

# Utilizing Viola Jones with Haar Cascade Along with Neural Networks for Face Detection and Recognition

Karan Arora<sup>1</sup>, Sarthak Arora<sup>2</sup>

<sup>1</sup>Department of Computer Science, Chitkara University, Rajpura, Punjab, India

<sup>2</sup>Department of IT, Maharaja Agrasen Institute of Technology, IPU, New Delhi, India

## ABSTRACT

**Viola-Jones object detection framework** introduced in 2001 by Dr. Paul Viola and Dr. Michael Jones is an object detection framework which can be trained for detecting a variety of object classes. It is primarily used for the problem of face detection. In most video recording or surveillance systems it became impossible for human beings to retrieve large image datasets and analyze them for potential results. Now-a-days accurate facial recognition has a great impact in our ecosystem be it face unlock or face recognition in cameras for auto adjust. Implemented in two stages our proposed methodology will first utilize one of the widely accepted methods to detect faces i.e. Viola Jones which utilizes Haar Classifiers and in the second stage we will recognize the face using Principal Component Analysis (PCA) and Feed Forward Neural Network. Bio ID-Face-Database is used as a training database. Test is conducted on webcam video and image snapshots.

**Keywords:** haar-features, Viola Jones, Image Analysis, face detection, feature extraction, face edge detection

**How to cite this paper:** Karan Arora | Sarthak Arora "Utilizing Viola Jones with Haar Cascade Along with Neural Networks for Face Detection and Recognition"

Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-1, December 2020, pp.284-291, URL: www.ijtsrd.com/papers/ijtsrd35848.pdf



Copyright © 2020 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



## 1. INTRODUCTION

Face detection due to its wide application in computer vision and also in image processing techniques plays a vital role in human to computer interaction. The recent advancements in Video processing, Image Compressing High Rate Frame Rendering facilitated diversified domains to utilize face detection and recognition techniques. It also made possible for us to avail the latest technology in daily operations like blazing face unlock. The process of correctly recognising a human face is a tough task as it exhibits multiple varying attributes like expressions, age, change in hairstyle etc. Though technology has grown but still it challenges many aspects in image processing such as blurry face detection and human-animal confusion. The challenge occurs because of multiple layers to filter the images or editing the image generally makes the face incomprehensible.

Now-a-days Face recognition is used in various domains and has multiple applications such as security systems, credit card verification, identifying criminals in airports, railway stations etc. Though various methods are researched on to detect and recognize a human face, developing a subtle model for a big database is still a challenging task. That is the reason face recognition is taken as a high level computer vision challenge to achieve accurate results multiple methods can be developed.

Few methods known for face recognition are group based tree neural networks, artificial neural networks (ANN) and principal component analysis (PCA).

The proposed methodology is executed in two phases - As the face can be characterized by special facial features, The first step will be to extract those features. Then they are quantized making it easier to recognize a face by referring to those features. For detection we will use Viola-Jones algorithm which works on Haar Cascades and we also used AdaBoost classifier as a modifier. The next step is to recognize the face for which we used (PCA) principal component analysis along with artificial neural networks (ANN). The aim of the paper is to use the methodology mentioned to detect and recognise a face from the database and then on the test set and webcam outputs.

## 2. RELATED WORK

A strategy that enhances the recognition rate as compared to PCA was introduced by Muhammad Murtaza Khan et al.,[8] which was outperformed by sub-Holistic PCA in all test scenarios with 90% recognition rate registered for the ORL database.

