ACTIVE APPEARANCE BASED POSE AND ILLUMINATION VARIANT FACE RECOGNITION FROM VIDEO SEQUENCES

Anupkumar Awaradi[1], Deepak Kumar[2]

[1]*Department of E & C, Canara Engineering College, Mangalore - 574219, India*
[2]*Department of E & C, DSATM, Bangalore - 560082, India*

Abstract

Surveillance video provides security by monitoring the people entering and leaving the premises. Face recognition in video may act as a tool to record the information with minimal manual intervention. The major difficulty is that the faces are captured in different pose and illumination in the video. The recognition of these faces from the video sequences is complex. In this paper, we have addressed the face recognition problem in different pose and illumination.We have created a set of ten video sequences for ten people with different pose and illumination and extracted the faces at different angles and illumination from these video sequences to create a face database. Active Appearance Model (AAM) is used to create a feature vector for each face that exists in the database. Artificial Neural Network (ANN) is trained to recognize the face from the appearance feature vectors. The proposed method reduces the computational complexity involved in recognizing the faces in different pose and illumination. We have achieved a recognition rate of 96.1% on the test set of the face database.

Keywords: *Face Recognition; Active Appearance Model; Neural Network; Video Sequences;*

----------***----------

1. INTRODUCTION

Identifying and recognizing a face from a stored database of images or video is termed as face recognition [1],[3],[4],[5]. The two main challenges in face recognition are illumination and pose variation [2],[6],[7]. Many methods have been developed to provide feasible solution to recognize face. The performance of the solution is seriously degraded when face undergoes variation in pose or illumination [9],[10],[12],[13], since the appearance of face is changed drastically for small changes in pose or illumination which makes the recognition task very difficult [2],[3],[14]. Face recognition may be used in applications such as Entertainment, Smart Cards, Information & Security, Law enforcement, and Surveillance.

The complexity of face recognition increases with different pose, illumination, facial expression, occlusion, hairstyle, ageing, scaling, and twins. The task of extracting features from face becomes difficult. In this paper, we propose a method to extract the features from face and recognize it. We have reduced the computational cost involved in recognizing the face [4].

2. RELATED WORK

Face recognition is not new to the field of pattern recognition. Several methods and algorithms are developed in past using novel techniques [3], [7]. Algorithms are also developed to solve the pose and illumination variation [7]. In general, principal component analysis (PCA) is used for face classification.

Jindal and Kumar have proposed a method for face classification using PCA with Artificial Neural Network (ANN) [5]. Eigen face is computed using the dimensionality reduction technique, PCA. ANN is trained to classify the faces using Eigen face as the input features. The proposed method can recognize only the upfront facial images and does not address the problem of pose and illumination variation.

Roy-Chowdhury and Xu proposed a framework to solve the pose and illumination using analysis-by-synthesis [7]. Video sequences are analysed to synthesize the face which is captured. 3D model of each face is constructed for classification. The construction of 3D model is computationally intensive and involves multiple frames with different pose.

Chai et al. proposed a method to handle both pose and lighting condition simultaneously [6]. The pose and lighting condition of each image is calibrated to a pre-set reference condition through an illumination invariant 3D face reconstruction.

To combine the advantages of 2D face image and 3D reconstructed face image, Rama and Tarres proposed a face recognition using Partial PCA (P^2CA) [1]. A gallery of images is created from a minimal set of pose variation (0°, ±45°°, ±90°) and for classification, varied angles (0°, ±30°, ±45°, ±60°, ±90°) and with different illumination(natural or environment lighting, strong light and frontal mid light) are used.

Prasanna and Hegde proposed a method to recognize a face using nearest neighbour classification to solve pose and illumination problem from the video sequences [4]. A database is created to store the features extracted from the images in the classification stage. Here, UPC Face database

is used for experimentation [8]. Since, the nearest neighbour approach evaluate the input sample against all the stored samples. This results in more computation and increases the time taken to classify a face.

The pose and illumination variation in face is solved by constructing 3D facial representation of face. In construction, same face with multiple pose and illumination are necessary and then all of them are aligned to a reference frame. This approach is computationally expensive.

