Facial Expression Based Music Player

Sushmita G. Kamble and **Asso. Prof. A. H. Kulkarni**
Department Of Computer Science and Engineering
KLSGIT
Belagavi, Karnataka, India
E-mail id: sushmita.gk90@gmail.com, ahkulkarni@git.edu

***Abstract*- Conventional method of playing music depending upon the mood of a person requires human interaction. Migrating to the computer vision technology will enable automation of such system. To achieve this goal, an algorithm is used to classify the human expressions and play a music track as according to the present emotion detected. It reduces the effort and time required in manually searching a song from the list based on the present state of mind of a person. The expressions of a person are detected by extracting the facial features using the PCA algorithm and Euclidean Distance classifier. An inbuilt camera is used to capture the facial expressions of a person which reduces the designing cost of the system as compared to other methods. The results show that the proposed system achieves upto 84.82% of accuracy level in recognizing the expressions.**

Keywords: Euclidean Distance classifier, Expression Recognition, Facial Feature Extraction, PCA

I. INTRODUCTION

Facial expressions are one of the natural means to communicate the emotions and these emotions can be used in entertainment and Human Machine Interface (HMI) fields [1] [2]. In today's world, with the advancements in the areas of technology various music players are deployed with features like reversing the media, fast forwarding it, streaming playback with multicast streams. Although these features satisfy the basic requirements of the user, yet one has to manually surf for the song from a large set of songs, according to the current circumstance and mood [3] [4]. This is a time consuming task that needs some effort and patience. The main objective of this work is to develop an intelligent system that can easily recognize the facial expression and accordingly play a music track based on that particular expression/emotion recognized. The seven universally classified emotions are Happy, Sad, Anger, Disgust, Fear, Surprise and Neutral. The algorithm that is used in developing the present system is Principal Component Analysis (PCA) which utilizes eigenfaces to extract the facial features. The designed algorithm is very efficient due to less computational time taken hereby increasing the performance of the system. This work finds its applications in various domains like Human Computer Interaction (HCI), therapeutic approach in health care etc.

II. LITERATURE SURVEY

Several methods have been proposed earlier to detect and recognize the facial features and audio features from an audio signal with certain algorithms. But there are very few systems that automatically generate a playlist based on the expressions detected which make use of some additional hardware like sensors or EEG. Some of which are discussed as below:

In [5], Brain Computer Interface (BCI) - based mobile multimedia controller is proposed. It makes use of an external EEG hardware to monitor the cognitive state of mind. These systems require the user's active mental command continuously to control the multimedia. Most of the BCI systems require huge and costly EEG machines and commercial software's which limit the feasibility of the system for daily applications. A MIDlet program that contains a cognitive detection algorithm is built in the mobiles that continuously monitor the EEG signals acquired from EEG machines, and then recognize the user's state of mind.

In [6], the emotions are recognized by our voice/speech. The major concern of this system is the environment in which it is set up because if it is set up in an open environment then the surrounding noise will have a severe effect on the system's performance. To reduce the noise a voice activity detection algorithm has to be integrated to advance the system bandwidth efficiency which detects the speech from the input signal. It also reduces the occurrence of speech mismatch that can lead to a significant degradation of the emotion recognition rate.

In [7], Electroencephalography (EEG) signals are used to detect the human emotions. EEG records the electrical activity of the brain within its neurons. EEG signals are mainly used due to the fact that it detects real emotions arising instantly from the mind ignoring all other outside characteristics like facial expressions or gestures. Two classifiers such as Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) are employed to categorize the EEG signals into seven emotions of an individual. EEG signals are recorded by placing some electrodes on the scalp which are further processed to extract some features like Energy and Power Spectral Density (PSD) and are then fed to the two classifiers for the emotion detection.

Correlation Method [8], is one of the simplest methods also known as the nearest neighbor method. It returns the similarity score as the angle between two images. In this method the training images and test images are converted into column vectors. Comparison of the test image and gallery image is made in a high dimensional space rather in a low dimensional space which requires more storage space and hence the recognition time also increases leading to the disadvantage of correlation method.

978-1-5090-2029-4/16/$31.00 @2016 IEEE

In Geometric and Appearance Based Methods [9] [10], it tracks the size and shape of the facial parts like eyes, eyebrows, lip corners, mouth, nose etc. To classify the expressions some shape models [11] [12], that are based on a set of characteristics points on the face are being used. However, the distance between the facial landmarks differ for different individuals which makes the system less reliable. Outward appearances include change in composition and hence filters like Gabor wavelets, Local Binary Pattern etc that are a portion of the appearance based methods are applied either to whole of the face or to a particular face region to encode the surface. The facial patches vary according to the training data in these approaches based on their positions and sizes. Hence, it becomes hard to consider a standard system using these approaches.

