Facial Expression based Person Authentication

S. Saravanan
Research Scholar
Dept. of Computer Science & Engg.
Annamalai University, India

S. Palanivel
Professor
Dept. of Computer Science & Engg.
Annamalai University, India

M. Balasubramanian
Assistant Professor
Dept. of Computer Science & Engg
Annamalai University, India

ABSTRACT

High quality automatic facial expression based person authentication system is practically difficult mainly due to poses in face. This work paves way to develop a more perfect automatic person authentication system using facial expressions. In this work, ways to extract automatically pose free face images from video taken in normal room condition, determining mouth region, extracting features along with performance comparison in person authentication during normal and smile facial expressions is explained. The system contains two stages. In first stage, automatic pose free image selector is used to collect pose free face images from videos of ten persons taken in two sessions each with normal and smile facial expressions with poses. Testing on images taken from forty videos of resolution 640 x 480 the system identified and extracted pose free face images automatically which are 100% perfect pose free face images. The rejected images may have pose free images, but it will not affect the working accuracy of the system even though may reduce its speed, but not significantly. In stage second, automatically selected pose free images of mouth during normal and smile facial expression from the twenty videos of first session is used for training an auto associative neural network. Images from the second session of twenty videos are used to test for person authentication. The results clearly show that normal face gives more performance than smile facial expression for person authentication by accepting authentic persons and rejecting impostors. Equal error rate is used to calculate the performance of the person authentication system. Equal error rate for person authentication using normal face is 0.32% whereas with smile facial expression is 0.4%. The person authentication system is considered more efficient if the equal error rate value is lower.

General Terms

Image processing, Neural networks, Authentication

Keywords

Automatic pose free image selector, Auto associative neural network, Smile facial expression, Person authentication.

1. INTRODUCTION

A lot of researches have been nowadays done in person authentication, mainly using face images. Person authentication applications can be used in many places practically like surveillance, criminal identification, polling stations and automated teller machine centers. The main hindrance to achieve accuracy in person authentication using face image is pose, when images are taken in a normal room condition like an automated teller machine center. Preexisting image databases are not used as nearly all have face area manually cropped. There are also scarcity of videos of same persons taken in two different sessions with normal and smile facial expressions with poses. Automatic pose free image selector is not used in most of the person authentication system, without which creating a perfect automatic person authentication system is practically difficult. Poses are of three types namely tilt, roll and yaw. Pitch or tilt is the up and down movement of head. Roll is tilting the head sideways facing the camera while the nose is in the same place. Yaw is the turning of head left and right side. After eliminating images with poses using automatic pose free image selector, person authentication is done using normal and smile facial expression and their performance in effectively authenticating person is compared. Equal error rate, a quicker way to calculate the efficiency of the person authentication system is used to compare the efficiency. Equal error rate is the rate at which errors in both accepting and rejecting are equal. The system is considered more efficient if the equal error rate is lower.

1.1 Related Work

Even though there can be many landmarks in a face, eyes and mouth are considered the most important for identifying pose of a face. Color details are used in a system to identify locations in a face. Person authentication system efficiency is usually calculated using equal error rate [1]. After identifying eyes, mouth center is calculated based on eyes location [2]. A practical face based authentication system should be able to manage automatically the variations in poses [3]. Several work confirmed that authentication based on more than one modality yields better results [4]. Pose variations reduces the performance of image based identification method a lot. Large amount of data can be obtained from video than still pictures [5]. Locating the mouth area correctly is an important issue for an automatic person recognizing system [6].

1.2 Outline of Work

The total work described in this paper consists of two stages. First stage is determination of pose free image mouth region, which is described in section 2. Second stage is facial expression based person authentication which is described in section 3. The first stage, determination of pose free image mouth region includes detection of face independently, detection of nose region independently, accepting the video frame only if one face and one nose is detected, detection of two mouth corners and nose tip from the detected face, determination of mouth region with pose free image using relative position of nose region, right mouth corner and nose tip, accepting or rejecting the video frame based on mouth region, mouth corners and nose tip. The second stage, facial expression based person authentication includes mouth feature extraction and person authentication using auto associative neural network along with performance comparison in person authentication while using normal and smile facial expression. Mouth feature extraction is described in section 3.1. Person authentication using auto associative neural network is described in section 3.2. Section 4 describes the capabilities of auto associate neural network. Section 5 gives the experimental results. Section 6 concludes the paper. Figure 1 shows the outline of first stage of the proposed system diagrammatically. Figure 2 shows the outline of second stage of the proposed system diagrammatically. Figure 3 shows the interface of facial expression based person authentication system.