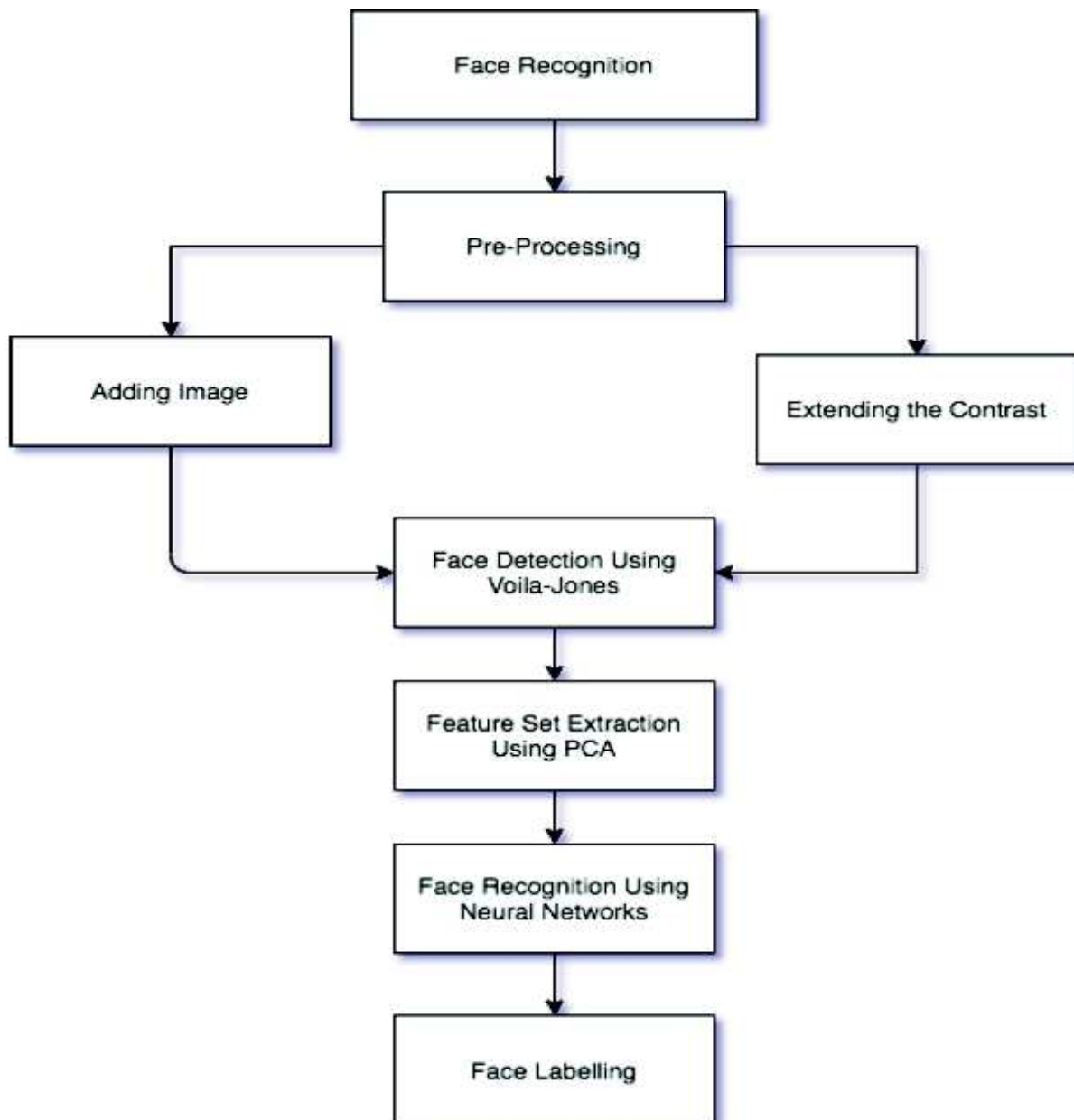
One method for face recognition based on preprocessing the face images was introduced by Patrik Kamencay et al.,[4] that used the Propagation Belief segmentation algorithm. The positive effect for face recognition was depicted by the algorithm with a face recognition rate of 84% for the ESSEX database. The use of linear and non linear techniques for feature extraction in face recognition was proposed by Hala M. Ebied et al., [5]. In the paper a high-dimensional feature space is mapped with nonlinear methods represented via

extension of Kernel-PCA from PCA, for the classification step K-nearest neighbor classifier with Euclidean distance is used. The SIFT-PCA method was proposed by Patrik Kamencay et al., [6] which implemented an impact of graph based segmentation algorithm on the recognition rate. SIFT related segmentation algorithms are used for preprocessing of the face images. The results depict a positive effect for face recognition for segmentation in combination with SIFT-PCA. A NP-hard problem of searching the best subnet of the available PCA features for face recognition is solved in the methodology proposed by Rammohan Mallipeddi et al.,[7].

