Face Recognition in video images using Neural Network

Gajanan Joshi[1], Prakash Khanale[2]
[1]Vidya Pratishthans Arts, science & commerce college, Baramati,
[2]DSM College, Parbhani.

Abstract – Face recognition being the foremost important biometric attribute it still faces many challenges, like variation in images, illumination variation etc. Once such variations are present in each pose and illumination, all the algorithms are greatly stricken by these variations and their performance gets degraded. The method to identify faces in given video sequence is presented in this paper. Identification of faces in the given frames can be done using viola Jones classification model. Frames are separated and categorized in two sections Good frames and Bad frames. Good frames mean the frames which containing face or faces, and Bad frames means frames, which does not contain face. Face detection is the first stage towards face recognition. This is called preprocessing for face recognition method. The frames obtained by this method can be put for further processing as far as face recognition is concerned.

Key Words: Illumination, Pose, Wavelet PCA, Frames Classification, Good frames, Bad frames, video sequence, skin-color, skin modelling, face detection.

INTRODUCTION

Face recognition from video has received full attention in recent computer vision analysis. In general, video provides further knowledge for recognition as compared to a still image. However, several troublesome problems still keep unsolved, like changes in illumination, pose, and occlusion. One essential drawback is matching corresponding pixels from overlapping face regions from ordered footage in an exceedingly video sequence beneath changes in illumination, pose, and occlusion. May this be a major drawback once alone a different of the face region is shown and so a similar region could appear in many poses and scales? One desires a method to correspond the elements of the overall faces or face patches, collect the face patches from video, and construct a full face or the utmost quantity of a face region as adaptable [1].

Variety of ways are planned to the standard face recognition, i.e. characteristic of face detected from one image. However, such ways could fail in video-based face recognition tasks. Many previous researches show that a naive scheme transferring from images to videos usually encounters a significant decrease in recognition accuracy.

Although videos give far more info than pictures, it is hard to exploit and utilize it effectively. Different video sequences of the same person may contain several kinds of variations, together with cause, face expression, illumination, image resolution and occlusion. Unlabeled and unpredictable variations result in mistakes in recognition. Motion blur and compression artifacts additionally deteriorate recognition performance. A way to acknowledge faces from videos captured in at environments has become exchange challenge for researchers [3].

Moreover, in video surveillance and e-passport identification, solely one still image of every person is out there. It is unfeasible to use intra-class information to explain variations in every class, which is important for classification. Meanwhile, differing from the gallery set consisting of a single still image per person, probe sets are typically collected on the spot within the kind of video sequences. It is crucial to match multiple probe images with the single gallery image correctly without being affected by possible outliers.

As a result, two major issues would like considering once implementing such a still-to-video face recognition system [3]. One is multi-modality or heterogeneousness between still pictures and video frames. Still pictures collected by digital cameras in strained environments area unit high-quality face samples with frontal face expression, neutral expression and traditional lighting. Still images and video frames have such different appearances that direct matching would obviously result in many mistakes.

Face recognition system is also sophisticated image-processing problem in universe applications with complicated effects of illumination, occlusion, and imaging condition on the live pictures. It is a combination of face detection and recognition techniques in image analyzes. Detection application is utilized to look out position of the faces throughout a given image. Recognition formula is employed to classify given pictures with known structured properties, which are used typically in most of the computer based vision applications. These pictures have some glorious properties like; same resolution, along with same facial feature elements, and similar eye alignment. These pictures are aiming to be referred as "standard image". Recognition applications uses customary pictures and detection algorithms detect the faces and extract face footage that embody eyes, eyebrows, nose, and mouth. That produces the formula harder than single detection or recognition formula. The first step for face recognition system is to accumulate an image from a camera or gallery. Second step is face extraction from the still image. Third step is, face detection. Forth step, face recognition that takes the face footage from output of detection half. Final step is person identity as results of recognition half.

STEPS INVOLED IN FACE RECOGNITION

1. Face Detection:

Detection can be location of face region as a whole, and location of face region with facial features (i.e. eyes, mouth, eyebrow, nose etc.). Detection methods in the literature are difficult to classify strictly, because most of the algorithms are combination of methods for detecting faces to increase the accuracy.

The basic principle of Viola-Jones algorithm is to scan a sub-window capable of detecting faces across a given input image. The standard image processing approach would be to rescale the input image to different sizes and then run the fixed size detector through these images. This approach turns out to be time consuming due to the calculation of the different size images. On the other side Viola-Jones rescale the detector instead of the input image and run the detector many times through the image each time with a different size. Viola-Jones has devised a scale invariant detector that requires the same number of calculations whatever the size. This detector is constructed using integral image and some rectangular features reminiscent of Haar wavelets.

Viola-Jones algorithm consist of four stages – Haar features selection, Creating an integral image, Adaboost Training, Cascading Classifiers.