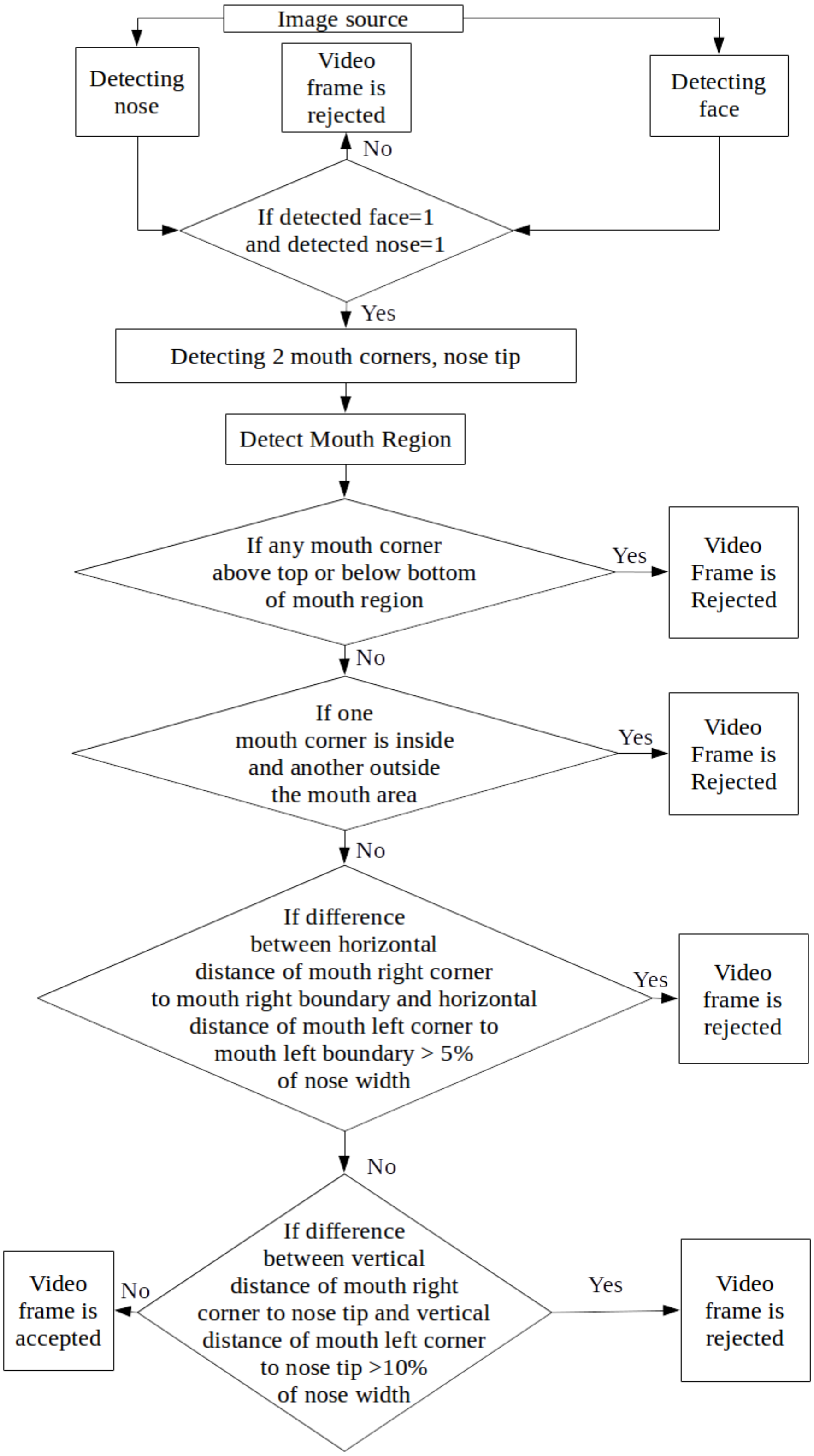


Image source
Detecting nose
Video frame is rejected
Detecting face
No
If detected face=1 and detected nose=1
Yes
Detecting 2 mouth corners, nose tip
Detect Mouth Region
If any mouth corner above top or below bottom of mouth region
Yes
Video Frame is Rejected
No
If one mouth corner is inside and another outside the mouth area
Yes
Video Frame is Rejected
No
If difference between horizontal distance of mouth right corner to mouth right boundary and horizontal distance of mouth left corner to mouth left boundary > 5% of nose width
Yes
Video frame is rejected
No
If difference between vertical distance of mouth right corner to nose tip and vertical distance of mouth left corner to nose tip >10% of nose width
No
Video frame is accepted
Yes
Video frame is rejected