The proposed method uses the differential equation algorithm called FS-DE. After maximizing the class separation in training data a feature set is obtained, further presenting an ensemble base for face recognition. A study of modified constructive training algorithms for Multi Layer Perceptron is proposed by Hayet Boughrara et al.,[8], which is applied to applications in face recognition. This paper contributed to depict the methods to simultaneously increase the output neurons with increasing input patterns. Perceived Facial Image is applied for feature extraction.

### 3. PROPOSED METHODOLOGY

We propose a robust methodology that is independent of facial variations like size, texture, feature position, facial expression etc using Viola-Jones, Principal component analysis and Neural Networks. Please refer to the flowchart stating the same in figure 1



**Figure 1: Flowchart**

#### 3.1. Data

Standard Data used is the BIOID face database. The dataset consists of 1521 gray scale images with 384\*286 pixel resolution. The front angle view images in the database consists of a face of 23 different persons. We have the test set with images with a variety of face size, lighting, background representing real life scenarios as shown in fig 2.



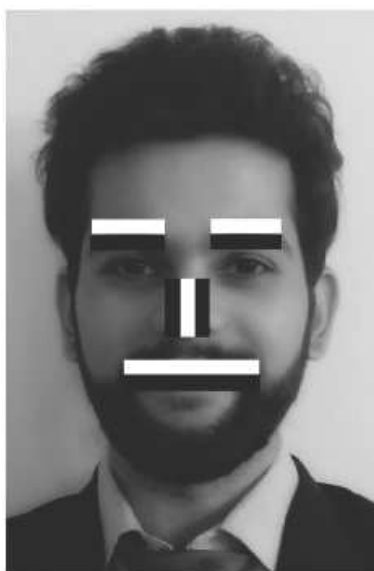
**Figure 2: Example set - Bio ID Face Database**

**PRE-PROCESSING:**

We considered an image database that is readily available in either gray scale or color. Contrast stretching is performed on the current image where white pixels were made whiter and dark pixels are made darker.

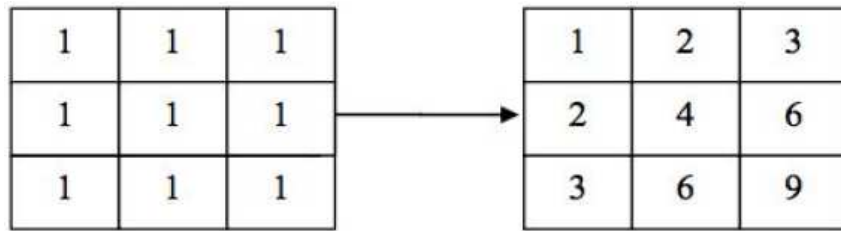
**FACE DETECTION:**

Right after Contrast-Stretching, Viola-Jones algorithm is utilized for detecting face in the image. We chose Viola-Jones detector as a detection algorithm because of its accuracy in detecting faces, and its ability to run in real time. The Viola-Jones detector works best with frontal images of faces and it can handle 45° face rotation both around the horizontal and vertical axis. There are three main concepts which allow it to run in real time first is Integral Image, Second is Ada-Boost and third is cascade structure Integral Image is a cost-effective generation algorithm that works on the sum of pixel intensities in a specified rectangle in an image. The main use-case is rapid computation of Haar-like features. The calculation done on the sum of a rectangular area inside the original image is extremely efficient for the initial step, requiring only four additions for any arbitrary rectangle size. The use of Ada-Boost is that it constructs strong classifiers as a linear combination of weak classifiers. voila-Jones uses Haar features, Haar Features used in the Viola Jones algorithm is shown in Fig 3.



