# Emotion Recognition from Facial Expressions using Images with Pose, Illumination and Age Variation for Human-Computer/Robot Interaction

**Suja Palaniswamy**[1,*] **& Shikha Tripathi**[2]

[1]Department of Computer Science and Engineering, Amrita School of Engineering,
Amrita VishwaVidyapeetham, Kasavanahalli, Carmelram Post, Bengaluru - 560035,
Karnataka, India
[2]Faculty of Engineering, PES University, Bangalore South Campus, Hosur Road,
Electronic City, Bengaluru - 560100, Karnataka, India
*E-mail: p_suja@blr.amrita.edu

**Abstract.** A technique for emotion recognition from facial expressions in images with simultaneous pose, illumination and age variation in real time is proposed in this paper. The basic emotions considered are anger, disgust, happy, surprise, and neutral. Feature vectors that were formed from images from the CMU-MultiPIE database for pose and illumination were used for training the classifier. For real-time implementation, Raspberry Pi II was used, which can be placed on a robot to recognize emotions in interactive real-time applications. The proposed method includes face detection using Viola Jones Haar cascade, Active Shape Model (ASM) for feature extraction, and AdaBoost for classification in real-time. Performance of the proposed method was validated in real time by testing with subjects from different age groups expressing basic emotions with varying pose and illumination. 96% recognition accuracy at an average time of 120 ms was obtained. The results are encouraging, as the proposed method gives better accuracy with higher speed compared to existing methods from the literature. The major contribution and strength of the proposed method lie in marking suitable feature points on the face, its speed and invariance to pose, illumination and age in real time.

**Keywords:** *classification; emotion; facial expression; feature extraction; illumination; pose; Raspberry Pi II.*

## 1 Introduction

Emotion recognition is an emerging field of research to provide support in applications like patient monitoring, driver fatigue detection, human robot interaction for children with autism [1], situation analysis in social interaction, affective computing, feedback during e-learning [2], psychology, contests [3], and entertainment industries. Emotion recognition can be employed in assistive technologies to provide assistance in medical or social contexts. Implementation of emotion recognition in robots in such environments will allow them to react to changes/moods of the assisted person, analogous to communication by

Received December 7th, 2016, Revised June 23rd, 2017, Accepted for publication January 16th, 2018.

humans [4]. In recent years, researchers have analyzed the origin of emotions in the human brain and stimulated the same in robots as robots are expected to operate in an environment that is shared with humans and to interact naturally with humans with a level of emotional perception.

Illumination and pose variation are two of the major challenges in emotion recognition. Developing a suitable method for recognizing emotions in people of different age groups is also a challenging task, as features of the face vary with age. Very limited work that considers these constraints has been done by researchers in this area. In this paper, a method is proposed for recognizing emotions from facial expressions using images under variation of head pose, background illumination and age group. This work is an extension of Suchitra, *et al.* work [5], where emotion was recognized from front pose only. The proposed method works in real time and addresses the challenges in emotion recognition under the aforesaid constraints. Feature extraction can be carried out by applying a suitable pre-processing method and devising a suitable technique for feature vector formation. Choosing a suitable classifier for classifying expressions into basic emotions is the next step.

A geometrical feature based approach with two stages of implementation was used for feature extraction in this work. In stage 1, a geometrical based method for emotion recognition from images with pose and illumination variation from the CMU-MultiPIE database was applied. Additional testing with encouraging accuracy was done using the in-house developed 'Amrita Emotion' database. Based on the obtained results, a suitable approach for real-time implementation of emotion recognition with simultaneous pose, illumination and age variation was implemented in stage 2. The results were compared with existing literature and the proposed method had better accuracy.

The major contributions of this work can be summarized as follows:

1. A method for head-pose invariant facial emotion recognition that can recognize emotions with head-pose in the range -45° to +45° pan rotation is proposed. It performs well for continuous head-pose variation in the range -45° to +45° even though training was conducted for discrete poses. The proposed method recognizes emotions under varying illumination conditions for subjects of all age groups as well.
2. The optimum number of feature points that represent the face geometry was identified after analyzing and experimenting with several feature point configurations. The selection of feature points contributes significantly to improved accuracy.
3. An in-house 'Amrita Emotion' database was developed, which includes subjects of children, adult and elderly expressing six basic emotions with

simultaneous illumination and continuous head-pose variation. The database is of Indian origin and the first of its kind with all mentioned variations.

The following are the strengths of the proposed method for emotion recognition:

1. It is implemented on Raspberry Pi II, which is low-cost equipment that can be deployed for suitable real-time applications.
2. It takes an average of 120 ms for recognizing emotions on Raspberry Pi II (ARM1176JZF, 900MHz). It also performs well for continuous change of emotions at an average rate of 0.2 ms and works satisfactorily even if the subject is speaking while expressing emotions.

The remainder of this paper is organized as follows: Section 2 outlines related work and Section 3 describes the proposed method in detail. Section 4 discusses the database used. Section 5 gives insight into results and analysis. The final section includes the conclusion and discusses future work.

## 2 Related Work

Limited work has been done on emotion recognition with respect to varying pose, illumination and age. The two most commonly used approaches for extracting features are appearance and geometry feature based methods. Using LBP, LDP, DT-CWT and Gabor wavelet transform techniques for feature extraction, experiments have been performed on both the Cohn-Kanade and the JAFFE database [6]. Encouraging accuracy was obtained with transform domain techniques, where neural networks performed better than K-NN classifiers. Illumination has a major role in recognizing emotions. A novel approach for preprocessing was proposed for images from the CMU-MultiPIE database with varying illumination by using contrast stretching, anisotropic smoothing and ratio based methods in Suja, *et al.*[7].

In 2015, Happy proposed a method by extracting salient facial patches followed by LBP for feature extraction and determined the optimum number of salient patches [8]. For images with pose variation, geometric feature based methods are used. Rodovic proposed a pose invariant emotion recognition method using coupled scaled Gaussian process regression [9]. Benta presented a new spontaneous facial emotion database and organized real-time facial emotion recognition solutions grouped into spontaneous and posed facial emotion databases [10]. Recognition of emotions with variation of head pose, illumination, age, hair occlusions, etc. is challenging. Most of the existing work is based on frontal-view images of faces without facial hair, glasses and young people without permanent wrinkles, which is unrealistic. Differences can be seen in facial features and expressions among people of different cultures and age groups. Emotion recognition systems must be robust against such variations.