All human faces share some similar properties. These regularities may be matched using Haar features. Few properties common to human faces - the eye region is darker than the upper-cheeks; the nose bridge region is brighter than the eyes. Instead of finding the faces, the algorithm should discard non-faces. The idea

behind this statement is that it is faster to discard non-faces than to find a face. With this idea, a detector consisting of only one (strong) classifier suddenly seems inefficient since evaluation time is constant. Hence, there is need of cascaded classifier. The cascaded classifier consists of stages each containing strong classifier. The job of each stage is to determine whether a given sub-window is definitely not a face or maybe a face. When a sub-window is classified to be a non-face by a given stage it is discarded. If sub-window classified as a maybe-face it is passed on to the next stage in the cascade. It means that the more stages a given sub-window passes, the higher the chance the sub-window actually contains face.

2. Face Pre Processing of Face Image

In this module, by means that of early vision techniques, face pictures are pre-processed and increased to boost the popularity performance of the system.

Based on demand a number of the following pre-processing techniques are used in the projected face recognition system. totally different varieties of pre-processing/enhancement techniques connected to the face recognition method are explained as fallows with the assistance of flow chart and corresponding face pictures.

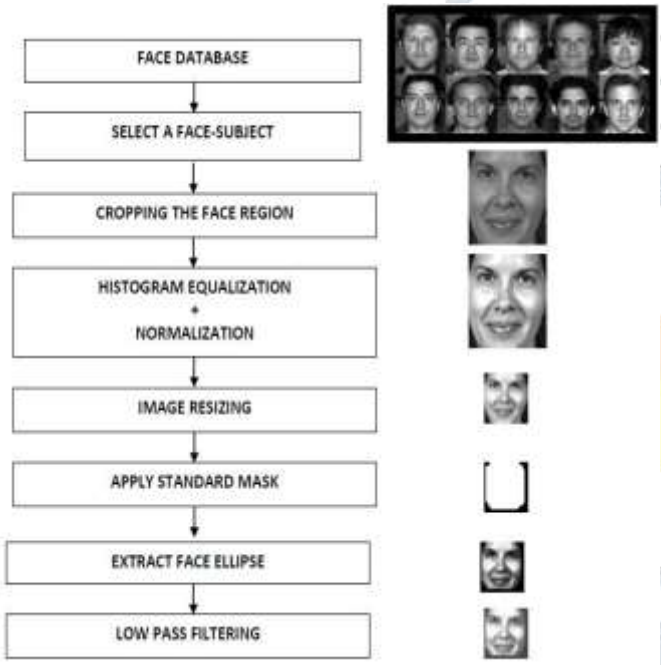


Figure 1 Flow chart representing the Pre processing

2.1 Cropping of images:

When the space of an image is abundant larger compared to that of a face, the region of the image wherever the face is found is cut out from the
image and solely this space is used in the method of face recognition. By using cropping technique solely main face will be extracted and the redundant information around the face, that detoriate the performance of recognition is removed [11].

In this study, the face space is determined manually primarily based on observation of completely different face pictures in the given database and rectangular or square window corner points are selected in such an approach that it includes main options of face such as eyes, nose, and mouth. All face images to be trained or tested in recognition process are cropped by using same corner points („imcrop" function in MATLAB is being used for cropping required part from each face image).

2.2 Histogram Equalization

It is typically done on low distinction pictures in order to enhance image quality and to improve face recognition performance. It changes the dynamic (contrast range) of the image and as a result, some necessary face expression become a lot of visible. Mathematically bar chart exploit are often expressed as:

$$S_k = T(r_k) = \sum_{j=0}^{k} \frac{n_j}{n} \quad \ldots\ldots\ldots \quad 2.1$$

Here in equation (2.1) 'n' is that the total range of pixels in a picture, 'nj' is that the range of pixels with grey level 'r k' ,and 'L' is that the total range of grey levels exist within the face image. The result when applying bar chart exploit to a sample face image is shown in Figure. 2.

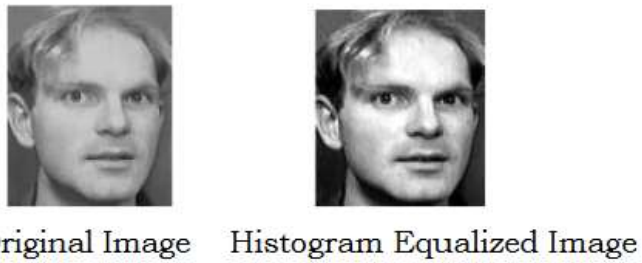


Figure 2 Before and after histogram

Two histogram plots are given in Figure 3. The histogram on the left is of the original face image
(between 6-250) and the one on the right is after histogram equalization is applied [11].