3. PROPOSED METHOD

A computationally inexpensive method is developed to solve the pose and illumination variation and also recognize the face. The proposed method is comprised of two stages: (i) Development of features database and (ii) Recognition of face image.

3.1 Development of Features Database

The variation in pose and illumination of a face is captured by taking several face images of the same person at multiple angles and light conditions. We have created videos for the ten different people and videos are captured using web camera under different lighting conditions. A person is made to sit in front of the camera for ten seconds and the person has to change the direction of his/her view from the right shoulder to the left shoulder or vice versa. A set of videos of the same person is captured again under different lighting conditions such as normal light, front facial lighting by placing a lamp behind the camera, and partial lighting by placing a lamp at the corners of the table. The procedure for the creation of features database is explained in the following subsections.

1) Normalized Images: The frame size of the captured videos is 640x480 pixels. The face in the frame can be localized by a face recognition algorithm. The localized face is cropped from the frame, but the size of the face may vary from frame to frame. The cropped images are resized to 100x100 pixels to have uniformity in the database.

2) Feature Extraction using AAM: Active appearance models (AAM) are generative models that capture the image properties of an object. The main properties of an object in an image are shape, texture, and pixel intensities. In order to extract the features, both the shape and pixel intensities are manually marked by placing landmarks in AAM. Face may have distinctive face regions like eyebrows, eyes, nose, lips, and chin. An image is marked with n landmark points around the face regions. Figure 1 shows the sample images from our database marked with landmark points for different pose and illumination.

Face is modelled as a shape and generally, a shape model can be represented as:

$$x = [x_1, x_2, ..., x_n, y_1, y_2, ..., y_n] \quad (1)$$

where, (x_i, y_i) is the 2D landmark position on the image.

We have manually labelled 60 landmarks on 900 images (the total number of images in our database). The facial features in each image is extracted from the landmark points placed on the image. Let S^j denote the facial vector for the j^{th} image in our database and can be represented as:

$$S^j = [p_1, p_2, ..., p_n] \quad (2)$$

where, p_i is the pixel value at the landmark point (x_i, y_i).

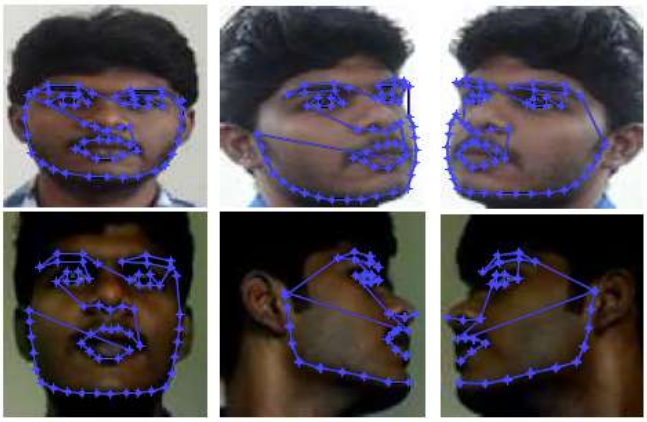

Fig. 1. Manually placed landmarks on the face images with different pose and illumination. Top row: The face images in normal lighting condition and the pose views are 0°, −45°, and +45°. Bottom row: The face images in weak lighting condition and the pose views are 0°, −90°, and +90°.

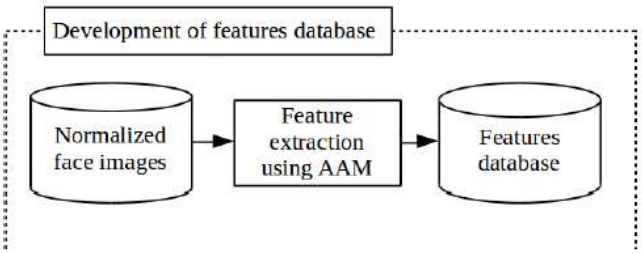

Fig. 2. Creation of database from the normalized face images. Normalized face images are images cropped from the video frames and rescaled to 100x100 pixels. Landmark points are manually placed to extract active appearance model (AAM) of the face. Features database consists of feature vectors from the 900 images.