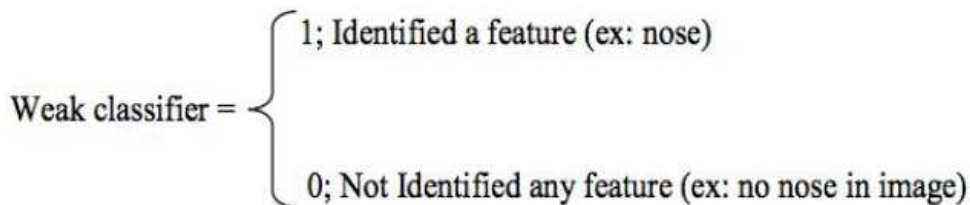
**Fig 3: Representing features with Haar Features**

Haar features shown above can be of various height and width. The working for the calculation for value is that from the applied Haar feature to the face, we calculate the sum of white pixels and also the sum of black pixels then it is subtracted to get a single value. If in the region the value is high, then it takes part of the face and is identified as eyes, nose, cheek etc. We have approximately 160000+ Haar features calculated all over per image. In real time application Summing up the entire image pixel and after that subtracting them to get a single value is not much efficient, which can be reduced by using the Ada-boost classifier. The major function of Ada-boost is reducing the redundant features. Integral image is used instead of summing up all the pixels as shown in figure 4.



**Fig 4: Integral Image**

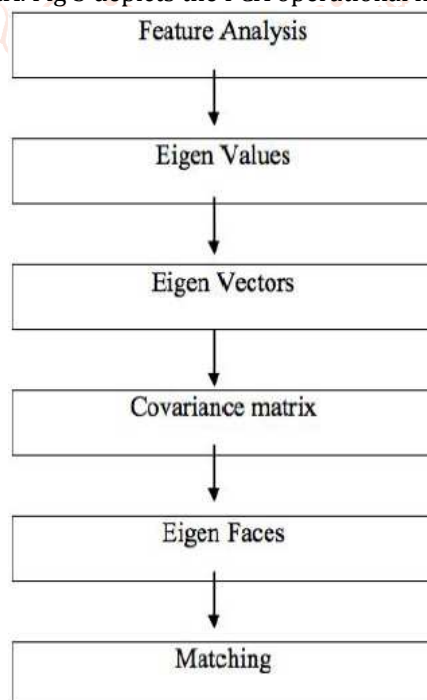
To obtain a new pixel value - pixels above and pixels to the left are added then all the values around the patch are added to obtain the sum of all pixel values. Ada boost will be determining relevant features and irrelevant features. Post identifying relevant features and irrelevant features, adaboost assigns a weight value to all of them, Which constructs a strong classifier as a linear combination of many Weak classifiers.



Nearly 2500 features are calculated, the number of computations can be further reduced by cascading. A set of features are here kept in another set of classifiers and so on in a cascading format. Using this method, one can detect if it is a face or not faster and can reject it if one classifier fails to provide a necessary output to the next stage. Then the detected face is cropped and resized to 100x100 that is the standard resolution. The step after that is to identify the detected image using Principal Component Analysis (PCA) and Artificial Neural Network Algorithm (ANN).

**FEATURE EXTRACTION:**

To extract human face features, we use PCA. Fig 5 depicts the PCA operational flow.