In *[1] [2]*, PCA algorithm is used to extract the facial features and classify the expressions of an individual and from *[3] [4]*, the idea of automatically generating the song based on the expression detected is been used and combined to get the desired and efficient and robust system.

III. System Design

A. System Architecture

The system architecture for the proposed system is given in fig 3.1 [13]. The input image is loaded into the system in .jpg format. Then each image undergoes pre-processing i.e. removal of unwanted information like background colour, illumination and resizing of the images. Then the required features are extracted from the image and stored as useful information. These features are later added to the classifier where the expression is recognised with the help of Euclidean distance. Minimum the value of the distance calculated, the nearest the match will be found. Finally, a music track will be played based on the emotion detected of the user.

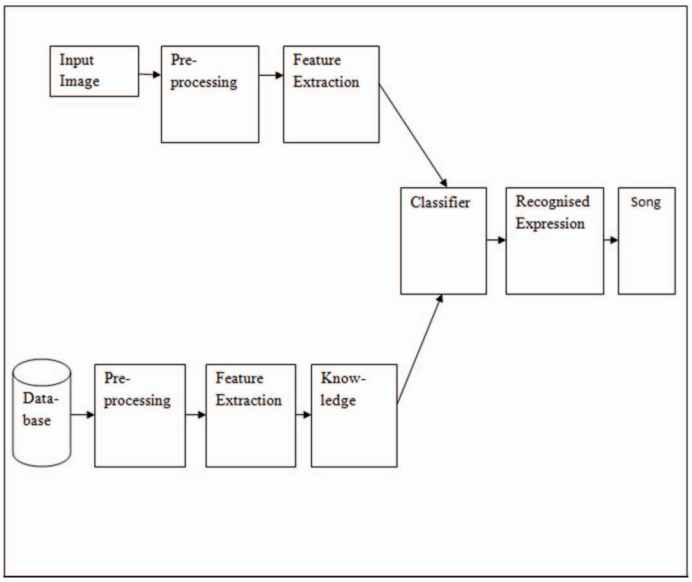

Fig. 3.1 . System Architecture

B. Steps involved to design the system

To design the system, training dataset and test images are considered for which the following procedures are applied to get the desired results. The training set is the raw data which has large amount of data stored in it and the test set is the input given for recognition purpose. The whole system is designed in 5 steps:

1. Image Acquisition

In any of the image processing techniques, the first task is to acquire the image from the source. These images can be acquired either through camera or through standard datasets that are available online. The images should be in .jpg format. The images considered here are user dependent i.e. dynamic images. The number of sample training images considered here are 112.

2. Pre-processing

Pre-processing is mainly done to eliminate the unwanted information from the image acquired and fix some values for it, so that the value remains same throughout. In the pre-processing phase, the images are converted from RGB to Gray-scale and are resized to 256*256 pixels. The images considered are in .jpg format, any other formats will not be considered for further processing. During pre-processing, eyes, nose and mouth are considered to be the region of interest. It is detected by the cascade object detector which utilizes Jones-Viola algorithm.

3. Facial Feature Extraction

After pre-processing, the next step is feature extraction. The extracted facial features are stored as the useful information in the form of vectors during training phase and testing phase. The following facial features can be considered “Mouth, forehead, eyes, complexion of skin, cheek and chin dimple, eyebrows, nose and wrinkles on the face”. In this work, eyes, nose, mouth and forehead are considered for feature extraction purpose for the reason that these depict the most appealing expressions. With the wrinkles on the forehead or the mouth being opened one can easily recognise that the person is either surprised or is fearful. But with a person’s complexion it can never be depicted. To extract the facial features PCA technique is used.

PCA Approach

Principal component analysis (PCA) or karhunen-loeve transformation is a statistical approach used for pattern recognition and signal processing to reduce the number of variables in face recognition technique. PCA technique has enormous potential as a feature extractor and is one of the approaches to improve the reliability of the recognition systems. PCA is hugely utilized in all forms of scrutiny because it is a simple method. It likewise lessens the dimension of a figure i.e. from higher dimension to lower dimension space effectively and yet holds the primary information of the images. In this technique, images in the

training phase are represented as weighted eigenvectors that are linearly combined and known as "Eigenfaces", and these eigenvectors are obtained from the covariance grid of a training dataset which is known as basis function. From the largely appropriate Eigenfaces the weights of an image are obtained. Similarity between the pixels among images in a dataset by means of their covariance matrix is one of the advantages taken by Eigen faces. By projecting the test image on the subspace spanned by the eigenfaces, the recognition of the facial expression is performed and then the further classification is carried out by a distance measure method known as Euclidean distance.

