ISSN 2348 – 7968

Audio Visual Mood Detection System By Using Various Dimension Reduction Techniques.

Rajneesh Singla, Gurmukh singh

Assistant Professor, UIET department, Panjab University Chandigarh.
Assistant Professor,UIET department, Panjab University Chandigarh

Abstract- In this proposed work, in order to improve the accuracy of mood detectoion system, the technique of Bimodal integration is applied on speech and facial expression database. The objective of this research work is that psychological state giving information about some disorders helpful with diagnosis of depression, mania or schizophrenia. The elimination of errors due to reflections in the image has not been implemented but the algorithms used in this project are computationally efficient to resolve errors. Endeavours are also put in this project to enhance the recognition rate of mood detection by adopting unique methodology. In this research work we have accepted five different moods to be recognized are: Happy, Sad, Neutral, Surprise and Fear. The main part of this project is to focus on image recognition and speed recognition. To recognize speech we have used PCA techniques and for Facial expression detection we use combination of ICA&FLD techniques.

Index terms- Bimodal fusion, FLD, ICA,, PCA and SVM

I. INTRODUCTION

Audio visual mood recognizes five basic facial expressions (e.g. happiness, neutral state, etc.) that have already been detected/tracked in video frames or camera. Mood Recognition Module provides functions to identify the facial expressive state of each person appearing in video at every video frame. The set of facial and speech expressions can be recognized includes happiness, sadness, surprise, fear and the neutral state. Mood recognition through speech and facial expression is an area which increasingly attracting attention within the engineers in the field of pattern recognition. Facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person; it plays a communicative role in interpersonal relations. Facial expressions, and other gestures, convey non verbal communication cues in face-to-face interaction. These cues may also complement speech by helping the listener to elicit the intended meaning of spoken words. The facial expressions have a considerable effect on a listening interlocutor; the facial expression of as speaker accounts for about 55 percent of the effect, 38 percent of the later is conveyed by voice intonation and 7 percent by the spoken words. As a consequence of the information that they carry, facial expressions can play an important role wherever humans interact with machines.

This paper presents a dual –mode recognition system based on speech and facial expression to increase the rate of single mode emotion recognition. Proposed system is shown in **Figure 1**. This system takes audio and image input and extracts the features as per requirement and then use two different classifiers. After this the output of two classifiers is combined together and final result was displayed. For emotion detection through speech SVM is used and for emotion recognition we use Baye’s classifier.

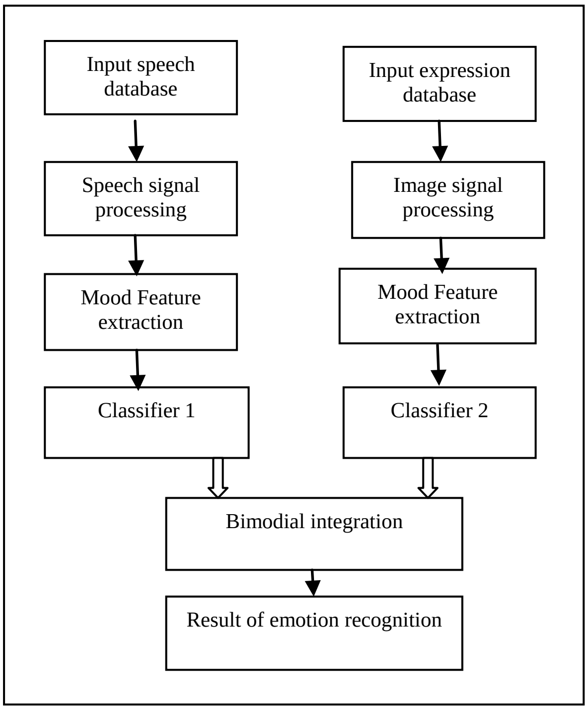

Fig. 1 Flow chart of bimodal fusion for emotion detection

The remaining paper is organized as section II describes related work and section III describes methodology for mood detection system. Section IV describes the methodology for speech based emotion detection. Section V gives information about database. Section six describes the conclusion.

II. RELATED WORK

The bimodal fusion method based on speech and facial expression was proposed to improve single-mode recognition rate.. Here emotion recognition rate can be defined as ratio of number of images properly recognized to the number of input images. Single mode emotion recognition term can be used either for emotion recognition through speech or through facial expression. Shruti and Savita (2014) used two methods to increase the recognition rate by using bimodal fusion. To do the emotion detection through facial expression they used adaptive sub layer compensation (ASLC) based facial edge detection method and for emotion detection through speech we use well known SVM. Then bimodal emotion detection is obtained by using probability analysis.

Peter and Joao (1997) developed a face recognition algorithm which is insensitive to large variation in lighting direction and facial expression. Taking a pattern classification approach, they considered each pixel in an image as a coordinate in a high-dimensional space. They took advantage of the observation that the images of a particular face, under varying illumination but fixed pose, lie in a 3D linear subspace of the high dimensional image space-if the face is a Lambertian surface without shadowing. They used eigenface and fisherface as a computation method.

James *et al.*, (2000) proposed most of the current work on automated facial expression analysis attempt to recognize a small set of prototypic expressions, such as joy and fear.. They developed the first version of a computer vision system that is sensitive to subtle changes in the face. The system includes three modules to extract feature information: dense-flow extraction using a wavelet motion model, facial-feature tracking, and edge and line extraction. The feature information thus extracted is fed to discriminant classifiers or hidden Markov models that classify it into FACS action units, the descriptive system to code fine-grained changes in facial expression. The system was tested on image sequences from 100 male and female subjects of varied ethnicity.