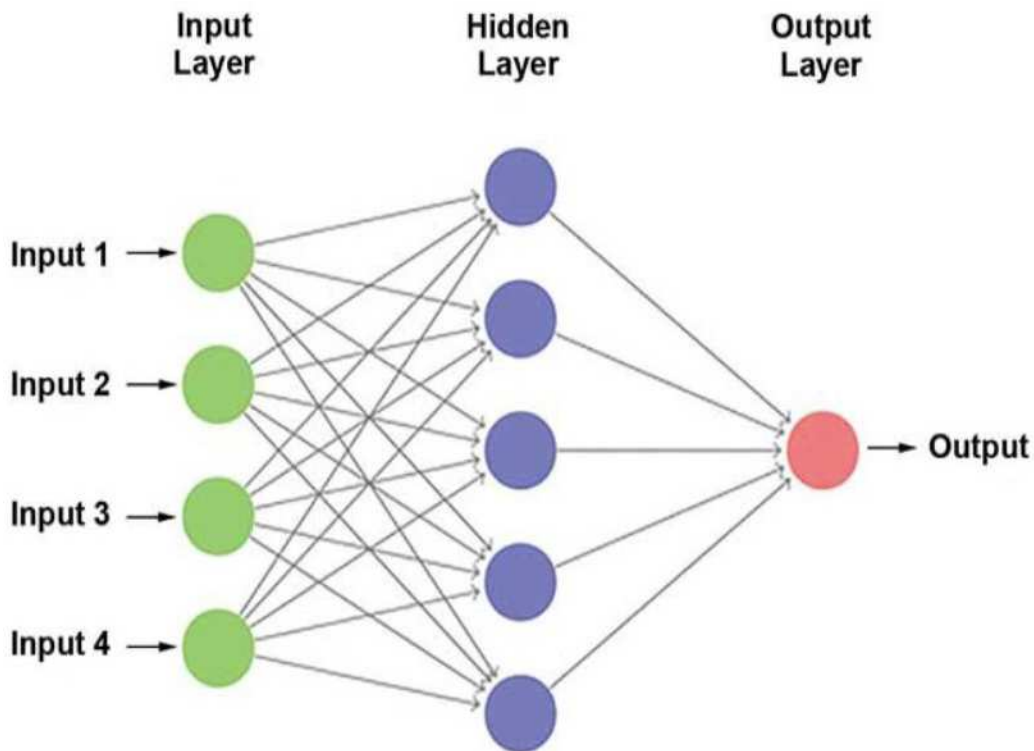
**Fig 5: PCA flow chart**

To extract features from a cropped and resized image Principal component analysis (PCA) is used. To transform higher dimensional data into lower dimensional data, It is used as a tool in exploratory data analysis and in predictive analysis. A bunch of M x M size facial images in a training are converted using principal component analysis technique into lower

dimensional face images. Principal Component Analysis (PCA) is used for the purpose of conversion of a set of correlated N variables into a set of uncorrelated k variables called principal components. The number of principal components (set of uncorrelated k variables) are less than or equal to the number of original values i.e.  $K < N$ . The above definition is modified as Principal component analysis for the application like face recognition application, it is one of the mathematical procedures used to convert a set of correlated N face images into a set of uncorrelated k face images called as Eigen faces.

Before calculating the principal components the dimension of the original images has to be Reduced, to reduce the number of calculations. Since principal components show more noise and less direction, only first few principal components (say N) are selected and the remaining components as they contain more noise can be neglected.

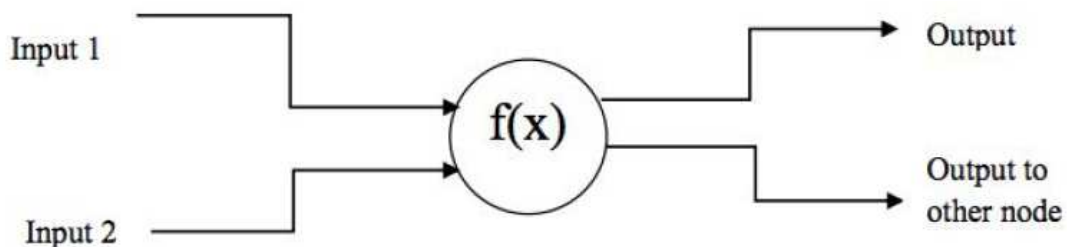
The M-image training set is represented by the best Eigen face with largest Eigen-values that accounts for the most variance with the set of best closely related feature-set and facial images. Each image in the training-set after finding Eigenfaces can be represented by a linear combination of Eigen Faces and will be represented as vectors. Standard database features are compared with input image features for Recognition.