A method for emotion recognition from facial expressions using images under pose, illumination and age variation in real time is proposed in this paper. The results obtained are encouraging and can be implemented in interactive human-computer/ robot applications. The next section describes the proposed method.

## 3 Proposed Method

An overview of the proposed method is shown in Figure 1. It involves two stages. The subsections in this section give a detailed insight into the proposed method.

### 3.1 Stage 1

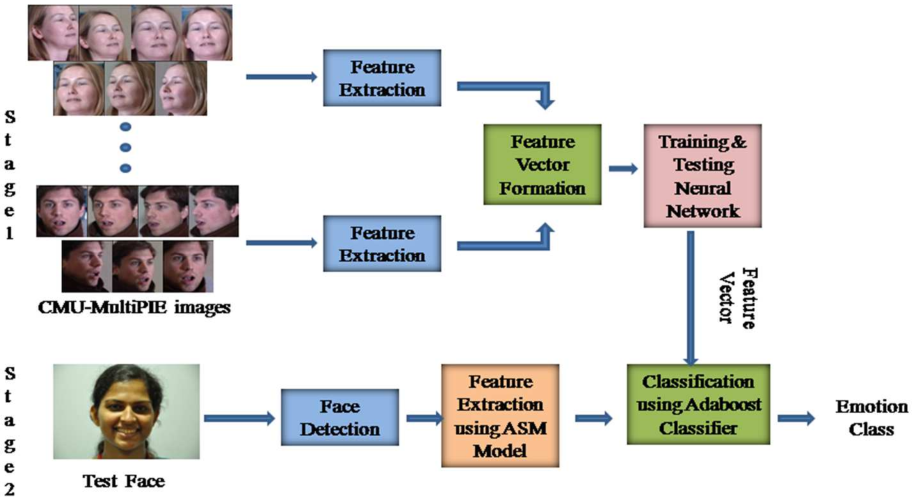


**Figure 1** Overview of the proposed method.

The algorithm for stage 1 works as follows:

1. Images with simultaneous pose and illumination variation from the CMU-MultiPIE database are given as input and preprocessed by cropping the face.

2. Features are extracted by marking 39 suitable feature points on the face. The selection of feature points is crucial as it determines the efficiency of the process.
3. Normalization is performed on the extracted feature points to remove effects of scaling and translation.
4. Feature vectors comprising a vast range of simultaneous poses and illumination variations are formed by combining feature points.
5. The feature vectors thus formed are fed into a neural network classifier for training and testing offline using the CMU-MultiPIE database and the in-house developed 'Amrita Emotion' database, respectively.

The algorithm is repeated by varying the number of feature points at 20, 24, 30, and 36. In stage 1, images with simultaneous pose and illumination variation were considered and a method for feature vector formation was developed. This feature vector was robust when tested with samples at arbitrary pose and varying illumination using 'Amrita Emotion' database. The algorithm is further explained in the following subsections.

### 3.1.1 Preprocessing

The size of the images in the CMU-MultiPIE database is 640 × 480. Cropping the face from the image is important, as features of the face convey essential information for emotion recognition. Each facial image was cropped to a size of 200 × 260. This process was repeated for all selected 10500 images (60 subjects x 5 emotions x 7 poses x 5 illuminations). The cropped images were then used for feature extraction.

### 3.1.2 Feature Extraction

Facial features were extracted using a geometric feature based approach by marking key feature points on the cropped facial images. These features can be extracted manually or automatically using Active Appearance Model tools. In this work, feature extraction was initially performed manually by using the am_tool developed by Cootes [11]. This tool enables the user to mark feature points on a face. In this work, initially 39 feature points were marked on eyebrows, eyes, mouth and nose, where emotions are found to be more prominent. The selection of 39 feature points was adapted from [9]. The movement of these points conveys information essential for recognizing emotions. Sample images displaying feature extraction using the am_tool are shown in Figure 2. Image 2a represents 39 feature points. Since the 39 points are marked and each point has 'x' and 'y' co-ordinates, the feature vectors extracted have 78 dimensions. This process was repeated for all images under consideration. These feature points have translation and scale components that need to be removed, which is described in the next subsection.

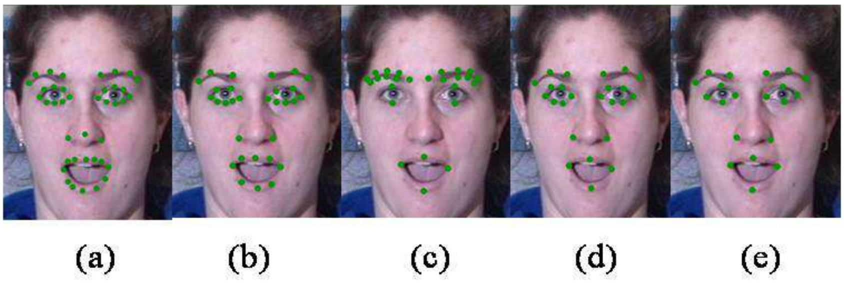


**Figure 2** Feature extraction by marking 39, 30, 26, 24 and 20 characteristic feature points on the face.

### 3.1.3 Normalization

The images are categorized into 5 different illumination levels and every emotion is expressed in different poses under each illumination level by each subject. For each illumination level, head pose space is divided into $M$ = 7 discrete poses ranging from -45° to +45° yaw rotation with an increment of 15°, as shown in Figure 1. The locations of $n$ facial points, marked on the facial images in pose $p$, where $p = 0,...,M\text{-}1$, were arranged in a vector, $\mathrm{n}^{\mathrm{p}}$ ϵ $\mathrm{R}^{2\mathrm{d}}$. The sample dataset is denoted by $D = \{D^0, \ldots D^p, \ldots D^{M-1}\}$, where $D^p = \{a_1^p, \ldots a_N^p\}$, and consisted of $N$ samples with $D^0$ sample data of frontal poses for each illumination level. The facial points were arranged into feature vectors as follows:

$$\mathrm{a} = [\mathrm{a}_1^{\mathrm{x}}, \ldots, \mathrm{a}_{\mathrm{L}}^{\mathrm{x}}, \mathrm{a}_1^{\mathrm{y}}, \ldots, \mathrm{a}_{\mathrm{L}}^{\mathrm{y}}] \qquad (1)$$