Figure 1: Outline of first stage - Determination of pose free image mouth region

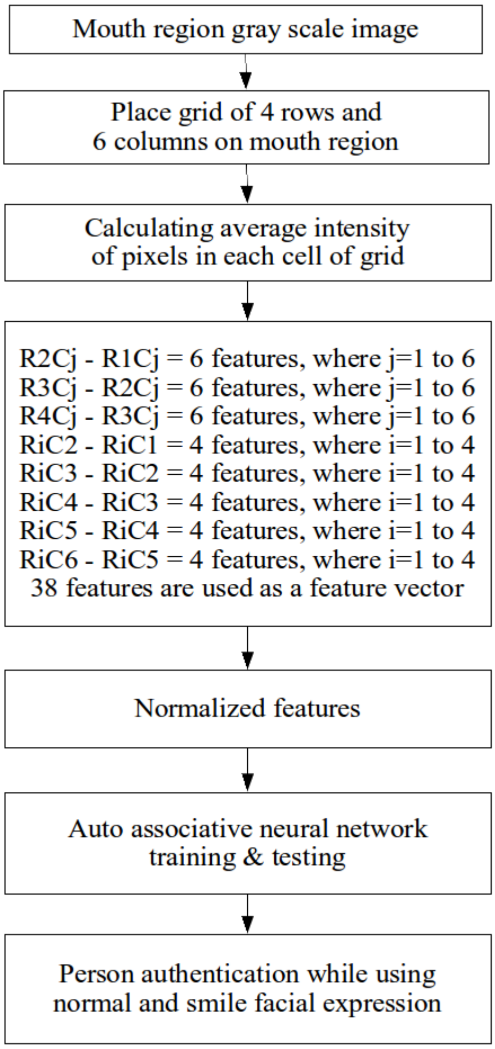
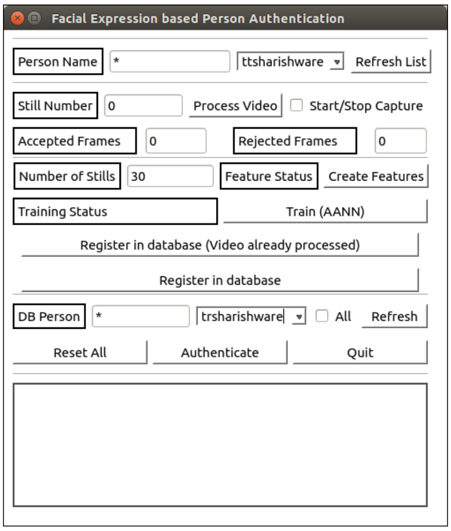

Mouth region gray scale image
Place grid of 4 rows and 6 columns on mouth region
Calculating average intensity of pixels in each cell of grid
R2Cj - R1Cj = 6 features, where j=1 to 6
R3Cj - R2Cj = 6 features, where j=1 to 6
R4Cj - R3Cj = 6 features, where j=1 to 6
RiC2 - RiC1 = 4 features, where i=1 to 4
RiC3 - RiC2 = 4 features, where i=1 to 4
RiC4 - RiC3 = 4 features, where i=1 to 4
RiC5 - RiC4 = 4 features, where i=1 to 4
RiC6 - RiC5 = 4 features, where i=1 to 4
38 features are used as a feature vector
Normalized features
Auto associative neural network training & testing
Person authentication while using normal and smile facial expression

Figure 2: Outline of second stage - Facial expression based person authentication



Person Name *
ttsharishware
Still Number 0
Accepted Frames 0
Rejected Frames 0
Number of Stills 30
DB Person *
trsharishware

Figure 3: Interface of facial expression based person authentication system

2. DETERMINATION OF POSE FREE IMAGE MOUTH REGION

Detection of face independently: Haar feature based cascade classifier proposed in [7] was improved in [8] and is used to detect face region. Figure 4 shows the detection of single face. In some situations, more than one person may be in the image. Or in rare situations some places in the image may be additionally detected as face even though only one face is available. Figure 5 shows multiple places detected as faces and even completely detecting the face wrongly. Ensuing procedures will automatically filter such video frames.

Detection of nose region independently: The detection of nose region is done independently using haar feature based cascade classifier proposed in [9] which is found to work better for nose based detection on empirical studies. Detection of multiple faces and multiple noses may happen when the face is moved faster and the camera is not capable of recording it perfectly. Ensuing procedures will automatically filter such video frames.

Accepting the frame only if one face and one nose is detected: A video frame is accepted only when an image is detected with only one face and only one nose by the above two methods. This multi modality approach increases the accuracy of the system. As the system is for single person authentication, if more than one face or more than one nose is detected in a video frame, that frame is rejected. This rejects many frames which increases the accuracy of person authentication.

Detection of two mouth corners and nose tip from the detected face: The automatically detected single face region is given as input to a detector of facial landmarks learned by the structured output SVM based on the Deformable Part Models proposed in [10]. It gives us eight landmarks from the face which includes face center, right eye inner corner, left eye inner corner, mouth right corner, mouth left corner, right eye outer corner, left eye outer corner and nose tip. Among these, three landmarks namely mouth right corner, mouth left corner and nose tip are taken into account for this proposed work.

Determination of mouth region in pose free image using relative position of nose region, right mouth corner and nose tip: The width of the detected nose region is set as the width of the mouth. As the ratio of width of nose by width of face will always be a constant, this method gives same location of mouth for a particular person face for any number of sessions. Moreover this will show the same location of mouth even when the scaling of the face occurs due to the variation in the distance between the camera and the person face. This is because when scaling of face occurs, both mouth and nose will also get scaled simultaneously. The middle of nose tip and mouth right corner is calculated and used as the top boundary of the mouth. For any number of sessions for a particular person this will mark the same location when the face is free from pose. Using of only one mouth corner to calculate the mouth top will act as an important point to find faces with poses. The mouth height is calculated as half of its width. As it is also a relative quantity to detected nose width, it varies according to the scaling of the face, but always shows the same location for a particular person. Figure 6 shows detected single face with marked mouth region. Figure 7 shows detected single face with marked three landmarks with a dot.