**Figure 6: Example of an Artificial Neural Network**

Count of the neurons in the input layer is equal to the count Eigen faces, the type of the network is Feed forward back propagation network.

Refer Figure7, For a single cell represented as  $f(x)$ , its output can be calculated as  $\text{output} = \text{input1} + \text{input2}$  as shown in Figure 7. The function  $f(x)$  is a neutral function because it won't add or amplify any value to the incoming inputs but it just adds the value of incoming inputs. One can use a mathematical function such as tanh to represent the above function.



**Figure 7: Single neuron cell**

Layered feed forward Artificial Neural Networks make use of the back propagation algorithm, where the neurons send their signals in forward direction and the errors are propagated backwards. Until ANN learns the training data, the back propagation reduces this error. Through the back propagation technique the neural networks learn and determine the connection weights between the inputs, outputs and hidden cells. To make the error minimal random weights are initially assigned to these networks which are to be adjusted.

$$\text{Error in network} = \text{Desired output} - \text{Calculated output}$$

The back propagation technique is used to minimize the error, using a formula which consists of weights, inputs, outputs, error and learning rate ( $\alpha$ ).

$$*W_x = W_x - \alpha \left( \frac{\partial \text{Error}}{\partial W_x} \right)$$

Old weight
Derivative of Error with respect to weight  
New weight
Learning rate

**Figure 8: Updating New Weights in Back propagation.**

**Training of the Neural Networks:**

For each individual in the database considered one ANN is used, twenty three networks are created since there are twenty three persons in the database. Face descriptors are used as input for the purpose of training ANN. The face descriptors relating to the same individual are used as positive examples for that individual network's output will be 1 and as for negative examples like others so that output will be 0. Our trained network will be utilized for the purpose of recognition.

**Neural Networks Simulation:**

The facial descriptions of the test image calculated on Eigen-faces are used as input in all networks and are simulated. The results produced were comparable and the output being much higher than the previously described level ensures that the test image belongs to a well-recognized person with a maximum output.

**FACE TAGGING:**

In the Face Tagging stage the result from the simulation is used by the recognition system to tag an appropriate name to the image of the person. The data is in binary form and hence this block is also responsible for evaluating the expression into a certain value and matching it to a person's name in the name list. However, if the interpreted value is not one of the values listed in the roster, then the name returned will be automatically predefined as "Unknown".

In the Face Marking phase the results from the simulation are used in the awareness program to mark the correct word in the person's image. Data is in binary form so this block also has a responsibility to test the expression into a certain number and compare it with the person's name in the word list. However, if the translated value is not one of the listed values, then the name to be returned will be automatically defined as "Unknown".

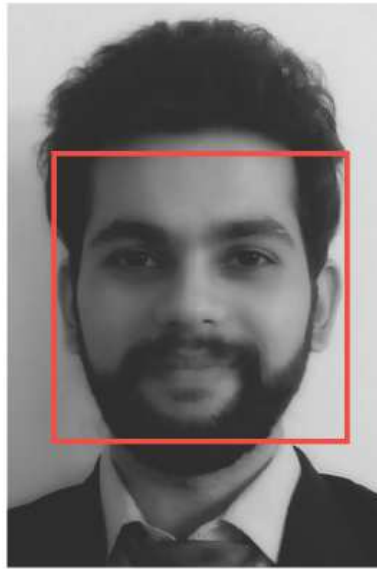
**RESULTS AND ANALYSIS**

Consider an image we have taken as test image as shown in Figure 9, it is preprocessed for identification