3) Database Creation: UPC Face database consists of images corresponding to 10 persons with 27 images per person acquired under different pose views (0°, ±30°, ±45°, ±60° and ±90°) and three different illuminations (natural light, Weak light source, and Less Brightness) [8]. In UPC Face database, the images are acquired at a particular view and illumination. In our database, the face of a person is acquired by moving the face from right to left or from left to right. We could not mark the face at exact view, but the frames used in normalization have approximated views. The shape model and facial features have positional and pixel value attributes, respectively. To form a single feature vector for a face, the position and pixel values are rescaled before concatenation. The landmark points are subtracted by

the mean value of all the landmark points and rescaled using the maximum value between the mean value and all the landmark points. The pixel values in the facial features are scaled by 255. Let f^j denote the feature vector for the j^{th} face in our database and can be represented as:

$$f^j = [x^j , y^j , s^j] \quad (3)$$

where, x^j , y^j , and s^j are the scaled attributes for a face. Figure 2 shows the overview of database creation. The feature vectors (f^j) are stored as features database. Our database consists of 900 images of 10 persons with different angles.

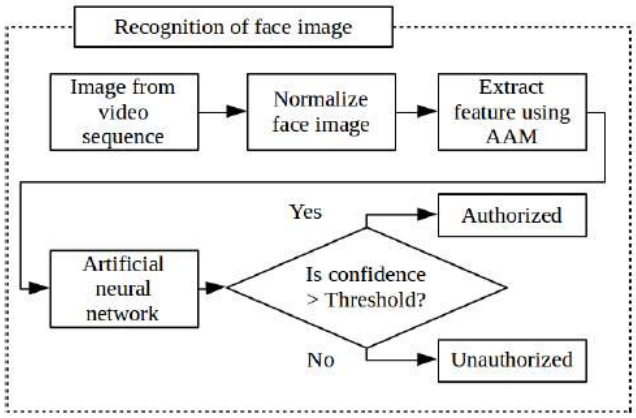

Fig. 3. A video frame is selected from the acquired video sequence. The face in the video frame is normalized and features are extracted from it. The extracted features are fed to the trained ANN for classification. The person in the video may be authorized or unauthorized depending on the confidence value provided by the ANN.

3.2 Recognition of Image

We may have to recognize a person at different pose and illumination. We have to follow similar process for extracting the feature vector from the video frame. The extracted feature vector is provided as input to artificial neural network (ANN) to identify the person. We consider the output values of ANN as confidence values for each person in the database.

A threshold value is fixed to authorize the person. If the confidence value of a test face is below the threshold value, then the person is classified as an unauthorized person. Figure 3 shows the block diagram of recognition process.

1) Artificial Neural Network: Human brain consists of complex interconnected neurons to process various tasks [5], [15]. New models were developed to mimic the biological neuron and resulted in Artificial Neural Network (ANN). An ANN learns the correlation between input and target values. There are different types of ANN for example, Multi-layered Perceptrons (MLP), Self Organizing Maps (SOM). A multilayered feed forward neural network consists of the three layers namely input layer, hidden layer, and output layer. Figure 4 shows an example for a multi-layered feed forward neural network. Back propagation is a supervised learning algorithm to train the neural network. Back propagation is an iterative process and the mean square error is reduced in each iteration by updating the weights of neurons. The iterative process is stopped when the error falls below the tolerance level. In our proposed method, we provide the features from the features database as input and the person as target to train the artificial neural network.

The multi-layered ANN has different processing elements in each layer. The description of each processing elements is as follows [5]:

- Input to the neuron: The input to the neurons are connected through the weights. The weight for the connection between i^{th} input feature and k^{th} neuron is represented as w_{ki}. The activation potential a_k for the k^{th} neuron is calculated as follows,