Accepting or rejecting the video frame based on mouth region, mouth corners and nose tip: Now the detected two mouth corners and nose tip from the detected face is used along with created mouth region to further reject video frames which has poses. A video frame is checked whether any one of its mouth corner is above the top or below the bottom of the marked mouth region. If it is so that video frame is rejected as this situation usually appears during roll. If one mouth corner is inside the marked mouth region and another outside the marked mouth region, that video frame is also rejected. Usually this situation occurs during yaw. If the difference between the horizontal distance of mouth right corner to marked mouth region right boundary and horizontal distance of mouth left corner to marked mouth region left boundary is more than 5% of the width of the nose, then that video frame is rejected. If the difference between the vertical distance of mouth right corner to nose tip and vertical distance of mouth left corner to nose tip is more than 10% of the width of the nose, then that video frame is rejected. These situations arise usually during yaw. Nose width based threshold is used instead of fixed pixels to overcome issues based on scaling. The remaining video frames are accepted as pose free video frames. Figure 8 shows an accepted video frame and Figure 9 shows a rejected video frame.

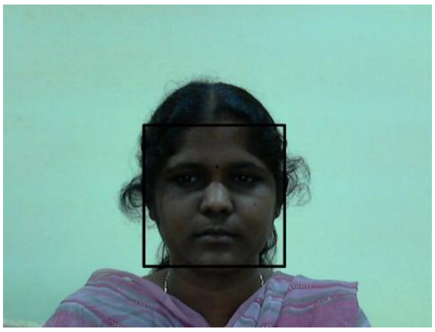


Figure 4: Detection of single face

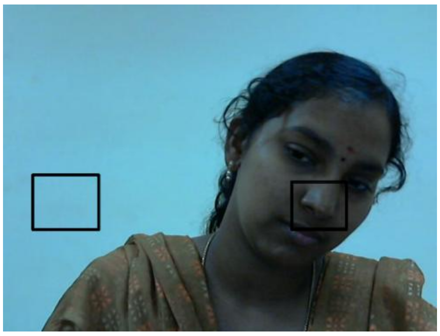


Figure 5: Multiple places detected as faces

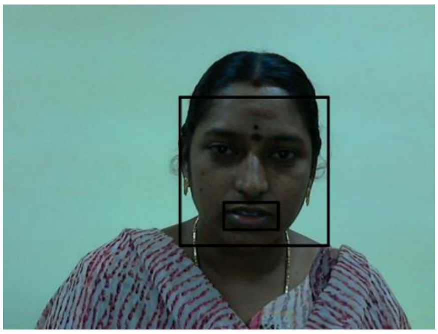


Figure 6: Detected single face with marked mouth region

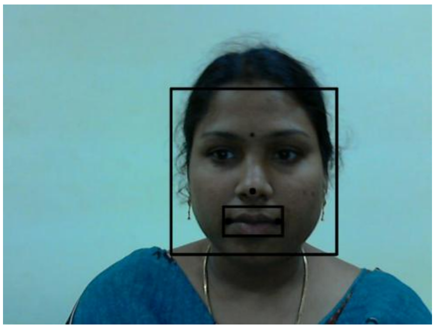


Figure 7: Detected single face with marked landmarks

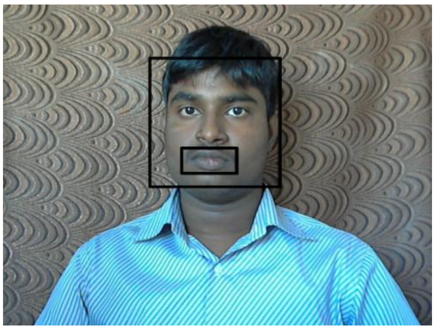


Figure 8: Accepted video frame

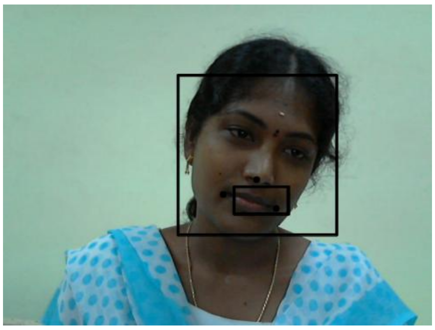


Figure 9: Rejected video frame

3. FACIAL EXPRESSION BASED PERSON AUTHENTICATION

3.1 Mouth Feature Extraction

The pose free image mouth region is taken and converted into gray scale. A grid of six columns and four rows is formed. Number of columns in the grid and number of rows in the grid decides the size of each cell in the grid. If size of cell in the grid is too large that is if number of columns and rows of cells are less in the grid, then number of features will be less which will affect person authentication efficiency. If size of cell in the grid is too small that is if number of columns and rows of cells are more in the grid, then number of features will be more, but even a slight change in position in the image between the train and test will lead to inaccurate results. Hence an optimum value is obtained by using empirical studies. The average intensity of pixels in each cell of the grid is calculated to form a grid of four rows and six columns with an average intensity value in each grid cell. The first set of six features is generated by subtracting value in every column of first row from value in corresponding column of second row. The second set of six features is generated by subtracting value in every column of second row from value in corresponding column of third row. The third set of six features is generated by subtracting value in every column of third row from value in corresponding column of fourth row.

