classification using data mining of facial features using Random forest algorithm. The geometric features such as distance between two eyes (D_{eyes}), distance between left eye and centre of nose (D_{LN}), distance between right eye and centre of nose (D_{RN}), mouth length (M_L) and lips (D_{Lips}) portions are extracted. The mathematical model of these extracted features is formed. The extracted features will be taken as input variables of the decision trees. The forest of various decision trees is created. Further, random forest classification algorithm is used for classification of the face patterns. Growing an ensemble of decision trees and allowing them to vote for the most popular class, can produce significant increase in classification accuracy for person classification.**

IndexTerms **- Person classification, facial feature extraction, random forest algorithm, data mining**

I. INTRODUCTION

Classification of persons based on their facial expressions is the key concerns of many researchers during last two decades. Different approaches have been exploited by various researchers for automatic human classification. 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]

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges as to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates that compare favourably to Adaboost but are more robust with respect to noise. Internal estimates monitor error, strength, and correlation and these are used to show the response to increasing the number of features used in the splitting. Internal estimates are also used to measure variable importance. These ideas are also applicable to regression.

This classifier involves in choosing a set of features randomly and creating a classifier with a bootstrapped sample of the training data. A large number of trees (classifiers) are generated in this way and finally un-weighted voting is used to assign an unknown pixel to a class. Further, the performance of the random forest classifier is compared with support vector machines in term of classification accuracy, training time and user-defined parameters.

1.1 Random Forest Classifier

The random forest classifier consists of a combination of tree classifiers where each classifier is generated using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector (Breiman, 1999). The random forest classifier used for this study consists of using randomly selected features or a combination of features at each node to grow a tree. Bagging, a method to generate a training data set by randomly drawing with replacement *N* examples, where *N* is the size of the original training set (Breiman, 1996), was used for each feature/feature combination selected. Any examples (pixel) are classified by taking the most popular voted class from all the tree predictors in the forest (Breiman, 1999). Design of a decision tree required the choice of an attribute selection measure and a pruning method. There are many approaches to the selection of attributes used for decision tree induction and most approaches assign a quality measure directly to the attribute. The most frequently used attribute selection measures in decision tree induction are Information Gain Ratio criterion (Quinlan1993) and Gini Index (Brieman et. al., 1984). The random forest classifier uses the Gini Index as an attribute selection measure, which measures the impurity of an attribute with respect to the classes. For a given training set T, selecting one case (pixel) at random and saying that it belongs to some class C_i, the Gini index can be written as:

$$\sum_{j \neq i} \sum (f(C_i, \mathrm{T})/|\mathrm{T}|)(f(C_j, \mathrm{T})/|\mathrm{T}|) \qquad (1)$$

where f (C_i , T)/ |T| is the probability that the selected case belongs to class C_i .

Each time a tree is grown to the maximum depth on new training data using a combination of features. These full-grown trees are not pruned. This is one of the major advantages of the random forest classifier over other decision tree methods like the one proposed by Quinlan (1993). As the studies suggest that the choice of the pruning methods, and not the attribute selection measures, affect the performance of tree based classifiers (Mingers, 1989; Pal and Mather, 2003a). Breiman (1999) suggests that as the number of trees increases, the generalisation error always converges even without pruning the tree and over-fitting is not a

problem because of the Strong Law of Large Numbers (Feller, 1968). The number of features used at each node to generate a tree and the number of trees to be grown are two user-defined parameters required to generate a random forest classifier. At each node, only selected features are searched for the best split. Thus, the random forest classifier consists of *N* trees, where *N* is the number of trees to be grown which can be any value defined by the user. To classify a new data set, each case of the data sets is passed down to each of the *N* trees. The forest chooses a class having the most out of *N* votes, for that case.

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 satisfied 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 researches 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] .