Following are the steps involved to recognize the facial expressions using PCA approach:

1. Prepare the data: A 2-D facial image can be represented as a 1-D image by concatenating each of its row (or column) into a thin vector. A set of images with M vectors of size N is considered.
2. Subtract the mean: The average grid is to be figured and subtracted from the first faces, and the result obtained is put away in a variable Φ.
3. Calculate the co-variance matrix: After the above step, the covariance matrix of Φ is calculated.
4. Calculation of the eigenvectors and values of covariance matrix: Now, the eigenvectors and the subsequent eigenvalues of particular images are calculated.
5. Calculate eigenfaces
6. Classifying the faces: The new image is changed into its eigenface segments. The weight vector is shaped from the subsequent weights. The distance between two weight vectors gives comparability measure between the consequent images i & j.

4. Expression Recognition

To recognize and classify the expressions of a person Euclidean distance classifier is used. It gets the nearest match for the test data from the training data set and hence gives a better match for the current expression detected. Euclidean distance is basically the distance between two points and is given by "(3.1)". It is calculated from the mean of the eigenfaces of the training dataset. The training images that correspond to various distances from the mean image are labeled with expressions like happy, sad, fear, surprise, anger, disgust and neutral. When the Euclidean distance between the eigenfaces of the test image and mean image matches the distance of the mean image and eigenfaces of the training dataset the expression is classified and named as per the labeled trained images. Smaller the distance value obtained, the closest match will be found. If the distance value is large enough for an image then the system has to be trained for that individual. The equation to measure Euclidean distance between two points, say p and q is given as:

$$d(p,q) = d(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad (3.1)$$

The following steps are involved in recognition and classification of the expressions:

1. The trained images are utilized to create a low dimensional subspace which is prepared by performing Component Analysis on training dataset and taking the segments that have more noteworthy Eigen.
2. The test images are also projected on the face space (subspace), and all these test images are represented as the selected principal components.
3. The intensity of the particular expression is determined by calculating its distance from the mean of the projected images.
4. The Euclidean distance of a projected test image from all the projected trained images is calculated and the minimum value among them is chosen to find out the trained image which is most similar to the test image.
   (Steps 1 - 4 are carried out for testing phase as well)
5. To find the closest match it is assumed that the test image belongs to the same category as that of the trained image.

5. Play Music

The last and the most important part of this system is the playing of music based on the current emotion detected of an individual. Once the facial expression of the user is classified, the user's corresponding emotional state is recognized. A number of songs from various domains pertaining to a number of emotions is collected and put up in the list. Each emotion category has a number of songs listed in it. When the user's expression is classified with the help of PCA algorithm, songs belonging to that category are then played.

IV. Proposed System

The proposed system goes through various stages in order to get the desired result. Fig 4.1 [2] shows the detailed flow of it. It explains the detailed steps involved in recognizing the expression of the user and playing a music track as according to the expression recognized.