To remove effects of scaling and translation, the feature vectors were normalized. For this, the feature vectors were transformed into image coordinates $(\mathrm{a}_{\mathrm{i}}^{\mathrm{x}}, \mathrm{a}_{\mathrm{i}}^{\mathrm{y}})$, $i$=1, 2,..,$L$, as in Eq. (6), using Eqs. (1) to (5), where $L = 39$ represents the number of key feature points. The gravity center $(a_c^x, a_c^y)$ and scaling parameter $sc$ were computed as given by Rudovic [12]:

$$a_c^x = \frac{1}{L}\sum_{i=1}^{L} a_i^x \text{ , } a_c^y = \frac{1}{L}\sum_{i=1}^{L} a_i^y \qquad (2)$$

$$sc = \frac{1}{L}\sum_{i=1}^{L}((a_i^x - a_c^x)^2 + \left(a_i^y - a_c^y\right)^2)^{\frac{1}{2}} \qquad (3)$$

where the normalized facial landmark positions were:

$$a_i^{x_n} = \frac{((a_i^x - a_c^x)^2 + (a_i^y - a_c^y)^2)^{\frac{1}{2}}}{sc}(a_i^x - a_c^x) \qquad (4)$$

$$a_i^{y_n} = \frac{((a_i^x - a_c^x)^2 + (a_i^y - a_c^y)^2)^{\frac{1}{2}}}{sc}\left(a_i^y - a_c^y\right) \qquad (5)$$

The normalized feature vector for one illumination level was arranged as:

$$a^n = [a_1^{x_n}, \ldots, a_L^{x_n}, a_1^{y_n}, \ldots, a_L^{y_n}] \tag{6}$$

In a similar manner normalization was performed on the images with all 5 levels of illumination. The feature vectors for the sample images under consideration were concatenated to form a single feature vector, which is explained in the next subsection.

### 3.1.4 Emotion Recognition

For classification, a neural network classifier was used as this can develop and train networks with a large volume of data accurately. This classifier was trained with feature vectors of 5 emotions for 7 poses and 5 illuminations of 60 subjects. The network had neurons of size 78, 10 and 5 for input, hidden and output layers respectively. The size of the input feature vector was 78 and the number of output classes was 5. The number of neurons for the hidden layer was chosen to be 10 after repeated experimenting with different sets of neurons. As images were ordered first according to illumination and then pose and the normalized feature vectors of all images under consideration were stacked one below the other according to the level of illumination, pose, subject, and emotion. Feature vector samples of all 35 categories were combined into a single spreadsheet and supervised training was performed for learning the patterns to recognize emotions. A single classifier was developed that classifies the emotion of the test samples for each particular pose and illumination combination. A confusion matrix generated by the Neural Pattern Recognition tool displays the classification results and is discussed in Section 5.

In this approach, 80% of the feature vectors for the 10500 samples from the CMU-MultiPIE database were taken for training. This input data set included a wide range of data that covered all possible feature points for 5 emotions with 5 levels of illumination, and head pose ranging from +45º to -45º. The tool developed a network by training using a scaled conjugate gradient algorithm in several iterations. This network was tested with the remaining feature vectors from the CMU-MultiPIE database and the in-house ‘Amrita Emotion’ database. Sample images from the ‘Amrita Emotion’ database are shown in Section 4. These images were taken with arbitrary poses and varying illumination. Neural network classifies emotion by giving emotion label (*l*) as output. The geometric feature based method considers the location of feature points and the background illumination will not affect the recognition of emotions. The novelty of proposed method lies in the formation of a feature vector that covers a wide range of sample images with simultaneous pose and illumination variation. This feature vector was used for training a real-time classifier, which recognizes emotions in persons of different ages with pose and illumination variation in real time. The real-time implementation is described in Section 3.2. (Stage 2).

### 3.1.5 Optimum Number of Feature Points

To find whether the accuracy varies with the number of points, the proposed method was repeated by varying the number of feature points at 30, 26, 24, and 20. The locations of the feature points are shown in Figure 5, represented by sub-images (b) to (e). It was found that the average accuracy for 26 and 39 feature points was 99% in training and the maximum average accuracy obtained in testing for 26 feature points using the CMU_MultiPIE and the 'Amrita Emotion' database was 95% and 92%, respectively. This indicates that appropriate selection of feature points on the face is very important for accurate classification. A comparison of the accuracy obtained for all feature points under consideration is discussed in Section 5. Additional testing results obtained using the 'Amrita Emotion' database with arbitrary pose and varying illumination encouraged and motivated implementation in real time. Hence, a method for real-time implementation is proposed the following section.

## 3.2 Stage 2

The procedure for performing emotion recognition in real time is discussed in this section.

**Step 1:** An image is captured in real time using a webcam
**Step 2:** The face is detected from the image using Viola-Jones face detection [13], which uses the concept of the Haar wavelet to develop a basic image to help detect the face. The Viola Jones face detection algorithm works well for non-frontal poses from +45° to -45°. The detected facial image is then used for performing further processing.
**Step 3:** After detecting the face, preprocessing is carried out, where the image is segmented into specified coordinates and normalization is done as discussed under stage 1. The cropped image is given to a Sobel filter for enhancing the edges of the face and to suppress the non-face regions.
**Step 4:** Extracting features is performed using the Active Shape Model (ASM) library.

Extracting features is performed using ASM, based on a geometric feature based approach [14]. The ASM automatic feature point location method is initially applied to the preprocessed image, followed by determining the Euclidean distance between the center of gravity (CoG) coordinate and annotated feature points on the face in the image. To extract geometric features, the difference between a person's basic emotional expression and a neutral expression is determined. In ASM, the shape of the input face is altered to iteratively obtain the shape of the model, as shown in Figure 3. From the 116 feature points that are marked automatically in the shape model, 26 feature points are extracted from the input facial image. Since 39 and 26 feature points give the same training accuracy (as discussed under stage 1) and as time is an important factor in real time emotion recognition, 26 feature points are extracted in real time implementation, as shown in Figure 4. Euclidean distances are determined for the points, as shown in Figure 4, using Eq. (7). These distances are combined to form a feature vector that is fed into AdaBoost for emotion classification. For training the model, the feature vector formed from the CMU-MultiPIE database discussed above was used. Euclidean distance is applied to calculate the distance between two feature points with P and Q coordinates for a 2D image, which is expressed as:

$$D= \sqrt{(P_x - P_y)^2 + (\, Q_x - Q_y)^2} \tag{7}$$

where x, y = 1, 2, 3… 26 and P, Q are the coordinates of the extracted feature points in the 2D image.