Biometrics is measurable characteristics specific to an individual. Face detection has diverse applications especially as an identification solution which can meet the crying needs in security areas. A. Dhanalakshmi *et al*. [2] presented use of measurable characteristics of persons. The traditionally 2D images of faces have been used, 3D scans that contain both 3D data and registered color are becoming easier to acquire. Before 3D face images can be used to identify an individual, they require some form of initial alignment information, typically based on facial feature locations. We follow this by a discussion of the algorithms performance when constrained to frontal images and an analysis of its performance on a more complex dataset with significant head pose variation using 3D face data for detection provides a promising route to improved performance. [2]

A biometric system which primarily based on the cues of unimodal biometric for individual identification is not always meet the desired results. The concept of multimodal biometrics for human Identification is an emerging trend. Radhey Shyam *et al.*[5] present state-of-the-art novel multimodal biometric system, for face recognition, which combines the similarity scores of the unimodal modalities such as appearance based and texture based techniques of face recognition, to cater the decisive results at the level of matching score. Formally, it includes the fusion of unimodal techniques to devise the multimodal models in four possible combinations such as (a) Eigen faces and local binary pattern (LBP), (b) Fisher faces and LBP, (c) organics' and augmented local binary pattern (A-LBP), and (d) Fisher faces and A-LBP. The performance of the multimodal face recognition systems is tested on the publicly available face databases such as the AT & T-ORL and the Labeled Faces in the Wild (LFW) using a new Bray Curtis dissimilarity metric. The experimental results show a significant improvement in the performance of recognition accuracies of multimodal face recognition techniques. [5]

Biometric systems are becoming increasingly important, as they provide more reliable and efficient means of identity verification. Human identification at a distance has recently gained enormous interest among computer vision researchers. Gait recognition aims essentially to address this problem by recognising people based on the way they walk. In this paper, we propose an efficient self-similarity based gait recognition system for human identification using modified Independent Component Analysis (MICA). Initially the background modelling is done from a video sequence. Subsequently, the moving foreground objects in the individual image frames are segmented using the background subtraction algorithm. Then, the morphological skeleton operator is used to track the moving silhouettes of a walking figure. The MICA based on eigen space transformation is then trained using the sequence of silhouette images. Finally, when a video sequence is fed, the proposed system recognizes the gait features and thereby humans, based on self-similarity measure. The proposed system is evaluated using gait databases and the experimentation on outdoor video sequences demonstrates that the proposed algorithm achieves a pleasing recognition performance.

III. CLASSIFICATION USING RANDOM FORESTALGORITHM

As random forest classifier is ensemble algorithm, we start with construction of a decision tree. This is a standard machine learning technique. The decision tree, in ensemble terms, corresponds to our weak learner. In a decision tree, an input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets. In the proposed work, the random forest algorithm is used for classification of persons into normal and abnormal categories. The process starts with the image dataset of normal and abnormal faces comprising of 650 samples. The step-wise procedure for classification process is given below.

1) Pre-processing of image dataset
2) Facial Feature Extraction
3) Formation of subsets of extracted features
4) Formation of Decision trees & forest

5) Traversing the random forest for classification

3.1 Pre-processing of image dataset:

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. This process is called pre-processing



Figure1: A)Color image b) Gray image c)Binary image

3.2 Facial Features Extraction:

The face image is a 3-D matrix of intensity values or gray level values lying in the range 0 to 255. We have used RGB color model for this work. In RGB color model, each color space is a 2-D matrix of intensity values and it is represented as,

$$I=f(x,y) \quad \text{for } 0 \le x \le 255 \text{ and } 0 \le y \le 255 \qquad (2)$$

Facial features extraction deals with extraction of various facial features viz. distance between two eyes, length of nose, width of mouth etc.

Let $p[x_1, y_1]$ → centre of left eye ball

$q[x_2, y_2,]$ → centre of left eye ball

$r[x_3,y_3]$ → mid-point of nose

$s[x_4,y_4]$ → left end of mouth and

$t[x_5,y_5]$ → right end of mouth

p,q,r,s and *t* be the points representing centre of left eye ball, centre of left eye ball , mid-point of nose , left end of mouth and right end of mouth respectively as shown in Figure 2.