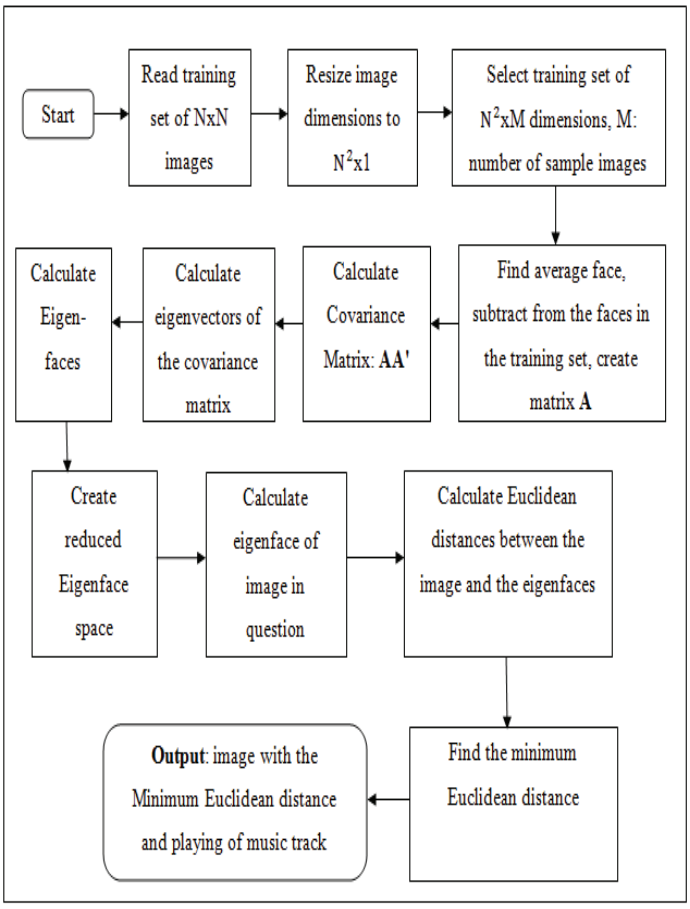

Fig. 4.1 . Flowchart of the proposed algorithm

Initially an image is uploaded as an input of size NxN. Then it undergoes pre-processing technique that is conversion of RGB images into grayscale images to enhance the appearance quality of the input sample, it also resizes the dimensions of these sample images in order to create a lower dimensional space. Here the images are resized to N^2x1. Also in this step, the noise is removed from the input image either by using some filters or by applying median filtering. With this an image will preserve the edges which help in localizing the faces.

Then these images are selected for the training set with dimensions N^2xM where M is the number of sample images. Consider the training set to be $\Gamma_1, \Gamma_2, \Gamma_3, \ldots\ldots, \Gamma_M$. Next an average face is found from this input image. Average face is a standard statistical mean calculated for every single pixel. Equation (4.1) gives the average face of an image.

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n \qquad (4.1)$$

where Ψ is the average face, M is the number of samples and n is the no. of variables.

Then, subtraction of these average faces is done from the faces in the training set and finally a matrix say A is created. Equation (4.2) gives the subtracted image.

$$\Phi_i = \Gamma_i - \Psi \qquad (4.2)$$

where Φ_i is the result obtained, Γ_i is the training image and Ψ is the average face.

Now, the covariance matrix of A is calculated that is AA′. Equation (4.3) calculates the covariance matrix

$$C = \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n' = AA' \qquad (4.3)$$

where A is the matrix of subtracted images $[\Phi_1, \Phi_2, \ldots, \Phi_M]$ and A' is the transpose of it and M is the number of sample images.

The next step after calculating the covariance matrix is to calculate the eigenvectors of the matrix obtained. Equation (4.4) is used to calculate the eigenvector.

$$Av = \mu v \qquad (4.4)$$

where A is a matrix, μ is a scalar known as eigenvalue and v is the vector.

The next step is to find the eigenfaces which is a weighted combination of some components or basis faces. These basis faces are differently weighted to represent different faces. And then the reduced eigenface space (subspace) is created by selecting a set of vectors that are multiplied by A. Equation (4.5) calculates the eigenfaces u_l.

$$u_l = \sum_{k=1}^{M} v_{lk} \Phi_k \qquad (4.5)$$

where v_l is the M eigenvectors with $l = 1, \ldots, M$

With this study, the calculations are significantly reduced from the order of number of pixels (N^2) to the number of images (M) in the training set.

All these steps are carried out for training phase and testing phase. Now in testing phase, the eigenfaces is calculated for the test image. Subsequently, Euclidean distance is calculated among the input image and the eigenfaces that are stored in the training dataset. Lesser the Euclidean distance, the nearest and the best match is found. Then the image with the minimum Euclidean distance is classified as the current expression of the user. Finally, a music track is played automatically on recognition of expression of the user.

V. Experimental Results

The implementation is carried out in Matlab2013a or above. Here, for the facial expression recognition purpose testing was carried out on dynamic images to achieve real time performance. The images were taken through the in-built camera for various individuals. Here, in the training phase 4 sets of 7 expressions were taken for 4 different individuals that resulted into 112 trained images and are stored in the database. When the input image is given to the system it finds for the smallest Euclidean distance between the test image and the trained image, and when the lowest value is found it is displayed as the recognized expression. Table 5.1 shows the accuracy level of all the seven expressions which are obtained after testing the input image.

Table 5.1: Accuracy level

Facial Expression	Accuracy level (%)
Happy	93.75
Sad	75
Anger	81.25
Fear	87.5
Disgust	81.25
Surprise	81.25
Neutral	93.75

The accuracy level is calculated using equation (5.1)

$$Accuracy = \frac{No.of\ relevant\ images}{Total\ no.of\ images\ retrived} X\ 100 \qquad (5.1)$$

where no. of relevant images are the correctly matched images and total no. of images retrieved = no. of relevant and irrelevant images.