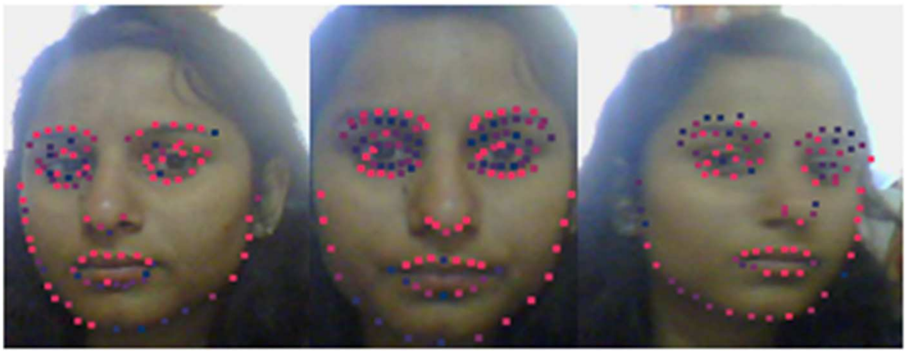


**Figure 3** Shape model.

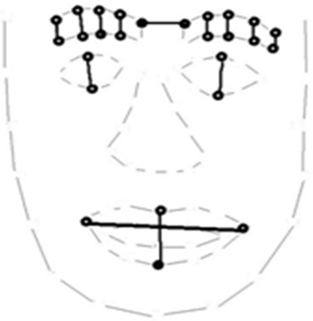


**Figure 4** Feature points extracted in real time from the shape model.

**Step 5:** Classification is carried out using the AdaBoost classifier [15]. Choosing a classifier depends on the type of data, the number of samples in the training dataset, the number of features, feature selection dimensionality reduction, the desired accuracy and the structure of the problem. Cascade classifiers have good computational and classification performance in real time [13]. The cascade classifier comprises several stages, where each stage is an ensemble of weak learners. Each stage is trained using a technique called boosting. Boosting combines several mildly powerful predictors, called weak classifiers, to form a highly accurate combined classifier. The classification error is reduced exponentially by boosting and for this reason AdaBoost is more widely used for real-time applications compared to SVM and Bayes classifiers. In our proposed method for emotion recognition in real time, we chose the best training samples as identified by NN and gave them to AdaBoost. In every iteration, AdaBoost increases the weight of the samples that were wrongly classified. The classifier takes only three iterations to achieve maximum accuracy.

The classifier was trained with the 50 training samples formed in stage 1. The input, a training set and the target $y_j$, with each element corresponding to 5 emotions, are given as $\left\{\left(x_i, y_j\right), \dots, \left(x_m, y_j\right)\right\}, y_j = \{1,2,\dots,5\}, i = 1,2,\dots,50$, respectively. AdaBoost creates a cost matrix $\boldsymbol{C}_t \in \mathbb{R}^{50 \times 5}$, specifying to the weak learner that the cost of classification of a sample $x_i$, denoted as $l$, is $\boldsymbol{C}_t(i,l)$.The cost matrix is not arbitrary but conforms to the restrictions given in Eq. (8).

The weak learner returns some weak classifiers $h_t: X \rightarrow \{1,2,\dots,k\}$ from a fixed space $h_t \in \mathcal{H}$ and the cost incurred is:

$$\boldsymbol{C}_t \bullet \mathbf{1}h_t = \sum_{i=1}^{m} \boldsymbol{C}_t\left(i, h_t(x_i)\right) \tag{8}$$

Here, $\mathbf{1}_h$ is the 50 × 5 matrix whose *(i,j)*th entry is 1 [*h(i)=j*]. AdaBoost computes weight $\alpha_t$ for the current weak classifier based on how much cost was incurred in that iteration. Finally, AdaBoost predicts the weight according to the weighted number of votes of the classifiers returned in each iteration (see Eq. (9)).

$$H(x) \triangleq argmax\, f_T(x,l), where\, f_T(x,l) \triangleq \sum_{t=1}^{T} 1[h_t(x) = l]\alpha_t \tag{9}$$

By carefully choosing the cost matrixes at each iteration, AdaBoost aims to minimize the training error of final classifier *H*. Here, the cost matrix is given by following Eq. (10):

$$\boldsymbol{C}(i,l) = \begin{cases} d(i,l) if l \neq y_i \\ -d(i,l) if l = y_i \end{cases} \tag{10}$$

where the initial value of *d* considered is 1/50.

The training feature vector makes the classifier suitable for recognizing emotions at any arbitrary pose and illumination. The feature vector formed in real time is compared with the training-stage feature vector and the classifier identifies the emotion class to which the distance is closest. ASM is used for feature extraction and fits the data consistent with the training set. ASM makes a best fit of the input data model to the training set in real time. It allows only minimum variability but is specific to the class of emotion it represents. After training the AdaBoost classifier, the feature vector formed from real-time facial images is used for testing.

**Step 6:** Real-time implementation using Raspberry Pi II
Raspberry Pi II is a small computer embedded on a chip (Broadcom BCM2835, ARM1176JZFS) with a video core 4 GPU and floating point running at 900 MHz that works in a Linux environment. The program developed for real-time emotion recognition was installed onto Raspberry Pi II. Input and output devices like keyboard and monitor were linked to Raspberry Pi II for input and display, as shown in Figure 5. Putty software and a virtual network connection (VNC) were used to use a laptop as a remote desktop for display and a keyboard for input. The method proposed under stage 2 performs operations on Raspberry Pi and classifies emotions expressed as output on the monitor.

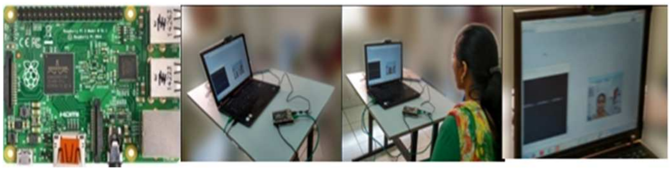


**Figure 5** Raspberry Pi connected with camera and display and subject expressing emotion.

The proposed method was tested with subjects of different age groups with varying pose and illumination in real time. The results were encouraging. Real-time emotion recognition with simultaneous pose, illumination and age variation is highly challenging and very limited work has been done in this area of research. The novel combination of ASM and AdaBoost enhances the performance. The proposed method recognizes emotions under all variations and also for -90º to +90º with good accuracy, even though the training data had pose variant images with variation from -45º to +45º. The proposed method recognizes emotions at an average time of 120 ms and is suitable for applications where time is an important criterion. A subject changing emotions continuously is also recognized within 0.2 ms on average. Even though the proposed method is applicable for images, it works well even if the subject speaks while expressing emotions. The results and analysis are discussed in Section 5.