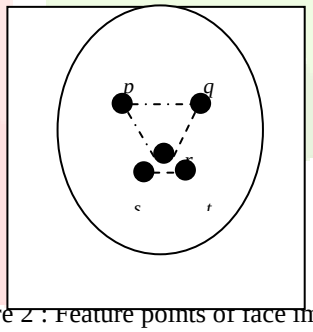


Figure 2 : Feature points of face image

Let the features of face images are :

D_{eyes} → distance between two eyes

D_{LN} →distance between left eye and centre of nose,

D_{RN} → distance between right eye and centre of nose

M_L → mouth length

D_{Lips}→ thickness of lips

$$D_{eyes} = |x1 - x2| \qquad (3)$$

$$D_{LN} = |y1 - y3| \qquad (4)$$

$$D_{RN} = |y2 - y3| \qquad (5)$$

$$M_L = |x4 - x5| \qquad (6)$$

$$D_{Lips} = |y6 - y7| \qquad (7)$$

Let $X1=D_{eyes}$, $X2=D_{LN}$, $X3=D_{RN}$, $X4= M_L$ and $X5= D_{Lips}$. Figure 3 shows the extracted features of a child.

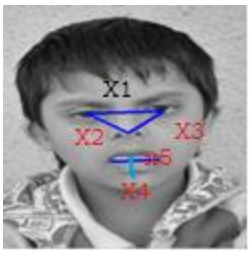


Figure 3 : Facial features of a child

Similarly, features of all the database images are extracted to form feature matrix. Some of features of normal and abnormal children are shown in Figure 4 and Figure 5 respectively.

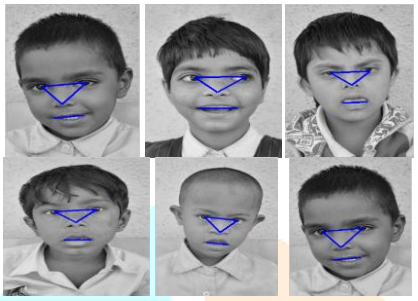


Figure 4 : Features of abnormal children

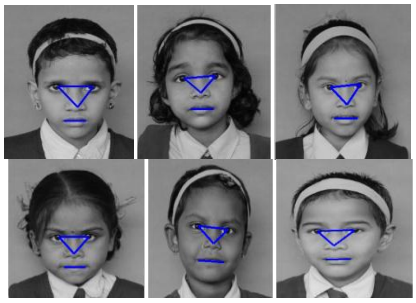


Figure 5 : Features of Normal children

Numeric values of some of the extracted features of abnormal and normal objects are shown in Table 1 which is given below. The values given in the table helps us to determine the range of values of each parameter for abnormal and normal person.

Table I : Extracted Features of normal and abnormal faces

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
1	75	46	50	45
2	80	47	63	40
3	90	50	41	42
4	84	43	39	48
5	61	45	47	49
6	85	38	45	40
7	36	35	50	36
8	62	42	60	40
9	63	45	51	45
10	56	49	49	39
11	60	44	40	48
12	32	41	48	40
13	27	44	36	42
14	60	45	45	40
15	60	42	50	36
16	63	42	49	42
17	40	30	50	40
18	55	40	44	33
19	64	40	48	40
20	64	48	48	36
21	87	39	39	39
22	63	41	45	42
23	60	40	50	40
24	65	42	42	42

25	60	45	45	41
26	64	64	48	48
27	71	51	37	33
28	61	44	44	33
29	60	40	48	36
30	61	42	39	39
31	68	62	42	38
32	57	57	39	39
33	61	42	42	42
34	80	40	48	40
35	60	45	45	41
36	64	40	48	48
37	61	41	37	33
38	62	42	42	38
39	63	41	39	39
40	60	40	48	40

3.3 Formation of decision trees

The input of the decision tree consists of a training set which is given as : [X1, X2, X3, X4,X5] with corresponding labels as [L1, L2, L3, L4,L5], From the observation of values from Table I, it has been found that we can form four different combinations or subsets of feature values. The geometric facial features of normal and abnormal persons are tend to form specific pattern. To check the deformity in the face, these four patterns can be observed and examined.