Table 5.2 shows the number of images matched from the total number of images present in the database for each category.

Table 5.2: Number of Recognised expressions

Facial Expression	No. Of Images per category	No. Of Recognised Images
Happy	16	15
Sad	16	12
Anger	16	13
Fear	16	14
Disgust	16	13
Surprise	16	13
Neutral	16	15

Fig 5.1 shows the accuracy rates for the recognized expressions in the form of bar graph.

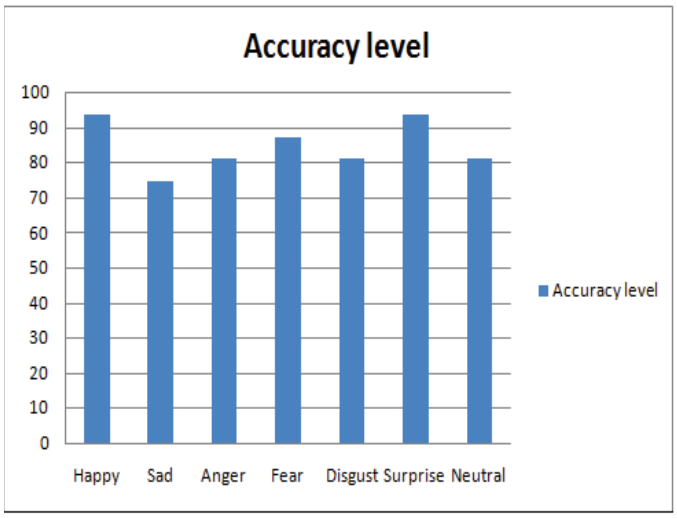

Fig. 5.1 . Bar graph of the expression recognition rate

From fig. 5.1 near accuracy of the results can be seen. The test image considered for expression recognition matches exactly with the images stored in the trained database, but gives less accuracy for sad as it depends on the extracted features during processing and also for the fact that it is difficult to pose for this expression.

Fig 5.2 shows the training phase, where the user poses for all the 7 expressions and saves them. These images are stored in the database. The 3 boxes around the eyes, nose and mouth are considered as the region of interest.

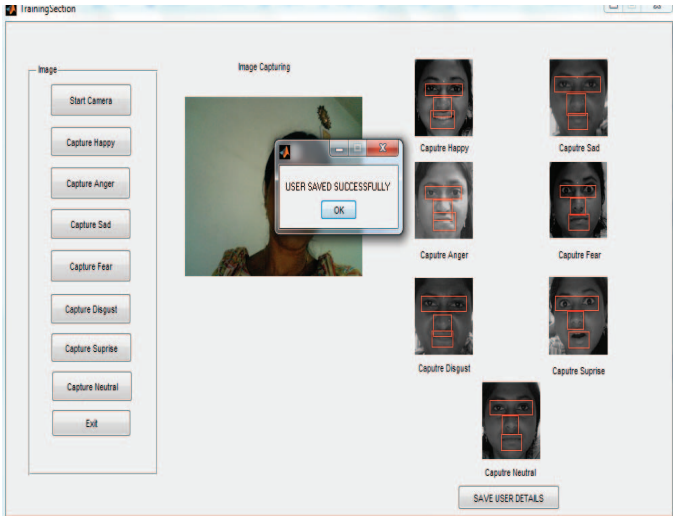

Fig. 5.2 . Training section

Fig 5.3 shows the testing phase, where one of the recognized expression is displayed. The recognized expression in fig 5.3 is surprise.