## 4 Experimental Setup

The CMU-MultiPIE and the in-house developed 'Amrita Emotion' database were used in this work. This section gives a description of both.

### 4.1 CMU-MultiPIE database

The CMU-MultiPIE facial expression database [16], developed by Carnegie Mellon University, is a collection of 2D images for which 337 subjects expressed emotions in four different sessions. Out of 337 subjects, 129 attended all 4 sessions. To systematically capture images with varying poses and illuminations, a system of 15 cameras and 18 flashes was used. Thirteen cameras were located at head height, spaced at 15° intervals, and two additional cameras were located above the subject, simulating a typical surveillance view. During one recording session, 20 images were captured by each camera: 1 image without flash illumination, 18 images with each flash firing individually, and then one more image without flash. For the present work, 60 subjects, of whom 40 males and 20 females, were chosen after excluding subjects with glasses and hair occlusions.

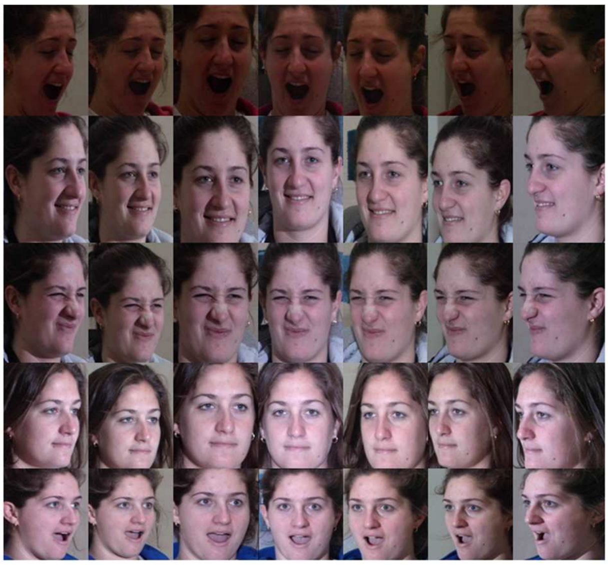


**Figure 6** Sample images from the CMU-MultiPIE database with simultaneous pose and illumination variation.

All chosen subjects expressed 5 basic emotions, i.e. neutral, happiness, surprise, disgust and anger, at 5 levels of illumination and 7 poses varying from -45º to +45º (0º, ±15º, ±30º, ±45º). The total number of images from the database used in this work was 10500. One subject from the CMU-MultiPIE database expressing emotions anger, happy, disgust, neutral and surprise under simultaneous illumination and pose variation is shown in Figure 6, where level 0 illumination is at the top and level 8 is at the bottom, the intermediate levels are 5, 6 & 7, and the head pose is in the range -45º to +45º from left to right.

## 4.2 Amrita Emotion Database

Apart from the CMU-MultiPIE database, an in-house developed database, named the 'Amrita Emotion' database, was also used for testing. It includes 35 subjects with emotions anger, disgust, happiness, neutral and surprise. Figure 7 shows subjects of different age groups expressing basic emotions and a subject expressing happiness in an arbitrary pose is shown in Figure 8. The 'Amrita Emotion' database has both male and female subjects of varying age groups (1-18, 19-39, 40-59, 60 and above). The images were recorded at ±45º yaw rotation at frontal view and arbitrary pose and illumination. This database was mainly used for testing the head pose and illumination invariance of the algorithm proposed in stage 1, which proved to be database-independent and head-pose and illumination invariant. The 'Amrita Emotion' database is introduced in this paper.

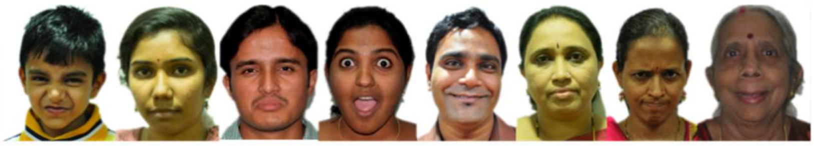


**Figure 7** Sample images from the in-house developed 'Amrita Emotion' database showing different emotions.

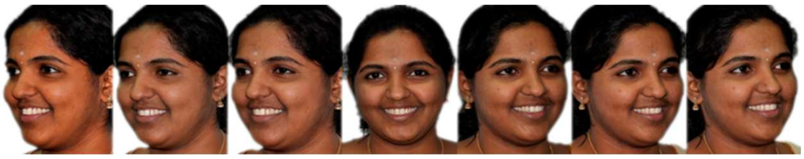
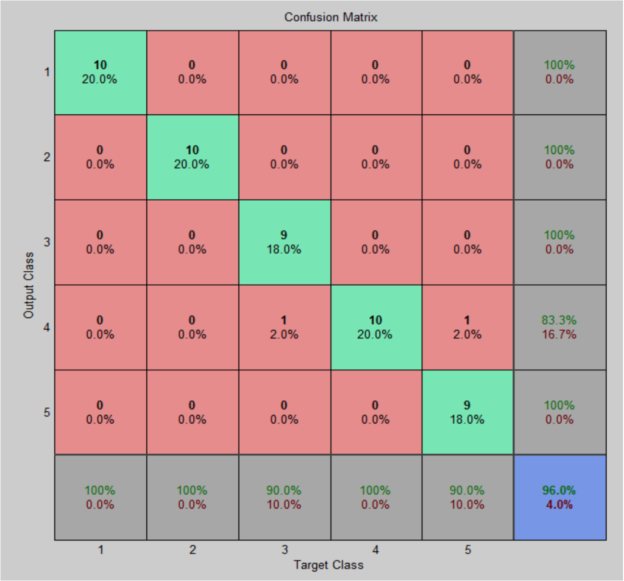


**Figure 8** Images of a subject showing happiness in an arbitrary ±45º head pose space from the in-house developed 'Amrita Emotion' database.