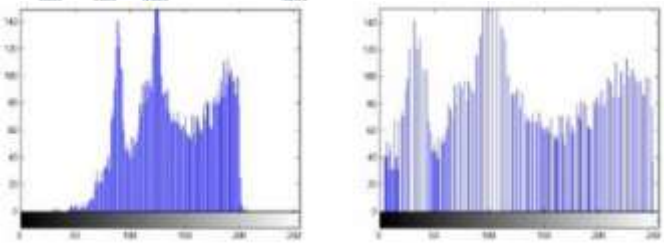


Figure 3 The histogram of an image before and after the histogram equalization

2.3 Illumination Normalization

Face pictures taken beneath completely different illuminations will degrade recognition performance, particularly for face recognition, systems supported the mathematical space analysis, during which entire face data is employed for recognition. The complete high and low illumination levels are adjusted so the image becomes abundant clearer and quiet. Normalization of face image is performed to urge zero mean and unity variance value [11].

2.4 Image re-sizing exploitation Bi-Cubic Interpolation methodology

The process of image resizing changes the dimensions of the image, during this work, the dimensions of the image is scaled all the way down to scale back the resolution of face image. This reduction in size is finished to scale back mathematical quality in PCA/LDA method and neural network method .For resizing face image, completely different interpolation techniques exist in literature, among them Bi-cubic interpolation methodology is being employed within the planned work, the advantage of resizing through Bi-cubic interpolation is that, it produces more smoother surfaces than any other interpolation technique[11] (image is being resized to desired dimension by using „imresize" function in MATLAB) . In Bi-cubic Interpolation technique, it considers sixteen neighboring components within the rectangular grid and calculates weighted average of those pixels to switch them with one pixel, it's that component that possesses the flavor of all the sixteen replaced pixels [11].

2.5 Masking

By employing a mask, that merely contains a face formed region; the result of background amendment is decreased. The form of the mask employed in this study is shown in Figure 4 [11].

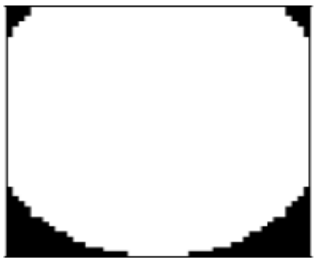


Figure 4 The shape of the face mask used in the pre-processing

2.6 Low Pass Filtering

The mean/averaging filter is applied so as to provide the result, as a result of within the later stages face recognition algorithmic rule embrace step of face image re-sizing (by exploitation down sampling method) whereas maintaining the standard of face image. The 5*5 filter is employed for this method [11].

$$R=\frac{1}{25}\sum_{i=1}^{25} Z_i$$

Equation above calculates the average value of the pixels, whereas 'z' is the mask, 'i' are mask elements. The mask is then convolved with image to produce filtering effect, for a 5*5 mask used in the implementation, it calculate the average of 25 pixels in that filter mask.

3. Wavelet transformation

One major advantage afforded by wavelets is that the ability to perform local analysis that is, to analyses a localized space of a large signal. In wavelet analysis, we often speak regarding approximations and details. The approximations are the high-scale, low-frequency parts of the signal. The details are the low-scale, high frequency components [11].

Using 2nd WT (Fig.5), the face image is decomposed into four sub images via the high-pass and low-pass filtering. The image is decomposed on column

direction into sub images to high-pass frequency band H and lowpass frequency band L. assumptive that the input image is a matrix of m x n pixels, the resulting sub images become m/2 x n matrices. At second step the pictures H and L are rotten on row vector direction and severally manufacture the high and low Frequency band HH and hl for H, and LH and LL for L. The four output pictures become the matrices of m/2 x n/2 pixels. Low frequency subimage LL(A1) possesses high energy, and is a smallest copy of original images (A0.). The remaining subimages LH, HL, and HH severally extract the dynamical components in horizontal (D11), vertical (D12), and diagonal (D13) direction.

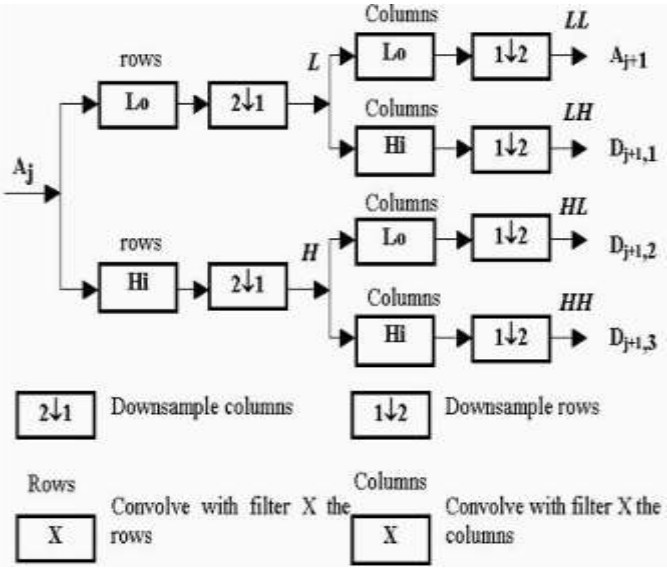


Figure 5 Scheme of one level two-dimensional Wavelet Transform

The effect of use 2D WT on real image of face shown on Fig.5. The size of output images is the same as the input image. It is result of down sampling that is rejection of every second row and column. This operation don't increase amount of data and simultaneously don't cause loss of information.