The fourth set of four features is generated by subtracting value in every row of first column from value in corresponding row of second column. The fifth set of four features is generated by subtracting value in every row of second column from value in corresponding row of third column. The sixth set of four features is generated by subtracting value in every row of third column from value in corresponding row of fourth column. The seventh set of four features is generated by subtracting value in every row of fourth column from value in corresponding row of fifth column. The eighth set of four features is generated by subtracting value in every row of fifth column from value in corresponding row of sixth column. So totally 38 (18+20) features from mouth region of horizontal and vertical variations are generated. Table 1 gives the pseudo code of the feature extraction algorithm. Figure 10 shows this feature extraction procedure diagrammatically. As the features are generated by subtracting average intensity it will have tolerance to intensity changes. To have more tolerance to intensity changes among images of same persons the resultant features are normalized between new minimum 0 and new maximum 1 using the formula

$$y_k = (f_k – f_{min}) * (y_{max} – y_{min}) / (f_{max} – f_{min}) + y_{min}$$

where

y_k= Normalized intensity

f_k = Current intensity

f_{min} = Current minimum value

f_{max} = Current maximum value

y_{min} = New minimum value, that is 0

y_{max} = New maximum value, that is 1.

Table 1. Pseudo code of feature extraction algorithm

```
FOR k = 1 to (m-1)*n
        FOR i = 1 to (m-1)
                FOR j = 1 to n
                        fk = R(i+1)Cj – RiCj
                ENDFOR
        ENDFOR
ENDFOR
FOR k = [(m-1)*n+1] to [(m-1)*n+m*(n-1)]
        FOR j = 1 to (n-1)
                FOR i = 1 to m
                        fk = RiC(j+1) – RiCj
                ENDFOR
        ENDFOR
ENDFOR

Here fk is the feature vector, where
fk = ( f1, f2, . . . , f[(m-1)*n+m*(n-1)] ), where
Ri is ith row
Cj is jth column
RiCj is average values in grid cell of ith row and jth
column
i is equal to 1 to m
j is equal to 1 to n
m is equal to number of rows in the grid
n is equal to number of columns in the grid
In this work, m = 4, n = 6
[(m-1)*n+m*(n-1)] = 38 and
fk = (f1, f2, . . . , f38)
```

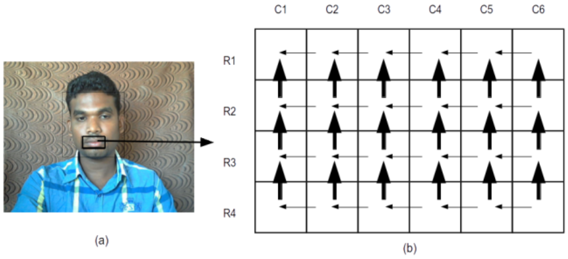


Figure 10: (a) Mouth region (b) Feature extraction

3.2 Person Authentication using Auto Associative Neural Network

In the first session, videos are recorded for ten persons, out of which five are males and five are females. The video is recorded in normal room conditions with normal and smile facial expression. During both normal and smile facial expression sessions, persons are asked to move their head enough so that to include the three types of poses namely tilt, roll and yaw. Pitch or tilt is moving the head up and down. Roll is tilting the head sideways by facing the camera while the nose is stationary. Yaw is left and right side turning of head. From each of the twenty recorded video, pose free images are automatically sensed and stills of mouth regions are extracted automatically in first stage as explained in section 2. Features are generated as explained in section 3.1 from the mouth region of the still images and used to train an auto associative neural network. Similar to first session, in the second session also twenty videos are generated from the same set of ten persons. Session one and session two has a time gap of twenty days to ensure a real life situation. No prior information is given to persons about the recordings during both the sessions for more real life situations. From each of the twenty recorded video, pose free images are automatically sensed and stills of mouth regions are extracted automatically and used for testing in the auto associative neural network for person authentication. To compare the performance of the person authentication system while using normal and smile facial expression equal error rate is used. The performance is considered better if the equal error rate is lower. Equal error rate, also called as crossover error rate is the rate at which false acceptance rate and false rejection rate are equal. False acceptance rate, also called as false match rate is the percent of impostors who are falsely accepted as authentic persons upon a particular threshold value. False rejection rate, also called as false non match rate is the percent of authentic persons falsely rejected as impostors upon a particular threshold value. While using normal face, equal error rate for person authentication is 0.32% at threshold value 0.73. The corresponding graph is shown in Figure 11. While using smile facial expression, equal error rate for person authentication is 0.4% at threshold value 0.83. The corresponding graph is shown in Figure 12.