## 5 Results and Analysis

Testing in stage 1 was performed for 35 individual combinations of pose (7 poses) and illumination (5 levels) variations and the test samples were categorized accordingly. Classification of emotions was performed by taking test samples from each category to test the network and the NPR tool generated a confusion matrix, as shown in Figure 9 (shown for 0 illumination level and +45°). Labels 1, 2, 3, 4 and 5 in Figure 9 represent the emotions anger, disgust, happy, neutral and surprise respectively. The average accuracy of 96% at the bottom right of the matrix is the overall accuracy for all emotions and test samples. Likewise, the overall accuracy of all 35 combinations of pose and illumination was consolidated as reported in Table 1. The average accuracy obtained for pose and illumination variations for 26 feature points is given in Table 1. The average accuracy presented in Table 1 for illumination level shows that illumination variation did not affect the recognition rate as it was similar for all levels. Overall average accuracy for all illumination levels was 95%. The accuracy for one illumination level for all poses varied at ± 2% with respect to 95%, which is acceptable. This infers that the proposed method using the geometric feature based approach is illumination invariant and suitable for real-time application in illumination variant environments. For pose variation, emotions expressed at 0° and +45° pose were recognized uniformly. This indicates that the feature points marked on the frontal face also hold good for +45° head poses and avoids mapping of feature points from non-frontal to frontal pose. The average accuracy was the same for the majority of non-frontal poses. The overall average accuracy for all poses was 95%. The average accuracy for every pose for all illumination levels varied at ± 3% with respect to 95%, indicating that the proposed method is pose invariant.


Confusion Matrix
Output Class
Target Class

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| 1 | 10<br>20.0% | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% | 100%<br>0.0% |
| 2 | 0<br>0.0% | 10<br>20.0% | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% | 100%<br>0.0% |
| 3 | 0<br>0.0% | 0<br>0.0% | 9<br>18.0% | 0<br>0.0% | 0<br>0.0% | 100%<br>0.0% |
| 4 | 0<br>0.0% | 0<br>0.0% | 1<br>2.0% | 10<br>20.0% | 1<br>2.0% | 83.3%<br>16.7% |
| 5 | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% | 0<br>0.0% | 9<br>18.0% | 100%<br>0.0% |
| | 100%<br>0.0% | 100%<br>0.0% | 90.0%<br>10.0% | 100%<br>0.0% | 90.0%<br>10.0% | 96.0%<br>4.0% |

**Figure 9** Confusion matrix for illumination level 0 and +45° pose.

**Table 1** Average accuracy at stage 1 for all emotions with pose and illumination variation using CMU-MultiPIE database.

| | -45° | -30° | -15° | 0° | +15° | +30° | +45° | Avg |
|---|---|---|---|---|---|---|---|---|
| **0** | 95 | 95 | 94 | 95 | 98 | 98 | 96 | 96 |
| **5** | 97 | 97 | 94 | 96 | 97 | 95 | 96 | 96 |
| **6** | 94 | 97 | 95 | 95 | 95 | 96 | 96 | 95 |
| **7** | 96 | 95 | 95 | 96 | 97 | 95 | 94 | 95 |
| **8** | 98 | 96 | 94 | 94 | 94 | 94 | 95 | 95 |
| **Avg** | 96 | 96 | 94 | 95 | 96 | 96 | 95 | |

For stage 1 of the proposed method, the experiment was repeated by varying the feature points. It was found that 26 feature points gave the maximum training and testing accuracy, as shown in Figure 10.

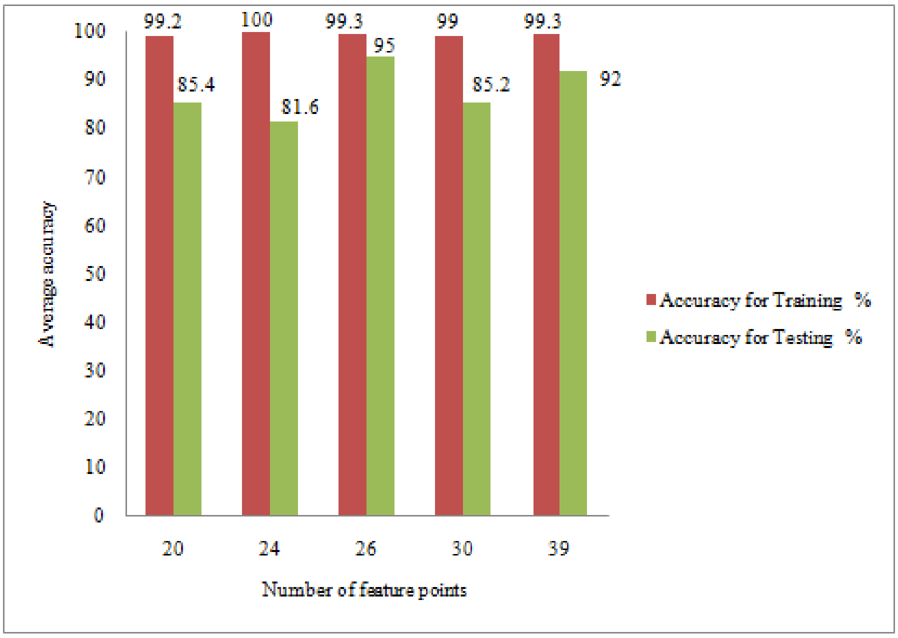


**Figure 10** Graph showing comparison of accuracy for 20, 24, 26, 30 and 39 feature points.

This indicates that the selection of appropriate feature points is very important for emotion recognition. The performance of the proposed method was tested on the CMU-MultiPIE and the 'Amrita Emotion' database. Figure 11 shows the emotion-wise accuracy obtained during testing for both databases. The testing result demonstrated that the proposed method performed well on the test samples of continuous poses even though the network was trained for discrete poses.