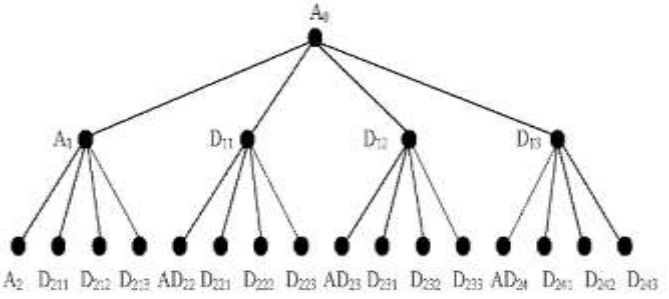


Figure 6 the wavelet decomposition tree

4. PCA

The Principal component Analysis (PCA) is one among the foremost successful techniques that have been employed in image recognition and compression. PCA may be a statistical method. The purpose of PCA is to scale back the big dimensionality of the data space (observed variables) to the smaller intrinsic spatial property of feature space (independent variables), that square measure required to explain the data economically. This can be the case once there is a strong correlation between observed variables [11].

The jobs that PCA will do square measure prediction, redundancy removal, feature extraction, data compression, etc. as a result of PCA may be a classical technique which might do one thing within the linear domain, applications having linear models square measure appropriate, such as signal processing, image processing, system and control theory, communications, etc.

PCA computes area that is represented by its training vectors. These basis vectors, truly eigenvectors, computed by PCA square measure within the direction of the most important variance of the testing vectors. Because it has been said earlier, we call them Eigen faces. Every Eigen face will be viewed a feature. Once a particular face is projected onto the face house, its vector into the face house describe the importance of every of these options within the face. The face is expressed in the face house by its Eigen face coefficients (or weights). We are able to handle an outsized input vector, facial image, solely by taking its little weight vector within the face house. This suggests that we are able to reconstruct the first face

with some error, since the dimensionality of the image house is far larger than that of face space.

5. Face recognition

Face recognition may be a terribly difficult task for the researchers. Its application is extremely helpful for private identification and verification. Face recognition concept includes biometric approaches appreciate fingerprint, iris/retina and voice recognition. Throughout recognition method we had taken an unknown image and at that time extract the options of that unknown image we had taken and next compared that image with the hold on example throughout training, the face recognition system can come the soul match list of possible matches. It will be divided into two approaches: Geometric feature based mostly approach and holistic approach [3].

Face recognition is that the most vital stage within the entire system. Videos are capable of providing a lot of unusable information than still image. The main blessings for using videos are first off the likelihood of using redundancy contained within the video sequence to boost still pictures recognition performance, second dynamic info is available and third to boost recognition effects from the video sequence using simpler representations appreciate a 3D face model or super-resolution pictures. Finally, video-based recognition permits learning or change the topic model over time. Though the benefits are obvious, there additionally exits some disadvantages. For example, poor video quality, low image resolution, and different influence factors (such as illumination, pose change, motion, occlusion, decoration, expression, large distance from camera, etc).

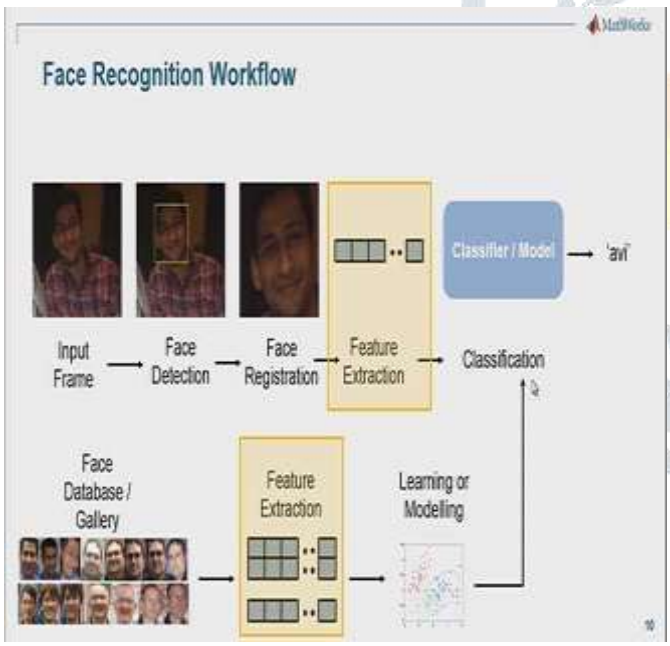


Figure 7 Face Recognition Workflow

The method for exploit face pictures depends upon the underlying application. Police work applications may best be served by capturing face pictures by means that of a video camera whereas image information investigations could require static intensity pictures taken by a customary camera.