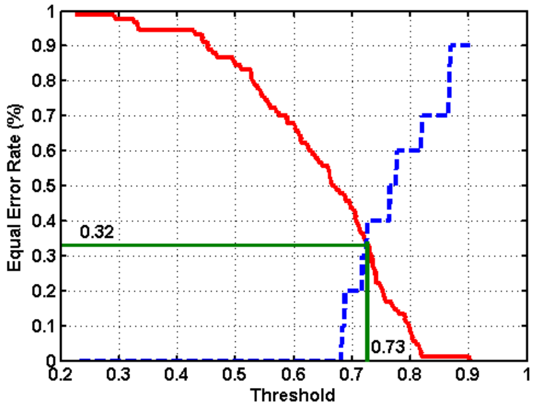


Figure 11: Equal error rate while using normal face

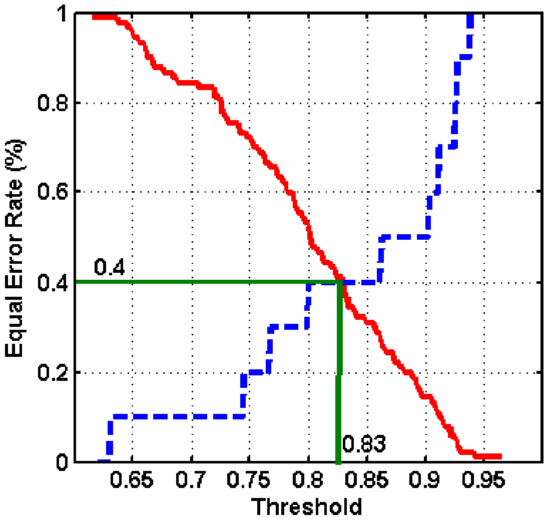


Figure 12: Equal error rate while using smile face

4. AUTO ASSOCIATIVE NEURAL NETWORK

A feed forward artificial neural network has an input layer, output layer, and one or more hidden layers. Each layer consists of processing units, where each unit represents the model of an artificial neuron, and the interconnection between two units has a weight associated with it. Artificial neural network models with different topologies perform different pattern recognition tasks. Auto associative neural network models are feed forward artificial neural networks. Distribution of the input data can be captured using it, by identity mapping the input space [1]. The ability of the auto associative neural network model to capture distribution is described in this section. The five layer auto associative neural network model shown in Fig. 13 has three hidden layers. In this network model the input or output layer has fewer units than the second and fourth layers. The input and output layer has more units than the third layer. Hence compression of input vectors to lower dimension occurs. The number of units in the input and output layers are same. The processing units in the compressed hidden layer can be linear or nonlinear, but the processing units in the hidden layers before and after the compressed hidden layer are nonlinear. The error between the desired output and the actual vectors is minimized. Hence the shape of the hyper surface obtained by the projection onto the lower dimensional space is determined by the cluster of points in the input space. For the two dimensional data shown in Figure 14 (a) for the network structure 2L 10N 1N 10N 2L, Figure 14 (b) shows the space spanned by the one dimensional compression layer. Here L denotes a linear unit, N denotes a nonlinear unit and the integer value indicates the number of units used in that layer. Back propagation algorithm is used for training the network. Mapping of the given input points due to the one dimensional compression layer is indicated by the solid lines in Figure 14 (b). Hence it can be said that the distribution of the input data is captured by auto associative neural network depending on the constraints imposed by the structure of the network. The error of each input data point can be plotted in the form of some probability surface to visualize the distribution better. In the input space, the error e_i for the data point i is plotted as $p_i=\exp^{-(ei/\alpha)}$, where α is a constant. We call the resulting surface as probability surface, even though i is not strictly a probability density function. For smaller error e_i, the plot of the probability surface shows large amplitude which indicates better match of the network for that data point. The shape the error surface takes in both the cases shows the constraints imposed by the network. The characteristics of the distribution of the input data captured by the network can be studied using the probability surface. Naturally, one would like to achieve the best probability surface, best defined in terms of some measure corresponding to a low average error. In this work, to capture the distribution of facial feature vectors, the distribution capturing ability of the auto associative neural network model is exploited. To capture the distribution of the facial feature vectors, a five layer auto associative neural network model is used. For capturing the facial features of a person, the structure of the auto associative neural network model used in our study is 38L 76N 19N 76N 38L, where L denotes a linear unit and N denotes a nonlinear unit. To minimize the mean square error for each facial feature vector, the back propagation learning algorithm is used to adjust the weights of the network.