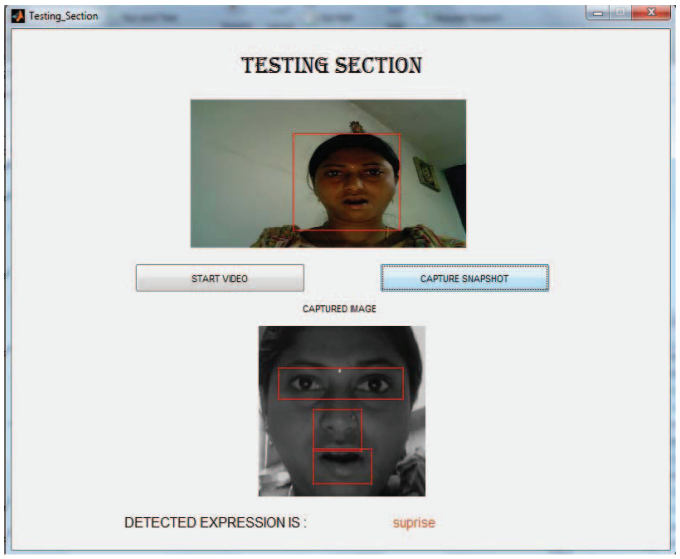

Fig. 5.3 . Testing section

VI. Conclusion

The proposed work presents facial expression recognition system to play a song according to the expression detected. It uses PCA approach to extract features, and Euclidean distance classifier classifies these expressions. In this work, real images i.e. user dependent images are captured utilizing the in-built camera. The final result shows the accuracy level obtained is upto 84.82%.

VII. Future Scope

The future scope for the proposed system would be to implement it on mobiles. To design a mechanism that would help in the music therapy treatment for the music therapists to treat the patients suffering from mental stress, acute depression and trauma. It can also be used to determine the mood of a physically challenged person.

In the proposed work, only one emotion can be detected at a time so it can be extended to mixed mood detection by continuously recording the face of the user.

References

[1] Debasmita Chakrabartia and Debtanu Duttab, "Facial Expression Recognition using Eigenspaces," in *Int. Conf. on Computational Intell.: Modeling Techniques and Applicat.*, 2013, pp. 755-761.

[2] Müge Çarıkçı and Figen Özen, "A Face Recognition System based on Eigenfaces method," *INSODE*, 2011, pp. 118-123.

[3] Hafeez Kabani, Sharik Khan, Omar Khan and Shabana Tadvi, "Emotion based Music Player," *Int. J. of Eng. Research and General Sci.*, Vol. 3, Issue 1, pp. 750-756, January-February 2015.

[4] Nikhil Zaware, Tejas Rajgure, Amey Bhadang and D. D. Sapkal, "Emotion based Music Player," *Int. J. Of Innovative Research & Develop.*, Vol. 3 , Issue 3, pp. 182-186, March 2014.

[5] Kevin C. Tseng, Yu-Te Wang, Bor-Shing Lin and Ping Han Hsieh, "Brain Computer Interface-based Multimedia," in *8th Int. Conf. on Intelligent Inform. Hiding and Multimedia Signal Process.,* Piraeus, 2012, pp. 277 - 280.

[6] Li Wern Chew, Kah Phooi Seng, Li-Minn Ang, Vish Ramakonar, and Amalan Gnanasegaran, "Audio-Emotion Recognition System using Parallel Classifiers and Audio Feature Analyzer," in *2011 3rd Int. Conf. on Computational Intell., Modelling & Simulation*, 2011, pp. 210-215.

[7] Aayush Bhardwaj, Ankit Gupta, Pallav Jain, Asha Rani and Jyoti Yadav, "Classification of Human Emotions from EEG signals using SVM and LDA Classifiers," in *2015 2nd Int. Conf. on Signal Process. and Integrated Networks (SPIN)*, 2015, pp. 180-185.

[8] Timothy Kevin Larkin, "Face Recognition using Kernel," *New Jersey Institute of Technology, Newark, NJ*, 2003.

[9] S L Happy and Aurobinda Routray, "Automatic Facial Expression Recognition using Features of salient Facial Patches," in *IEEE Trans. On Affective Computing*, January-March 2015, pp. 1-12.

[10] Aliaa A. A. Youssif and Wesam A. A. Asker, "Automatic Facial Expression Recognition System based on Geometric and Appearance Features," *Computer And Inform. Science*, Vol. 4, No. 2, pp. 115-124, March 2011.

[11] M.Pantic and I.Patras, "Dynamics of Facial Expression: Recognition of Facial Actions and their Temporal Segments from face profile image sequences," *IEEE Trans. Syst., Man, Cybern.,* vol. 36, no. 2, pp. 433-449, April 2006.

[12] M. Pantic and Leon J. M. Rothkrantz, "Facial Action Recognition for Facial Expression analysis from static face images," *IEEE Trans. Syst., Man, Cybern.,* vol. 34, no. 3, pp. 1449-1461, June 2004.

[13] Ajit P. Gosavi and S. R. Khot, "Facial Expression Recognition Using Principal Component Analysis," *Int. J. of Soft Computing and Eng. (IJSCE),Volume-3, Issue-4, September 2013.*

[14] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. of Cognitive Neuroscience*, vol. 3, no. 1, March 1991.