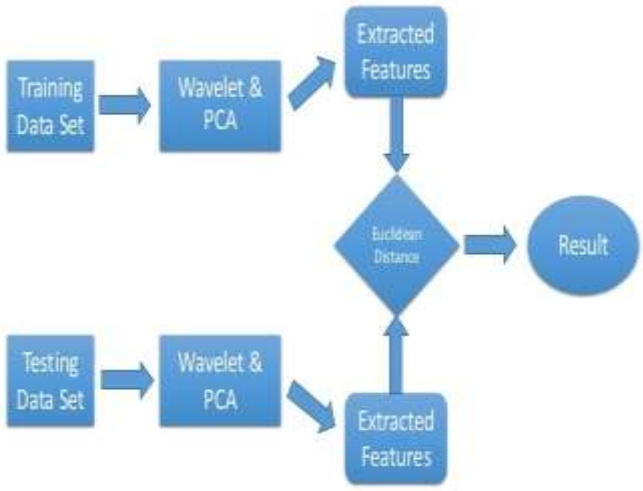


Figure 8 Block Diagram

Therefore, looking on the face knowledge acquisition methodology, face recognition techniques is broadly speaking divided into three categories: ways that operate intensity images, people who modify video sequences, and those that require different sensory knowledge appreciate 3D data or infrared imagination [4].

An overview of some of the well-known methods are given below.

5.1 Featured-based Approach:

Feature-based approaches or method the input image to identify and extract (and measure) distinctive facial features such as the eyes, mouth, nose, etc., yet as different fiducial marks, so compute the geometric relationships among those facial points, therefore reducing the input facial image to a vector of geometric options. Customary applied math pattern recognition techniques and then utilized to match faces using these measurements. Early work distributed on machine-controlled face recognition was largely supported these techniques. One in every of the earliest such tries was by Kanade, who employed easy image process ways to extract a vector of 16 facial parameters - that were ratios of distances, areas and angles (to make amends for the varied size of the pictures) -and used an easy euclidian distance live for matching to realize a peak performance of 75th on a database of 20 completely different individuals using a pair of images per person (one for reference and one for testing)[4].

5.2 Holistic Approach:

Holistic approaches attempt to determine faces using global representations, i.e., descriptions supported the complete image instead of on native features of the face. These schemes may be divided into two groups: statistical and AI approaches.

5.2.1 Statistical Approach

In the simplest version of the holistic approaches, the image is represented as a 2d array of intensity values and recognition is performed by correlational statistics comparisons between the input face and every one the opposite faces within the database. Although this approach has been shown to figure below restricted circumstances (i.e., equal illumination, scale, pose, etc.), reminiscent of sensitivity to face orientation, size, variable lighting conditions, background clutter, and noise.

5.2.2 AI Approach:

AI approaches utilize tools such as neural networks and machine learning techniques to recognize faces [4].

6. Soft Computing:

Soft computing is an rising approach to computing which parallel the outstanding ability of the human mind to reason and learn in a setting of uncertainty and imprecision.
Some of it's principle components includes:

- Neural Network(NN)
- Fuzzy Logic(FL)
- Genetic Algorithm(GA)

These methodologies form the core of soft computing.

6.1. Neural Network:

An NN, in general, is a highly interconnected network of a large number of processing elements called neurons in an architecture inspired by the brain [5].
NN Characteristics are:-

- Mapping Capabilities / Pattern Association
- Generalization
- Robustness
- Fault Tolerance
- Parallel and High speed information processing

Neuron

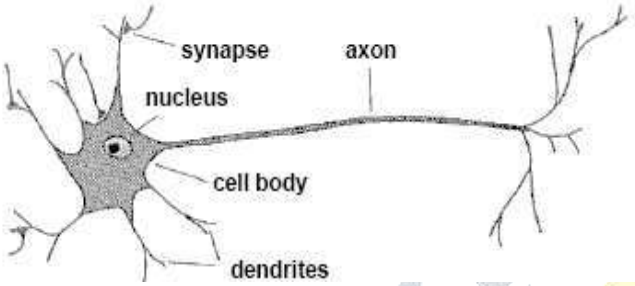


Figure 9 Biological neuron

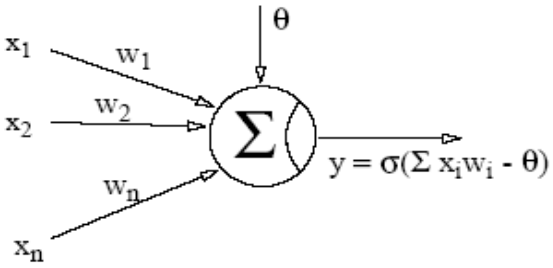


Figure 10 Model of a neuron

Architecture of ANN:

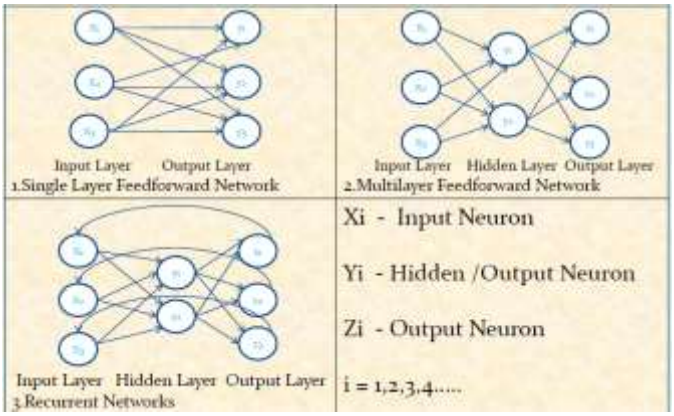
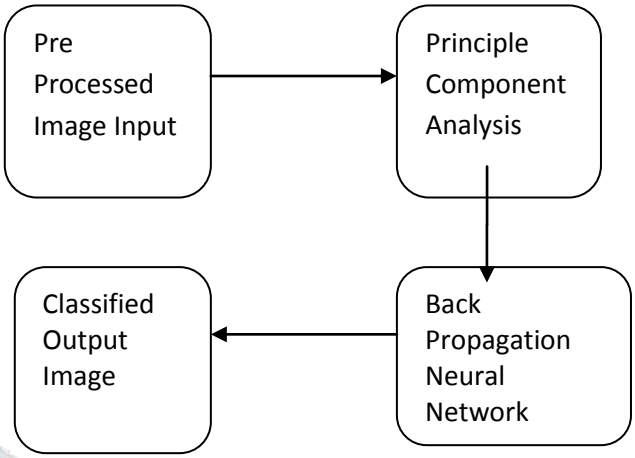


Figure 11 ANN Architecture

6.2 Steps involved in recognition Feed Forward NN:

- Pre-Processing stage
- Principle Component Analysis
- Back Propagation Neural Network



PRE-PROCESSING:

- To reduce or eliminate some of the variations in face due to illumination.
- It normalize and enhance the face image to improve the recognition performance.
- By using the normalization process system robustness against scaling, posture, facial expression and illumination is increased.

PRINCIPLE COMPONENT ANALYSIS(PCA)

- It involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components.

BACK-PROPAGATION NEURAL NETWORK(BPNN)

- It trains the network to achieve a balance between the ability to respond correctly to the input patterns that are used for training & the ability to provide good response to the input that are similar.
- It requires a dataset of the desired output for many input, making up the training set.
- These are necessarily Multilayer Perceptron's (MLPs) [5].

MLPs:
1. Set of input layers
2. One or more hidden layers
3. Set of output layers

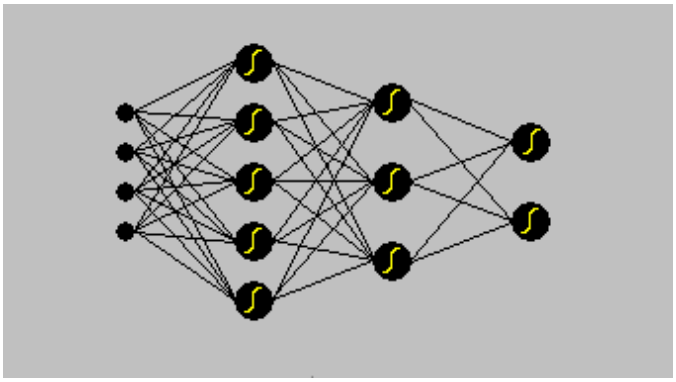


Figure 12 MLP's

FLOW CHART
Below flow chart describes the system flow.