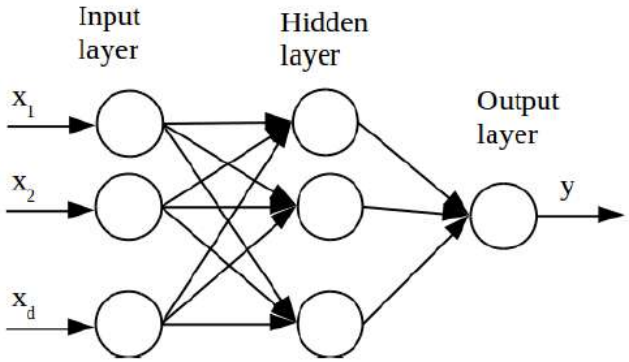

Fig. 4. An example for a multi-layered feed forward neural network. The three layers namely input, hidden and output layers perform specific task pertaining to that layer. Input layer passes the input values to the hidden layer. Hidden layer maps the input values to a set of patterns. Output layer uses these patterns with the trained weights to classify the input values as the probable target.

$$a_k = \sum_{i=1}^{d} w_{ki} x_i + w_{k0} \quad (4)$$

where w_{k0} is the bias for the k^{th} neuron and d is the number of inputs connected to the neuron. In our proposed method, d = 180.

- Mathematical Function: The output value of the neuron is a non linear function of the activation potential. The output of k^{th} neuron is calculated as follows,

$$y_k = f(a_k) \quad (5)$$

where, f () is the non linear function used to map the activation potential into the output.

The most commonly used non linear functions are binary, ramp, log-sigmoid, and tangent hyperbolic functions. In our proposed method, the input layer neurons use ramp function, the hidden layer neurons

use tangent hyperbolic function, and the output layer neurons use log-sigmoid function. The output of log-sigmoid function is considered as a confidence value of the neuron for an input face.

- Back propagation and weight update: The input provided to the ANN is processed to generate an output at the output layer. The output may not always match with the target or reference output. The weights of the neurons have to updated or modified to match the generated output from the ANN with the target or reference output. The error or difference between the generated output and the reference output is used to train the neural network. Figure 5 shows the block diagram for weight update in the neural network. In other words, the error value is used to update the weights of the neurons. To stop the iterative process of weight update, The mean square error (MSE) is used as the criteria. The MSE is calculated from all the M training samples.

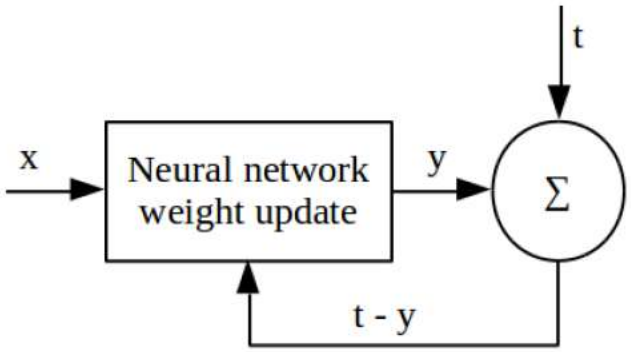

Fig. 5. Update of weights in the multi-layered neural network through back propagation. Here, x is the input value, y is the generated output by the neural network, and t is the target value.

$$MSE = \sum_{j=1}^{M} \| t_j - y_j \|^2 \qquad (6)$$

where, t_j is the target value for the j^{th} input sample and y_j is the output value from the neural network. In our proposed method, the j^{th} input sample is f^j stored in the features database.

4. EXPERIMENTAL RESULTS

The proposed face recognition method is implemented in the working platform of MATLAB (version 8.1) and tested in 2.20 GHz. In our proposed method, we have created a features database of 900 faces with different pose and illumination of ten persons. Each face sample has 180 attributes stored a feature vector in the features database. In our performance evaluation, we have split the features database in the ratio 0.6, 0.2, and 0.2 as training, validation, and testing set, respectively.

Table I provide the list of parameters used to train the neural network. Ten neurons are used in the output layer to classify the ten people in the database. The neuron which outputs the maximum value among the neurons, after training, is considered as the classified face which is mapped to the person.

4.1 Performance Analysis

Figures 6, 7, and 8 show the confusion matrices for the training, validation, and testing images, respectively. From the figures, we can observe that only 2, 6, and 7 samples are misclassified in training, validation, and testing set, respectively. In total 15 faces have been misclassified out of 900 faces. This shows our proposed method has high performance rate. During the training period the number of neurons in the hidden layer was set to 20. The overall performance achieved with 20 neurons was not good. To improve the performance, we increased the number of neurons in the hidden layer to 25. With the increase in the number of neurons, ANN could map the features in the better way and resulted in improved overall performance. Table II lists the performance value for the different number of neurons in the hidden layer.