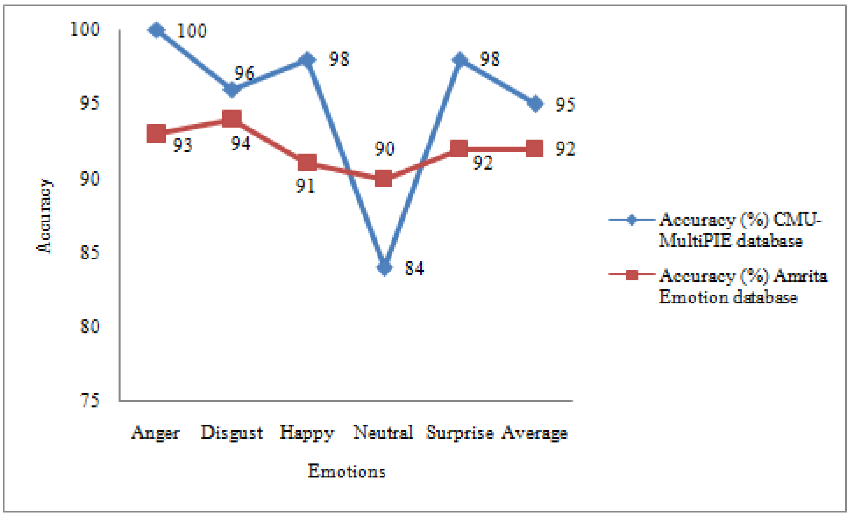


**Figure 11** Emotion-wise average accuracy in stage 1 for all pose and illumination variations for both databases.

Owing to the accuracy obtained offline, the feature vector formed in stage 1 was used for training the AdaBoost classifier in real time. The real-time module installed on Raspberry Pi II processed the input images and identified emotions as described above. The method proposed in Section 3.2 (Stage 2) was tested in real time with subjects of different ages expressing emotions with different illuminations and continuous head pose in the range -45° to +45°. Geometric feature based feature extraction is invariant of illumination and skin wrinkles as the coordinates of feature points are used to form the feature vector instead of texture characteristics of the facial features. For pose variation in real time, the Viola Jones face detection method detected faces in the range -45° to +45°. The ASM automatic feature detection algorithm performed well for feature extraction in real time. The feature extraction and face shape model performed well for any input face irrespective of size. As the geometric feature based feature extraction method does not include texture characteristics, the proposed method works well for varying age groups, poses and illuminations. Snapshots of subjects of varying age groups expressing emotions in real time and recognition displayed on a monitor are shown in Figure 12(a) to 12(d). The expressed emotion is displayed in words and percentages for other subtle emotions that are present in the expressed emotion are shown. The distance between the subject and the camera was approximately 0.5 m to 1.2 m. The average time taken for real time recognition was 120 ms. Continuously changing emotions were also recognized within 0.2 ms on average. Even though the proposed method is applicable for still images, it works well even if the subject speaks while expressing emotions.

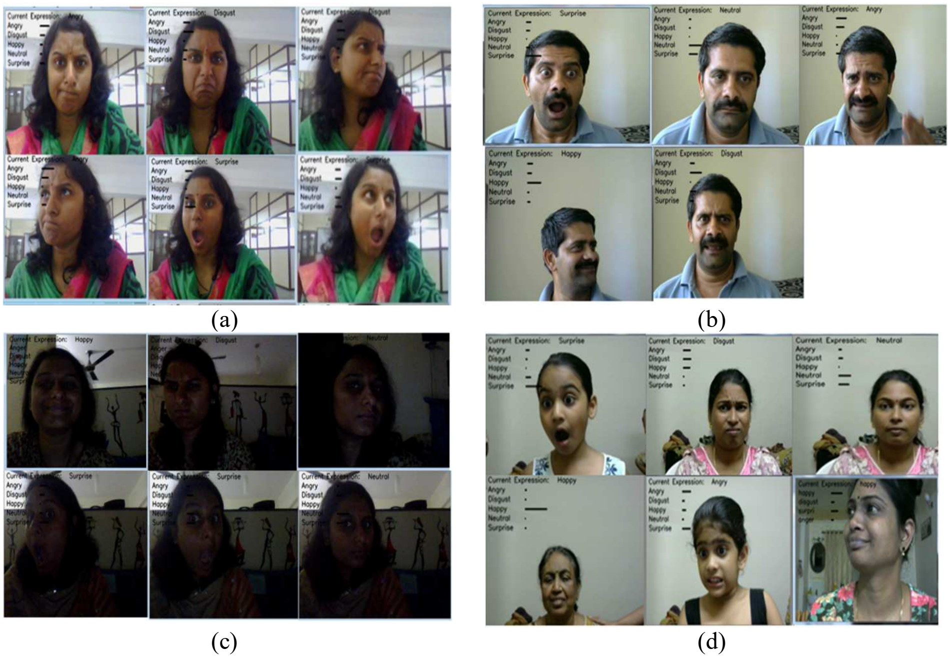


**Figure 12** (a) Snapshots of a subject expressing different emotions at varying poses with bright illumination. (b) Snapshots of a subject expressing different emotions at varying poses with medium illumination. (c) Snapshots of a subject expressing different emotions at varying poses with poor illumination. (d) Snapshots of subjects of different age groups expressing different emotions at varying poses with low illumination.

The confusion matrix in Table 2, with the average accuracy for all emotions tested in real time, was obtained by counting the true positives and misclassifications during the recognition of emotions for all subjects in real time. It represents the average accuracy for 30 subjects in real time. A comparison of these results with results from the existing literature is shown in Table 3.

**Table 2** Confusion matrix for real-time emotion recognition.

| | Anger | Disgust | Happy | Neutral | Surprise |
|---|---|---|---|---|---|
| **Anger** | 92.24 | 0.0 | 0.0 | 0.0 | 7.76 |
| **Disgust** | 3.7 | 96.3 | 0.0 | 0.0 | 0.0 |
| **Happy** | 0.0 | 0.0 | 97.36 | 1.32 | 1.32 |
| **Neutral** | 0.0 | 0.0 | 1.1 | 98.9 | 0.0 |
| **Surprise** | 1.54 | 0.0 | 3.02 | 0.0 | 95.44 |

During real-time testing, the emotions disgust, neutral, and happy were recognized with high accuracy. The accuracy of anger recognition in real time was lower compared to the results in Figure 11. One possible reason for this drop in accuracy could be the extent of expression of the emotion. In the CMU-MultiPIE database all subjects invariably close their eyes while expressing anger, whereas during real-time testing, the subjects did not close their eyes and expressed anger through their eyes. Overall, the real-time accuracy was close to the accuracy obtained offline. The number of subjects chosen for testing in real time was far smaller compared to the offline sample size. Increasing this would improve average accuracy in real time.

Table 3 shows a comparison of the results obtained by the proposed method with results from the existing literature. The proposed method obtained 96% accuracy in real time with 26 points, which is comparable to Abdat [17] using 38 points. The AdaBoost classifier used in the proposed method yielded better accuracy than SVM [18]. The geometric feature based approach implemented in the proposed method outperformed the appearance based method in [19]. In accordance with Barnard and Botha [20], a large sample size and training set with an equal number of samples for each group showed improved performance in this work. The average time taken for real-time recognition was 120 ms, which is highly suitable for real-time applications.