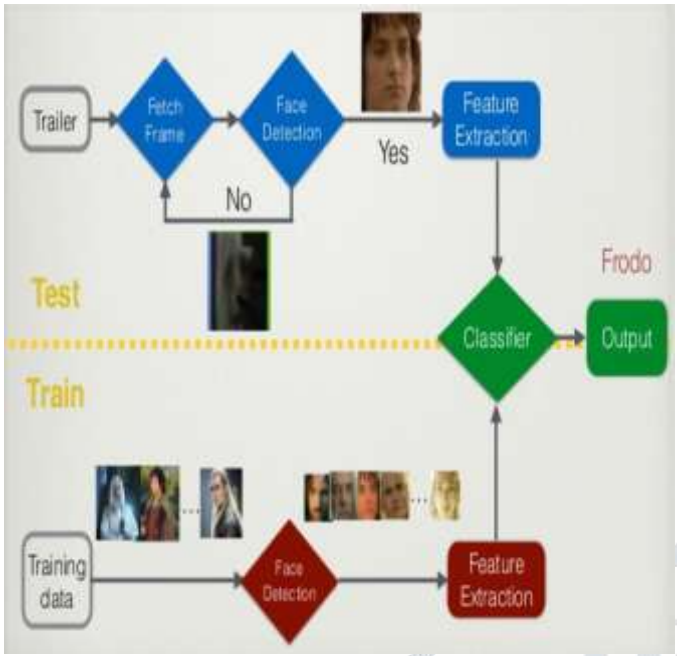


Figure 13 Flow Chart

EXPERIMENTAL WORK/ PROCESS
In our work we had used database, created using standard NRC-IIT data set, video sequence from computer science department. Total Images in dataset are100 images of 10 different persons.

a) First, we are going to take video as input for our system after that we are able to take separate frames which containing faces from that video sequence.
b) Now we are able to separate good frames and bad frames (good frame means frame with face bad frame means frame without face).
c) After catching good frames, we are able to apply cropping process; because of this, we have main face without redundant information around the face.
d) After cropping process, we are going to apply resizing process on that cropped image or frame, for getting proper coordinates of that image.
e) Next, we have to convert that resized image or frame to gray scale image.
f) Now we are applying Histogram equalization process on that image or pictures in order to enhance image quality and to improve face recognition performance.
g) After that Wavelet transformation carried out on that image for perform native analysis--that is, to analyses a localized space of a bigger signal. In wavelet analysis, we often speak regarding approximations and details.
h) Based on wavelet transformation result we obtained approximations and details of that image or picture then we apply PCA to get Eigenvectors.
i) From that Eigenvectors we had created feature vectors.
j) Then we train the network and simulate the network, training data set and test data set (video for face recognition).

Below shows the result screen of all steps,

1. Input Data from NRC-IIT

Figure 14 Result of different stages as skin detection- Feature Extraction- Face Detection-Face segmentation

Figure 15 Result of Face Recognition Process

2. Input Data from NRC-IIT

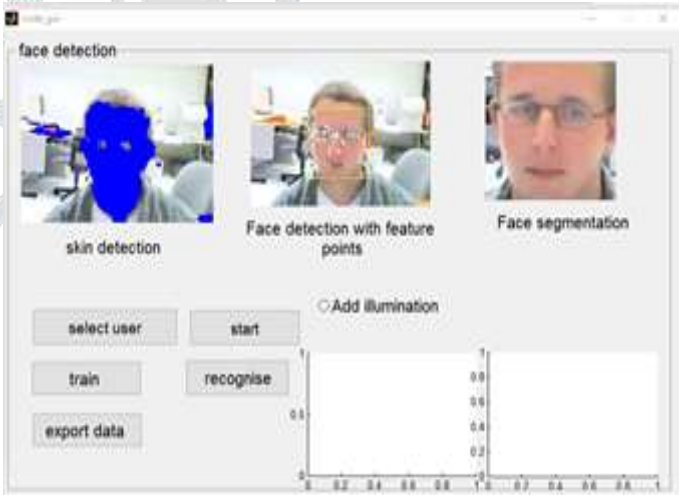


Figure 16 Result of different stages as skin detection- Feature Extraction- Face Detection-Face segmentation

Figure 17 Result of Face Recognition Process

3. Input Data from Local Dataset

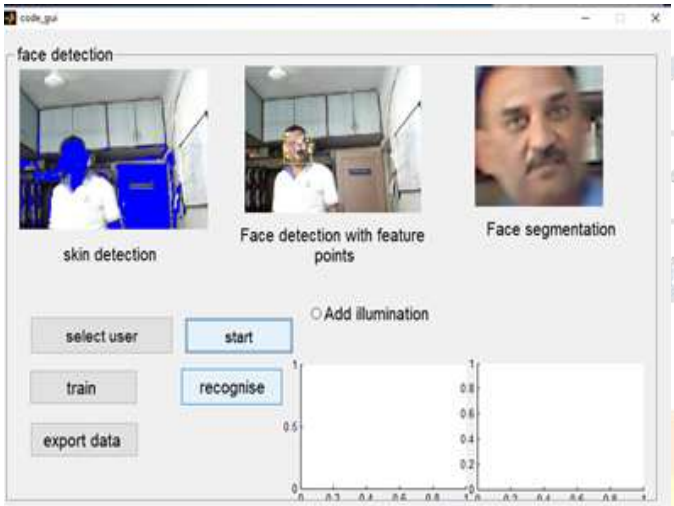


Figure 18 Result of different stages as skin detection- Feature Extraction- Face Detection-Face segmentation