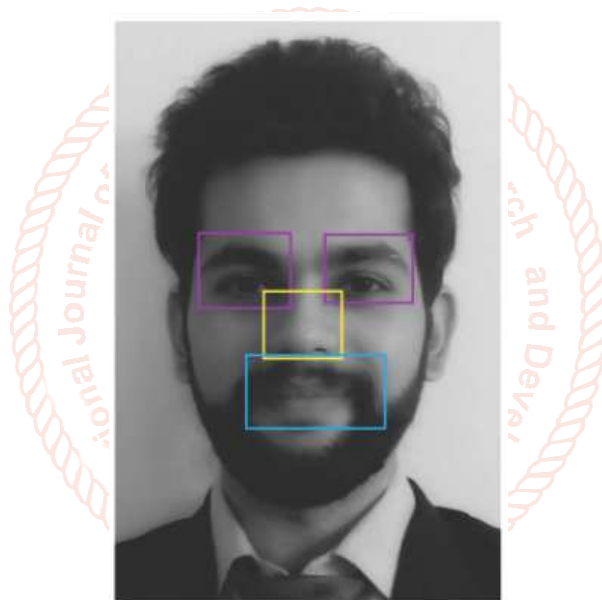
**Figure 9: Test Image**

Image 9 is the Test image taken and depicted for analysis for the paper, after applying the Viola-Jones algorithm to the image in Figure 9, Identified face image shown in Figure 10 is obtained (bounding box on identified face). Then it is resized to 100x100 pixels, that is the Haar features are calculated and all the related features are extracted.



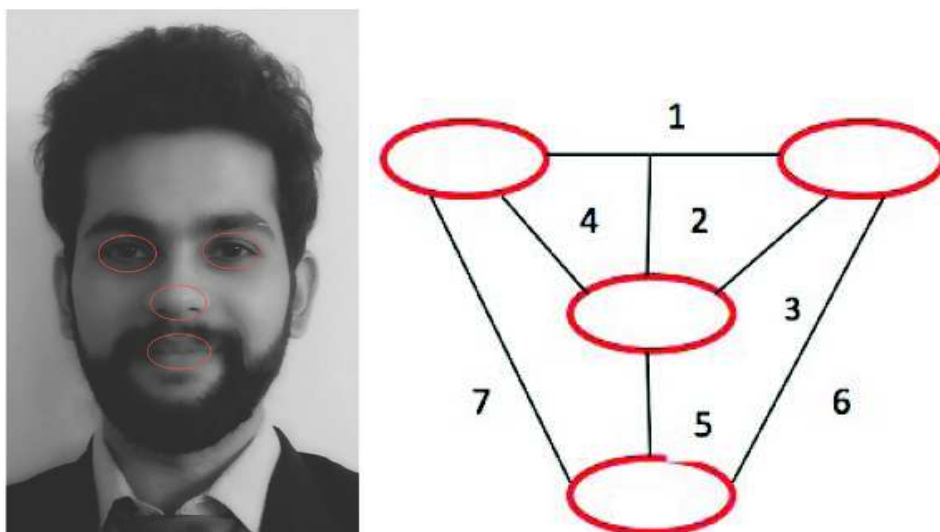
**Figure 10: Face recognised by Viola-Jones algorithm in Red boundary**

As shown by the bounding box in figure 11 main features of the face are identified by Viola-Jones algorithm and is used for deciding the nodes corresponding to the identified part of the face.



**Figure 11: Facial features identified by Viola-Jones algorithm (Boundary box)**

The features extracted by Viola-Jones algorithm are represented as nodes, these nodes are joined to form a shape making it sure that all nodes are well connected and the connected lines are named with reference numbers as shown in Figure 12.



**Figure 12: Facial Feature Calculation**

Figure 12 shows that in order to identify the person how the specifications of features are calculated. Each detail is tabulated after the features are calculated from various angles. The person in the image is identified correctly based on this calculation. The tabulated results of the various features taken-into-account are shown in Table 1.