**Table 3** Comparison of results from proposed method with existing literature.

| Reference | Accuracy % | Number of points | Time taken | Variations considered |
|---|---|---|---|---|
| Abdat [17] | 95 | 38 | 31 ms<br>Intel Pentium 3.4GHz | Nil |
| Myunghoon [18] | 72 | 77 | 421.6 ms | Nil |
| Anderson [19] | 82 | Appearance based methods | 18 fps on Matrox Genesis DSP boards | Pose and illumination |
| **Proposed Method** | **96with Raspberry Pi II** | **26** | **120 ms** | **Pose, illumination & age** |

## 6 Conclusion

A method for real-time recognition of emotions under varying constraints, i.e. pose, illumination and age, was proposed in this paper. The major contribution of this work is in identifying the optimum number and location of suitable feature points to form a feature vector and to choose the right combination of feature selection and classifier to make the method efficient in a real-time scenario with all aforesaid variations. The proposed method uses the CMU-MultiPIE and the ‘Amrita Emotion’ database, which include images with pose and illumination variation. A geometric based approach is used for feature extraction and the feature vector includes a wide range of poses and illumination variations that are used for training the classifier in real time. A combination of ASM and AdaBoost enhances the performance. The proposed method is user independent, takes very little time for recognition and works under poor illumination conditions as well as for all ages with varying head pose. This makes the system robust and suitable for real-time applications.

The strength of the proposed method lies in the number and location of the feature points on the face, its speed and invariance to pose, illumination and age in real time. Continuously changing emotions are also recognized within 0.2 ms on average. Even though the proposed method is applicable for still images, it works well even if the subject speaks while expressing emotions. The results are encouraging and suggest that the Raspberry Pi II can be placed on a moving social robot to be deployed in senior citizen's homes or health care environments that can recognize emotions expressed by people under aforesaid variations and alert personnel present. Emotion recognition robots could also be used in the education field to identify the mood of students in listening and also to observe their behavior during studying. The level of complexity in teaching can be adjusted by the observations made by robots from students' facial expressions. The aspects of making this into a commercial product for assisting in special environments for children will also be looked into in the future.

## Acknowledgments

The authors wish to express their gratitude to Mr. Pradeep L.N., Ms. Suchitra and Dr. Karthik R.M. for their effort and contributions in this work.

## References

[1] Leo, M., Coco, M.D., Carcagni, P., Distante, C., Bernava, M., Pioggia, G. & Palestra, G., *Automatic Emotion Recognition in Robot-Children Interaction for ASD Treatment,* International Conference on Computer Vision Workshops, IEEE, pp. 145-153, 2015.

[2] Happy, S.L., Dasgupta, A., Priyadarshi, P. & Routray, A., *Automated Alertness and Emotion Detection for Empathic Feedback during E-Learning,* 5th International Conference on Technology for Education (T4E), IEEE, pp. 47-50, 2013.

[3] Ian J. Goodfellow, Erhan, D., Carrier, P.L., Courville A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Taylor, J.S., Milakov, M., Park, J., Ionescu, R., Popescu, M., Grozea, C., Bergstra, J., Xie, J., Romaszko, L., Xu, B., Chuang, Z. & Bengio Y., *Challenges in Representation Learning: a Report on Three Machine Learning Contests*, Workshop in Challenges in Representation Learning (ICML), Atlanta, United States, pp. 1-8, 2013.

[4] Leo, M., Medioni, G., Trivedi, M., Kanade, T. & Farinella, G.M., *Computer Vision for Assistive Technologies,* Computer Vision and Image Understanding, **154**, pp. 1-15, 2017.

[5] Suchitra, Suja, P. & Shikha, T., *Real Time Emotion Recognition from Facial Images using Raspberry Pi II*, 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, pp. 666-670, 2016.

[6] Suja, P. & Shikha, T., *Analysis of Emotion Recognition from Facial Expressions using Spatial and Transform Domain Methods,* International Journal of Advanced Intelligence Paradigms, **7**(1), pp. 57-73, 2015.

[7] Suja, P., Thomas, S.M., Shikha, T. & Madan, V.K., *Emotion Recognition from Images under Varying Illumination Conditions*, 6th International Workshop on Soft Computing Applications (SOFA), Springer, pp. 913-921, 2016.

[8] Happy, S.L. & Routray, A., *Automatic Facial Expression Recognition using Features of Salient Facial Patches,* IEEE Transactions on Affective Computing, **6**(1), pp. 1-12, 2015.

[9] Rudovic, O., Maja P. & Ioannis (Yiannis), P., *Coupled Gaussian Processes for Pose-invariant Facial Expression Recognition,* IEEE Transactions on Pattern Analysis and Machine Intelligence, **35**(6), pp. 1357-1369, 2013.

[10] Benta, K.I. & Vaida, M.F., *Towards Real-life Facial Expression Recognition Systems*, Advances in Electrical and Computer Engineering, **15**(2), pp. 93-102, 2015.

[11] Cootes, T.F, Edwards, G.J. & Taylor, C.J., *Active Appearance Models,* IEEE Transactions on Pattern Analysis and Machine Intelligence, **23**(6), pp. 681-685, 2001.

[12] Rudovic, O., P. Ioannis (Yiannis) & Maja, P., *Facial Expression Invariant Head Pose Normalization using Gaussian Process Regression,* International Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 28-33, 2010.

[13] Paul, V. & Micheal., J., *Robust Real-time Face Detection,* International Journal of Computer Vision, **57**(2), pp.137-154, 2004.

[14] Cootes, T.F., Taylor, C.J. & Graham, J., *ASM – Their Training and Application*, Computer Vision and Image Understanding, **61**(1), pp.38-59, 1995.

[15] Indraneel, M. & Schapire, R.E., *A Theory of Multiclass Boosting*, Journal of Machine Learning Research, **14**, pp. 437-497, 2013.

[16] Gross, R., Matthews, I., Cohn, J., Kanade, T. & Baker S., *Guide to the CMU Multi-PIE database,* The Robotics Institute, Carnegie Mellon University, Technical report, 2007.

[17] Abdat, F., Maaoui, C. & Pruski, C., *Human-computer Interaction using Emotion Recognition from Facial Expression*, 5thUKSim European

Symposium on Computer Modeling and Simulation. IEEE, pp. 196-201, 2011.

[18] Myunghoon, S. & Prabhakaran, B., *Real-time Mobile Facial Expression Recognition System – a Case Study,* Conference on Computer Vision and Patten Recognition Workshops (CVPR),IEEE, pp. 132-137, 2014.

[19] Anderson, K. & McOwan, P.W.,*A Real-time Automated System for Recognition of Human Facial Expressions,* IEEE Transactions on Systems, Man, and Cybernetics, Part B, **36**(1), pp. 96-105, 2006.

[20] Barnard, E. & Botha, E.C., *Back-propagation Uses Prior Information Efficiently*, IEEE Transactions on Neural Networks, **4**(5), pp. 794-802, 1993.