Figure 19 Result of Face Recognition Process

4. Input Data from Local Dataset

Figure 20 Result of different stages as skin detection- Feature Extraction- Face Detection-Face segmentation

Figure 21 Result of Face Recognition Process

CONCLUSIONS

Facial features are extracted from video sequence. These feature or feature vector are used for further face recognition process. Neural network is trained using this feature vector (Eigen vectors). For testing, video sequence is taken as input. This input is nothing but the real time video data of surveillance camera. The system will identify the person from input video, display his name (or class name to which he /she belongs) and if person is not in the dataset then it is identified as unknown object or unknown person. With the help of this system, one can check the number of times the person appears or present before the camera. The record has been maintained that name of person,(class of person), date and time at which he/she appear in front of the camera, total number of time that particular object (face/ image/person) appear during the given time is calculated.

REFERENCES

[1] Changbo Hu, Josh Harguess and J. K. Aggarwal , "PATCH-BASED FACE RECOGNITION FROM VIDEO," Computer & Vision Research Center / Department of ECE The University of Texas at Austin.

[2] Praseeda Lekshmi V, Dr. SasiKumar, Divya S. Vidyadharan, "Face Detection and Localization of Facial Features in Still and Video Images", 978-0-7695-3267-7/08, 2008 IEEE DOI 10.1109 / ICETET. 2008. 138.

[3] Rabia Jafri* and Hamid R. Arabnia, "A Survey of Face Recognition Techniques", DOI : 10.3745/JIPS.2009.5.2.041,

Journal of Information Processing Systems, Vol.5, No.2, June 2009.

[4] J.SUNEETHA, "A survey on Video-based Face Recognition Approaches ", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 2, February 2014.

[5] Manisha M. Kasar, Debnath Bhattacharyya and Tai-hoon Kim, "Face Recognition Using Neural Network: A Review", International Journal of Security and Its Applications, Vol. 10, No. 3 (2016), pp.81-100 http://dx.doi.org/10.14257/ijsia.2016.10.3.08

[6] Sonam Kesharwani and Sanjivani Shantaiya, "A Survey on Face Recognition Techniques in Video", European Journal of Advances in Engineering and Technology, 2015, 2(5): 112-116.

[7] Haoyu Wang, Changsong Liu and Xiaoqing Ding, "Still-to-video face recognition in unconstrained environments," State Key Laboratory of Intelligent Technology and Systems Tsinghua National Laboratory for Information Science and Technology Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.

[8] Shaokang Chen, Erik Berglund, Abbas Bigdeli, Conrad Sanderson, Brian C. Lovell "Experimental Analysis of Face Recognitionon Still and CCTV images", NICTA, PO Box 10161, Brisbane, QLD 4000, Australia ITEE, University of Queensland, Brisbane, QLD 4072, Australia.

[9] António Mota Ferreira and Miguel Velhote Correia , "Face Detection Based on Skin Color in Video Images with Dynamic Background", Instituto de Engenharia Biomédica, Laboratório de Sinal e Imagem Universidade do Porto, Faculdade de Engenharia, Departamento de Engenharia Electrotécnica e de Computadores Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal.

[10] Dmitry O. Gorodnichy "Videobased framework for face recognition in video". Second Workshop on Face Processing in Video (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), pp. 330338, Victoria, BC, Canada, 911 May, 2005. ISBN 0769523196. NRC 48216.

[11] www.wikipedia.com

[12] Lixin Fan and Kah Kay Sung School of Computing National University of Singapore, Singapore, 117543, "Model-Based Varying Pose Face Detection and Facial Feature Registration in Video Images", ACM Multimedia 2000 1-58113-198-4/00/10, Los Angeles CA USA.

[13] Wikipedia: "Viola-Jones object detection framework".

[14] Xinzhu Wang, Yantao Tian, Shuaishi Liu, Cheng Peng, "Face detection and Tracking Algorithm in Video Images with Complex Background", Proceeding of the IEEE International conference on Robotics and Bioinformatics, December 14-18, 2010, Tianjin, China.

[15] Lijing Zhang, Yingli Liang, " A Fast Method of Face Detection in Video Images", Vol. 978-1-4244-5848-6/10, 2010 IEEE.

[16] Dmitry O. Gorodnichy Videobased framework for face recognition in video. Second Workshop on Face Processing in Video (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), pp. 330338, Victoria, BC, Canada, 911 May, 2005. ISBN 0769523196.NRC48216. [Abstract & Pdf]

[17] Dmitry O. Gorodnichy, Associative neural networks as means for lowresolution videobased recognition. International Joint Conference on Neural Networks (IJCNN'05), Montreal, Quebec, Canada, July 31August 4, 2005. NRC 48217. [Abstract & Pdf]