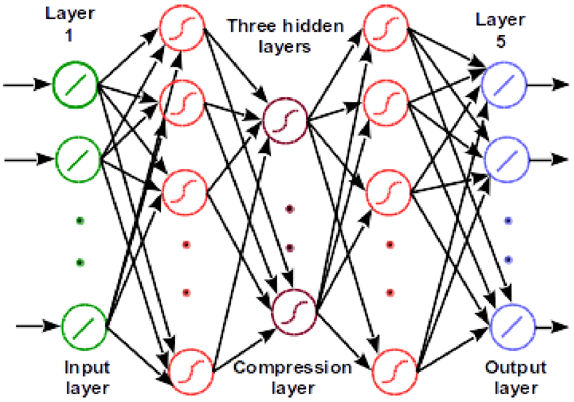


Figure 13: A five layer auto associative neural network

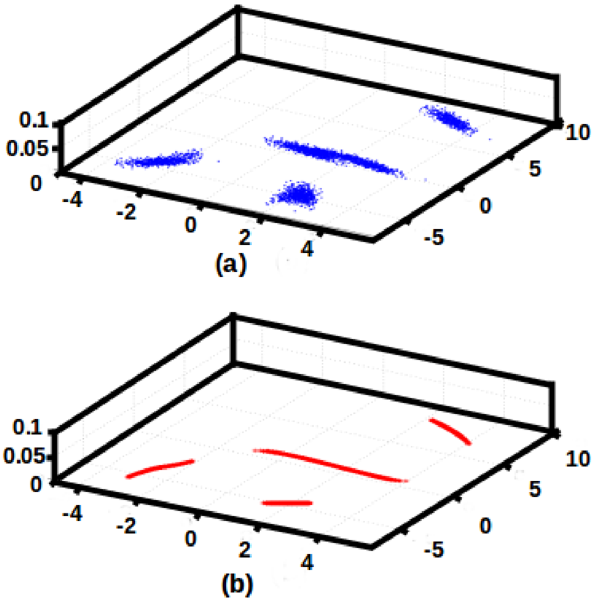


Figure 14: Distribution capturing capability of auto associative neural network

5. EXPERIMENTAL RESULTS

The performance of the system is tested in forty videos of resolution 640 x 480 taken at fifteen frames per second of ten subjects taken under two different normal room conditions in two sessions with variations in poses like yaw, pitch and roll. The duration of the videos are not fixed, as it is decided by number of pose free images collected by the system. Once thirty pose free images have been collected, the processing of video stops. Out of thirty video frames, all the face images collected by the automatic pose free image selector as pose free face images are pose free face images. But among the video frames rejected as images with pose, there are some images which can be accepted as pose free images. Even though such false rejection may little slow down the total speed of a facial expression based person authentication system which can be build using this automatic pose free image selector, it will not reduce its accuracy. Audio is not recorded which reduces the file size. Background need not be bothered about, as the mouth area alone will be sensed automatically and used for feature extraction. Variation in the size of the face due to the difference in distance from the camera does not disturb the accuracy as it is sensed by using improved haar feature based cascade classifier. From the collected pose free images, facial features are extracted. For training a subject, thirty pose free frames are selected automatically as explained in first stage and 38 facial feature vectors are extracted from each pose free video frame. All the extracted facial feature vectors are now normalized and given as input to the auto associative neural network model 38L 76N 19N 76N 38L for capturing the distribution. A single presentation of all the training vectors to the auto associative neural network is called one epoch. The network is trained for 3000 epochs. The training takes less than a minute in a Pentium Dual Core 2.3GHz processor laptop loaded with Ubuntu 12.04LTS, OpenCV 2.3 and Qt 4.8.0. For the best matching network structure, if the same data given for training is used in testing, the confidence score will be near one for all the persons. It means that the auto associative neural network model network structure captures the distribution of the features perfectly. Using this concept, the best network structure is decided through empirical observation. For computing the normalized squared error, the output of the model is compared with the input. The normalized squared error (e) for the feature vector y is given by, $e=||y-o||^2/||y||^2$, where o is the output vector given by the model. The confidence score (c) is calculated from error (e) using the formula c = exp(-e). The average confidence score is calculated from the claimant auto associative neural network model of all the facial features of the automatically detected thirty pose free face images. For person authentication, if the confidence score for training and testing data for the specific person is more than a particular threshold, then that claimant is authenticated or else considered as pretender. A common threshold is set, but threshold can also be arrived for individual persons. To measure the performance of the system, the confidence score is calculated for testing data of each person with training data of all the persons. The performance is decided by using equal error rate. The equal error rate for person authentication using normal face is 0.32% at threshold value 0.73 and smile facial expression is 0.4% at threshold value 0.83. To compare the performance of facial expression based person authentication system, the equal error rate with its corresponding threshold values are shown in table 2. Lower equal error rate implies better performance. The comparative equal error rate chart for performance of normal and smile facial expression is shown in Figure 15.