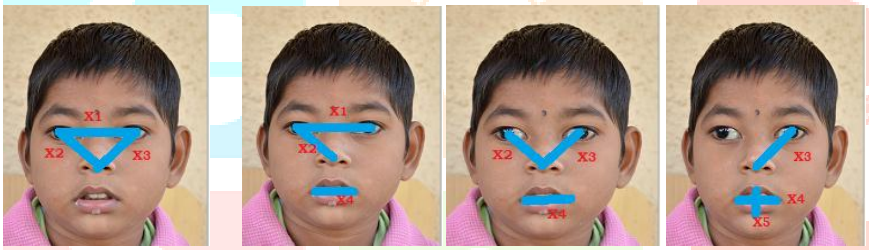


Figure 6 : Subset formation for the facial features.

The Random forest may create three decision trees taking input of subset for example,

1. [X1, X2, X3]
2. [X1, X2, X4]
3. [X2, X3, X4]
4. [X3,X4,X5]

On the basis of values evaluated for x1,x2,x3,x4,x5 during feature extraction the range can be decided as {60-72},{40-60},{38-50},{30-50}, {[20-30], [30-50]} respectively and on the basis of these ranges the following decision trees are formed.

The split at each decision node will be decided on the basis some machine learning rules given as follows. Experimentation is carried out on different test images

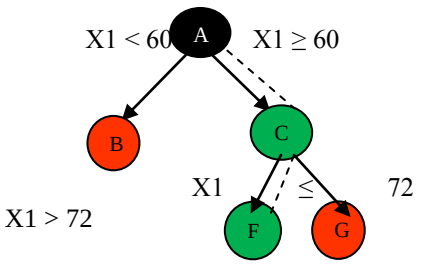


Fig 7 : Decision tree for X1

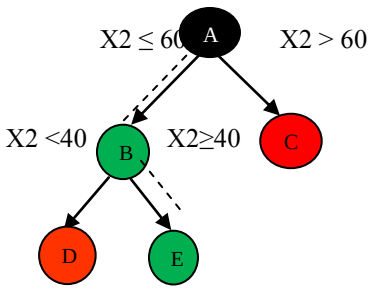


Fig 8 : Decision tree for X2

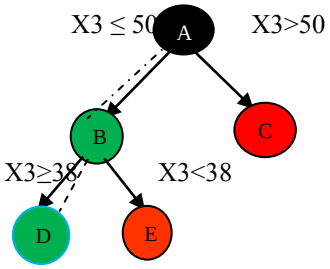


Fig 9 : Decision tree for X3

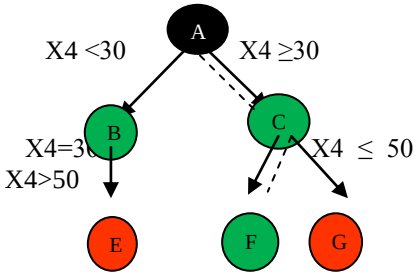


Fig 10 : Decision tree for X4

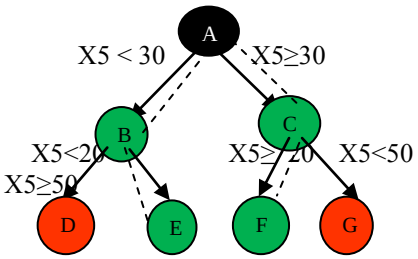


Fig 11 : Decision tree for X5

The decision paths :

P1: A-C-F

P2 : A-B-E

P3 : A-B-D

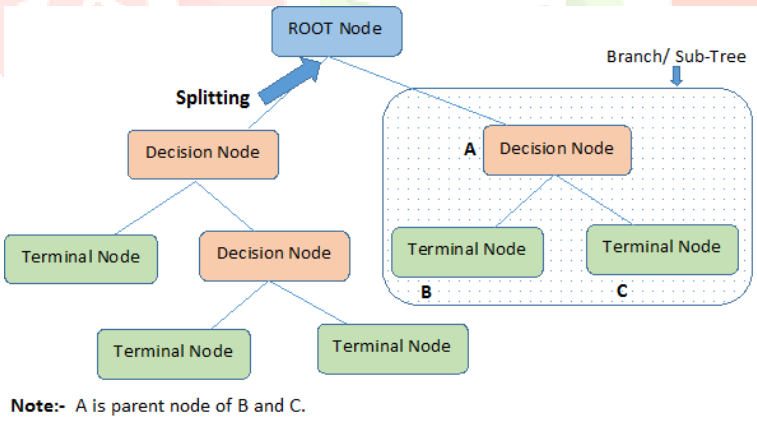
P4 : A-C-F

P5 : A-B-E/A-C-F

Figure 12 : decision paths

The types of nodes used in Decision trees is given below:

1. **Root Node:** It represents entire population or sample and this further gets divided into two or more homogeneous sets. We have represented the root node by black color.
2. **Splitting:** It is a process of dividing a node into two or more sub-nodes.
3. **Decision Node:** When a sub-node splits into further sub-nodes, then it is called decision node. We have represented the decision nodes with green color
4. **Leaf/ Terminal Node:** Nodes do not split is called Leaf or Terminal node. The leaf nodes are represented in red color.



5. **Pruning:** When we remove sub-nodes of a decision node, this process is called pruning. You can say opposite process of splitting.
6. **Branch / Sub-Tree:** A sub section of entire tree is called branch or sub-tree.
7. **Parent and Child Node:** A node, which is divided into sub-nodes is called parent node of sub-nodes where as sub-nodes are the child of parent node.

3.4 Traversing the random forest for classification

The algorithm selection is also based on type of target variables. Let's look at the four most commonly used algorithms in decision tree:

Gini Index

Gini index says, if we select two items from a population at random then they must be of same class and probability for this is 1 if population is pure.

1. It works with categorical target variable "Success" or "Failure".
2. It performs only Binary splits
3. Higher the value of Gini higher the homogeneity.
4. CART (Classification and Regression Tree) uses Gini method to create binary splits.

Steps to Calculate Gini for a split

1. Calculate Gini for sub-nodes, using formula sum of square of probability for success and failure (p^2+q^2).
2. Calculate Gini for split using weighted Gini score of each node of that split

The five decisions tree are traversed by using Gini index for split. The solution paths are P1,P2,P3,P4 and P5 as shown in figure in 12.

IV. RESULTS AND DISCUSSION

The extracted features have been taken as input variables of the decision trees. The forest of various decision trees is created. Further, random forest classification algorithm is used for classification of the face patterns. The result of classification is given below.

Table II : Classification result

Face Image	Classified as
1	Abnormal
2	Abnormal
3	Abnormal
4	Abnormal
5	Normal
6	Abnormal
7	Abnormal
8	Normal
9	Normal
10	Abnormal
11	Normal
12	Abnormal
13	Abnormal
14	Normal
15	Normal
16	Normal
17	Abnormal
18	Abnormal
19	Normal
20	Abnormal
21	Abnormal
22	Normal
23	Normal
24	Abnormal
25	Abnormal
26	Abnormal
27	Abnormal
28	Normal
29	Normal
30	Normal
31	Abnormal
32	Abnormal
33	Normal
34	Abnormal
35	Normal

36	Normal
37	Normal
38	Normal
39	Normal
40	Normal

V. CONCLUSION

In this work, we have provided an approach towards person classification using random forest algorithm. The modelling of user dataset is discussed followed by the formation of ensemble of decision trees. As decision tree are based on the rule-based system, the result of classification can be definitely improved. We have carried out the experimentation of 650 images and classified the persons into normal and abnormal categories. The classification of face images if involves the decision tree traversals by using the path [p1,p2,p3]or[p1,p2,p4] or[p1,p2,p4] or [p3,p4,p5] then face images are classified as normal otherwise abnormal.

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 PROENÇA, "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 Perspective[J]. IEEE Transactions on Knowledge and Data Engineering, l996, 8(6):866-883.

[9] R Agrawal ,T 1 mielinski, A Swami. Database Mining: A Performance Perspective[J]· 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 José Alberto Hernández-Aguilar[1], Crispin Zavala[1], Ocotlán Díaz[1], Gennadiy Burlak[2], Alberto Ochoa[3] and Julio César 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