**Table 1: Calculation of various features Different angles VS Face features of Images**

INDEX	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7
1	884	356	523	522	655	568	580
2	862	383	523	524	654	572	590
3	1923	370	530	512	650	638	646
4	1926	354	512	599	633	618	613
5	1120	419	696	612	720	639	654
6	1943	385	555	543	712	585	588
7	1912	362	560	546	666	700	612
8	1958	342	514	517	656	619	617
9	1192	326	449	466	632	630	632
10	1982	333	518	527	658	612	626
11	1943	320	472	489	626	579	692
12	1997	364	517	512	660	576	584
13	1934	245	492	420	542	613	619
14	1932	392	439	443	613	622	638
15	1868	360	512	599	632	552	548

**Table 2: Results**

Techniques / Authors	Accuracy
Neural Networks	92%
Principal Component Analysis	72%
Kamencay [2] a) Segmented	90%
b) Non Segmented	84%
Fernandez [3] (Artificial Neural Nets and Viola-Jones)	88.64%
Mohammad Da'san[4] (Viola-Jones , Neural-Networks)	90.31%
Proposed Method	94%

We compared the accuracy of our proposal with existing models as shown in Table 2. The accuracy of the proposed method turned out to be 94%, thus the proposed method is more accurate in recognising a person in an image when compared to other methods.

## CONCLUSION

The paper presents an efficient approach i.e. fusion of preprocessing PCA then Viola-Jones and utilizing neural networks for face detection and recognition. The accuracy is compared with other existing models that perform the same operations, where It is observed that the performance of the model is superior. Facial Detection plays an important role in a plethora of applications, where-in most cases there is a desire to utilize the high rate of accuracy in recognition of people hence the proposed method can be considered after taking account of the results with other existing methods

## REFERENCES

- [1] Maria De Marsico, Michele Nappi, Daniel Riccio and Harry Wechsler, " Robust Face Recognition for Uncontrolled Pose and Illumination Changes" IEEE Transaction on Systems, Man and Cybernetics, vol.43,No.1,Jan 2013.
- [2] Kamencay, P, Jelsovka, D.; Zachariasova, M., "The impact of segmentation on face recognition using the principal component analysis (PCA)," IEEE International Conference in Signal Processing Algorithms, Architectures, Arrangements, and Applications, pp.1-4, Sept. 2011.
- [3] Ma. Christina D. Fernandez, Kristina Joyce E. Gob, Aubrey Rose M. Leonidas, Ron Jason J. Ravara, Argel A. Bandala and Elmer P. Dadios "Simultaneous Face Detection and Recognition using Viola-Jones Algorithm and Artificial Neural Networks for Identity Verification", pp 672-676, 2014 IEEE Region 10 Symposium, 2014.
- [4] Mohammad Da'san, Amin Alqudah and Olivier Debeir, " Face Detection using Viola and Jones Method and Neural Networks" IEEE International Conference on Information and Communication Technology Research, pp 40-43,2015.
- [5] Anil K. Jain, "Face Recognition: Some Challenges in Forensics", IEEE International Conference On Automatic Face and Gesture Recognition, pp 726-733,2011.
- [6] Ming Zhang and John Fulcher "Face Perspective Understanding Using Artificial Neural Network Group Based Tree", IEEE International Conference On Image Processing, Vol.3, pp 475-478, 1996.
- [7] Hazem M. El-Bakry and Mohy A. Abo Elsoud "Human Face Recognition Using Neural Networks" 16th national radio science conference, Ain Shams University, Feb. 23-25, 1999.
- [8] Muhammad Khan, Jocelyn Chanussot, Laurent Condat, Annick Montanvert. Indusion : Fusion of Multispectral and Panchromatic Images Using Induction Scaling Technique. IEEE Geoscience and Remote Sensing Letters, IEEE - Institute of Electrical and Electronics Engineers, 2008, 5 (1), pp.98- 102. ff10.1109/LGRS.2007.909934ff. fffal-00348845f