Table 2. Performance of person authentication system

Feature	Equal error rate (%)	Threshold
Normal	0.32	0.73
Smile	0.4	0.83

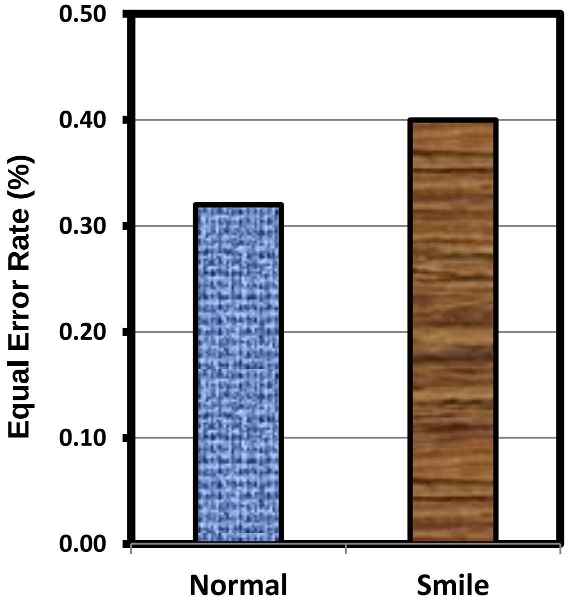


Figure 15: Comparison of equal error rate between normal and smile facial expression

6. CONCLUSION

In this paper, automatic pose free image selector was used to extract automatically pose free face images from video taken in normal room condition. Mouth region was determined automatically and features were extracted. The system used two stages. In first stage, automatic pose free image selector was used to collect pose free face images from videos of ten persons taken in two sessions each with normal and smile facial expressions with poses. Testing on images taken from forty videos of resolution 640 x 480 the system identified and extracted pose free face images automatically which were 100% perfect pose free face images. In stage second, automatically selected pose free images of mouth during normal and smile facial expression from the twenty videos of first session were used for training an auto associative neural network. Images from the second session of twenty videos were used to test for person authentication. The results clearly shows that normal face gives better performance than smile facial expression for person authentication by accepting authentic persons and rejecting impostors. Equal error rate was used to calculate the performance of the person authentication system. Equal error rate for person authentication using normal face is 0.32% at threshold value 0.73 whereas with smile facial expression it is 0.4% at threshold value 0.83. The person authentication system would be considered more efficient if the equal error rate value was lower. Based on the result it was clear that smile facial expression was less efficient than normal for automatic facial expression based person authentication. The reason may be because of too much variation in the mouth region in between smiles of the same person. In other words, smiles may be less similar between sessions. The future work is to compare the performance in person authentication using other facial expressions like anger, visual speech with normal face. The concepts in this work can lead to create a more perfect facial expression based person authentication system.

7. REFERENCES

[1] Palanivel, S. and B. Yegnanarayana, "Multimodal person authentication using speech, face and visual speech", Computer Vision and Image Understanding, vol. 109, no. 1, pp. 44–55, Jan. 2008; doi:10.1016/j.cviu.2006.11.013.

[2] Balasubramanian, M., S. Palanivel and V. Ramalingam, "Real time face and mouth recognition using radial basis function neural networks", Expert Systems with Applications, vol. 36, no. 3, pp. 6879-6888, Apr. 2009; doi:10.1016/j.eswa.2008.08.001.

[3] Xudong Xie and Kin-Man Lam, "Face recognition using elastic local reconstruction based on a single face image", Pattern Recognition, vol. 41, no. 1, pp. 406-417, Jan. 2008; doi:10.1016/j.patcog.2007.03.020.

[4] Roland Hu and R.I. Damper, "Optimal weighting of bimodal biometric information with specific application to audio-visual person identification", Information Fusion, vol. 10, no. 2, pp. 172-182, Apr. 2009; doi:10.1016/j.inffus.2008.08.003.

[5] Federico Matta and Jean-Luc Dugelay, "Person recognition using facial video information: A state of the art", Journal of Visual Languages and Computing, vol. 20, no. 3, pp. 180-187, Jun. 2009; doi:10.1016/j.jvlc.2009.01.002.

[6] Meng Li and Yiu-ming Cheung, "Automatic lip localization under face illumination with shadow consideration", Signal Processing, vol. 89, no. 12, pp. 2425-2434, Dec. 2009; doi:10.1016/j.sigpro.2009.05.027.

[7] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hawaii, IEEE Xplore Press, vol.1, pp. I-511-I-518, 08-14 Dec. 2001; doi:10.1109/CVPR.2001.990517.

[8] Rainer Lienhart and Jochen Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection", Proceedings of the 2002 International Conference on Image Processing, USA, vol. 1, pp. 900-903, 22-25 Sep. 2002; doi:10.1109/ICIP.2002.1038171.

[9] Castrillon, M., O. Déniz, C. Guerra and M. Hernández, "ENCARA2: Real-time detection of multiple faces at different resolutions in video streams", Journal of Visual Communication and Image Representation, vol. 18, no. 2, pp. 130-140, Apr. 2007; doi:10.1016/j.jvcir.2006.11.004.

[10] Michal Uřičář, Vojtěch Franc and Václav Hlaváč, "Facial Landmarks Detector Learned by the Structured Output SVM", Proceedings of the 7th International Joint Conference on Computer Vision Theory and Applications, on Computer Graphics Theory and Applications and on Information Visualization Theory and Applications, Springer Berlin Heidelberg, Italy, pp. 383-398, 24-26 Feb. 2012; doi:10.1007/978-3-642-38241-3.