Table 1: Parameters used to train the Neural Network.

Sl. No	Parameter	Value
1	Number of layers	3
2	Number of inputs	180
3	Number of outputs	10
4	Number of neurons in the input layer	180
5	Number of neurons in the hidden layer	25
6	Number of neurons in the output layer	10
7	Maximum number of epochs	1000
8	Transfer function in the output layer	Log-sigmoid

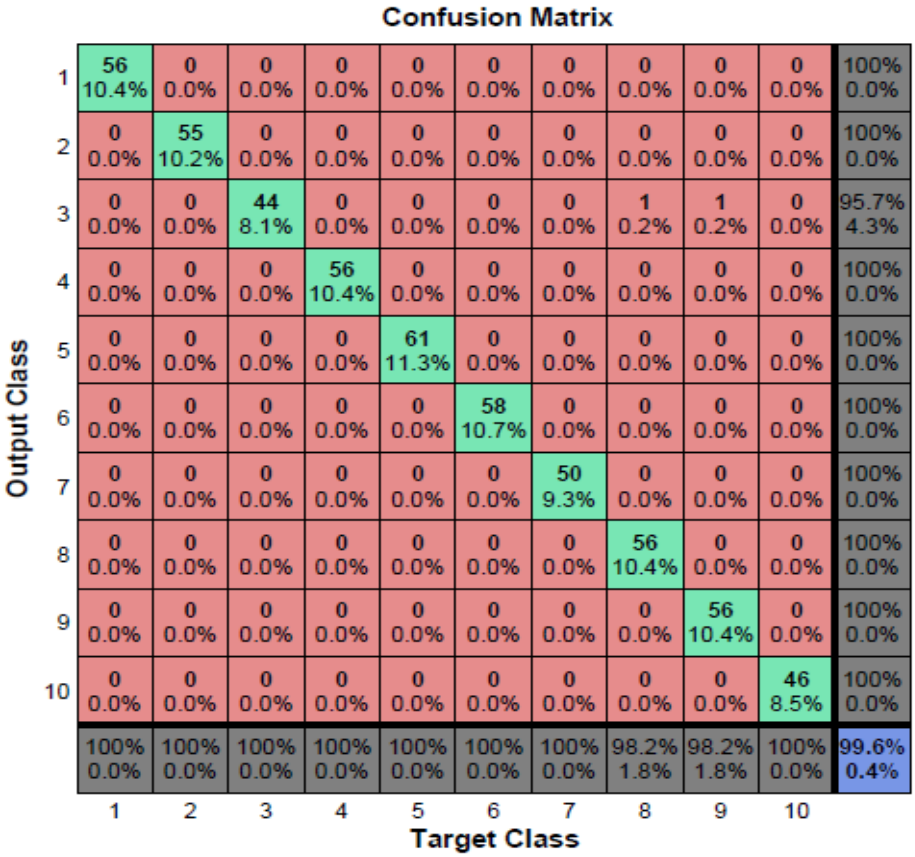

Fig. 6. Confusion matrix for the training images

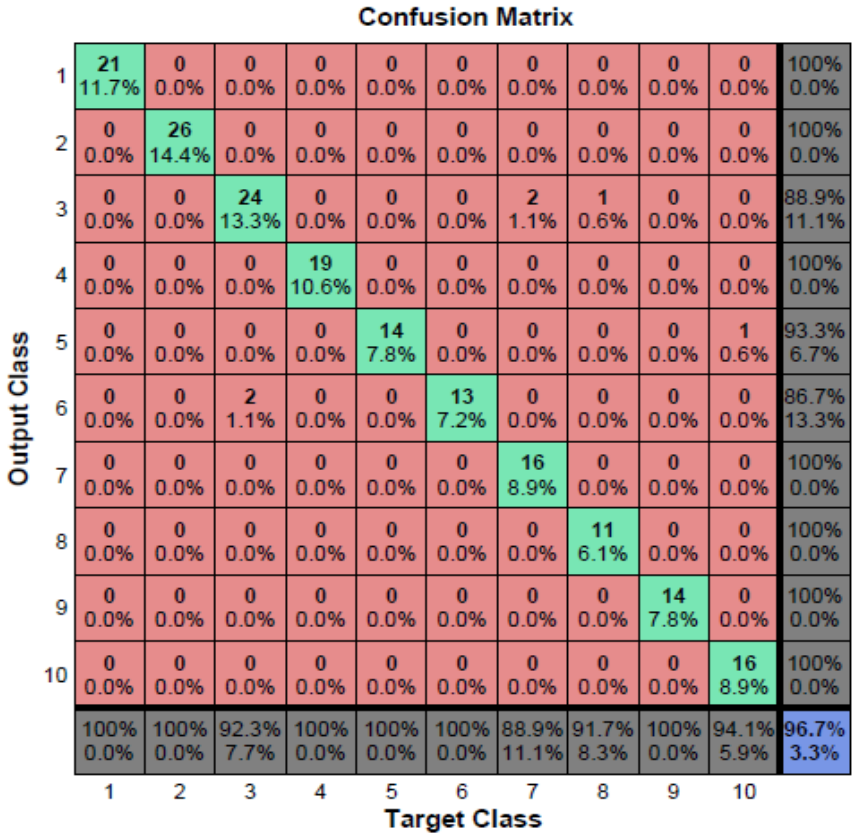

Fig. 7. Confusion matrix for the validation images

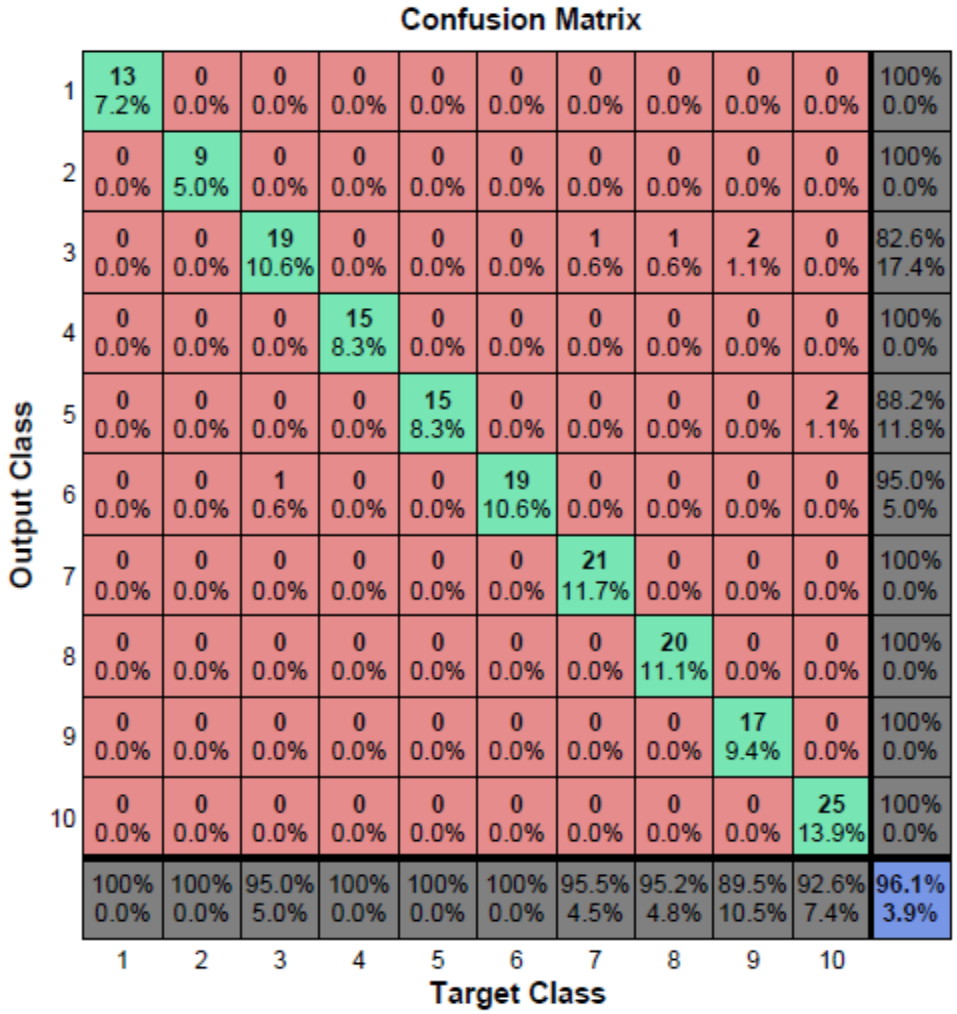

Fig. 8. Confusion matrix for the testing images

Table 2: Performance evaluation with different number of neurons in the hidden layer

Number of neurons in the hidden layer	Overall performance (%)
20	94.0
25	98.3

5. CONCLUSION

We have proposed a simple face recognition method which has reduced computational complexity cost. The proposed method overcomes the drawback of high computational cost such as building 3D facial model and nearest neighbour classifier. The proposed method has overall accuracy of 98.3% using ANN. In our future work, we would like to change the pattern recognition algorithm from ANN to support vector machines (SVM). SVM provides more optimal classification than ANN. Face recognition is non-invasive mode of authentication and may be used as CCTV application for making entries in workplaces.

REFERENCES

[1]. A. Rama and F. Tarres, P2CA: a new face recognition scheme combining 2D and 3D information, In Proc. of 16th IEEE *International Conference on Image Processing*, pp. 776–779, September 2005.

[2]. W. Li, C. Wang, D. Xu and S. Chen, Illumination invariant face recognition based on neural network ensemble, In Proc. of 24th IEEE *International Conference on Tools with Artificial Intelligence*, pp. 486–490, November 2004.

[3]. W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, Face recognition: a literature survey, *Journal of ACM Computing Surveys*, Vol. 35, No. 4, pp. 399–458, December 2003.

[4]. K. M. Prasanna, N. Hegde, A fast recognition method for pose and illumination variant faces on video sequences, *IOSR Journal of Computer Engineering*, e-ISSN: 2278-0661, p- ISSN: 2278-8727, Vol. 10, Issue 1, pp. 8–18, March - April 2013.

[5]. N. Jindal,V. Kumar, Enhanced face recognition algorithm using PCA with artificial neural networks, *International Journal of Advanced Research in Computer*

Science and Software Engineering, Vol. 3, Issue 6, pp. 864–872, June 2013.
[6]. X. Chai, L. Qing, S. Shan, X. Chen and W. Gao, Pose invariant face recognition under arbitrary illumination based on 3D face reconstruction, In Proc. of AVBPA 2005, LNCS 3546, pp. 956–965, 2005.
[7]. A. K. Roy-Chowdhury and Y. Xu, Pose and illumination invariant face recognition using video sequences, *Face Biometrics for Personal Identification*, part I, pp. 9–25, 2007.
[8]. UPC Face Database. http://gps-tsc.upc.es/GTAV
[9]. K. R. Singh, M. A. Zaveri and M. M. Raghuwanshi, Illumination and pose invariant face recognition: a technical review, *International Journal of Computer Information Systems and Industrial Management Applications*, ISSN: 2150-7988, vol. 2, pp. 29–38, 2010.
[10]. P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3D face model for pose and illumination invariant face recognition, In Proc. of 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 296–301, September 2009.
[11]. R. Rajalakshmi, M. K. Jeyakumar, A review on classifiers used in face recognition methods under pose and illumination variation, *International Journal of Computer Science Issues (IJCSI)*, Vol. 9, Issue 6, No. 2, pp. 474–485, November 2012.
[12]. F. Kahraman, B. Kurt and M. Gokmen, Robust face alignment for illumination and pose invariant face recognition, In Proc. of *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–7, June 2007.
[13]. S. Biswas, G. Aggarwal, and P. J. Flynn, Pose-robust Recognition of low-resolution face images, In Proc. of *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 601–608, June 2011.
[14]. R. Gross, S. Baker, I. Matthews and T. Kanade, Face Recognition Across Pose and Illumination, *IN HAND BOOK OF FACE RECOGNITION*, 2004.
[15]. B. Yegnanarayana, Artificial neural networks for pattern recognition, Sadhana, vol. 19, Part 2, pp. 189–238, April 1994.