Fadi and Franck (2003) addressed the recognition of facial expressions in continuous videos. They introduced a view- and texture independent approach that exploits facial action parameters estimated by an appearance-based 3D tracker. They also represented the learned facial actions associated with different facial expressions by time series. These time series are then efficiently and compactly represented in Eigenspace and Fisherspace for subsequent recognition. The developed approach was fast and can be used online. Experiments demonstrated the effectiveness of the developed method.

A) FACIAL MOOD DETECTION

Facial moods are one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions. Facial moods carries crucial information about the mental, emotional and even physical states of the conversation. . It is a desirable feature of the next generation human-computer interfaces. Computers that can recognize facial expressions and respond to the emotions of humans accordingly enable better human-machine communication development of information technology Recognition of facial expression in the input image needs two functions: locating a face in the image and recognizing its expression. We believe recognition of human facial expression by computer is a key to develop such technology. In recent years, much research has been done on machine recognition of human Facial expressions. Conventional methods extract features of facial organs, such as eyes and a mouth and recognize the expressions from changes in their shapes or their geometrical relationships by different facial expressions when we watch two photos of a human face, we can answer which photo shows the facial expression more strongly. Accordingly, as extending the step of facial expression recognition, we think it is important to develop a measurement method of the strength of facial expressions. One of the key **re**maining problems in face recognition is to handle the variability in appearance due to changes in pose, expression, and lighting conditions. There has been some recent work in this direction. The increasing progress of communication technology and computer science has led us to expect the importance of facial expression in future human machine interface and advanced communication, such as multimedia and low-bandwidth transmission of facial data In human interaction, the articulation and perception of facial expressions form a communication channel, that is additional to voice and that carries crucial information about the mental, emotional and even physical states of the conversation (Hyung and Daijin, 2006). Face localization, feature extraction, and modelling are the major issues in automatic facial expression recognition (Michel *et al*.,2011).

III. METHODOLOGY FOR MOOD DETECTION SYSTEM

A) ICA Algorithm

Independent component analysis used for finding underlying factors or components from multivariate (multidimensional) statistical data. There is need to implement face recognition system using ICA for facial images having face orientations and different illumination conditions, which will give better results as compared with existing systems. The ICA is similar to blind source separation problem that boils down to finding a linear representation in which the components are statistically independent. The comparison of face recognition using PCA and ICA on FERET database with different classifiers were discussed and found that the ICA had better recognition rate as compared with PCA with statistically independent basis images and also with statistically independent coefficients. Face recognition using ICA with large rotation angles with poses and variations in illumination conditions was proposed in. In ICA each face image is transformed into a vector before calculating the independent components. RC_ICA reduces face recognition error and dimensionality of recognition subspace becomes smaller. Kailash and Sanjay (2011) proposed a novel technique for face recognition combined the ICA model with the optical correlation technique. This approach relied on the performances of a strongly discriminating optical correlation method along with the robustness of the ICA model. This model had sparked interest in searching for a linear transformation to express a set of random variables as linear combinations of statistically independent source variables. ICA provided a more powerful data representation than PCA as its goal was that of providing independent rather than uncorrelated image decomposition and representation.

B) FLD Algorithm

Fisher's linear discriminant are methods used in statistics, pattern recognition and machine learning to find a linear combination of features which characterize or separate two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

In the other two methods (ICA and PCA) however, the dependent variable is a numerical quantity, while for LDA it is a categorical variable (i.e. the class label). Logistic regression and probit regression are more similar to LDA, as they also explain a categorical variable. These other methods are preferable in applications where it is not reasonable to assume that the independent variables are normally distributed, which is a fundamental assumption of the LDA method.

LDA is also closely related to principal component analysis (PCA) and factor analysis in that both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made.

LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis.(Rajneesh and Simpel 2011)

C) MECHANISM

Flowchart of Algorithm

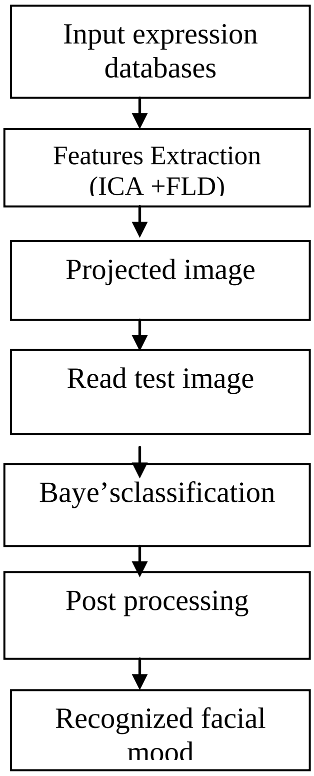

Fig. 2 Proposed system for facial expression based emotion detection

As the above mentioned figure explains the overall flow of my thesis from starting to the end point. This Flow chart depicts all the process of mood detection which are explained step wise below. This flowchart represents also depicts the type of resources used in the process of mood detection system. According to this flowchart we have designed the basic code of our mood detection system in Matlab.

For a research work a database of 50 images is created with having a subject of different moods this process was done with the help of a camera (5 megapixels) and the images collected were stored in a file system rather than any sophisticated software like Oracle simply because it is an academic research work and there is no issue of security. This folder containing different moods stored in it is called as a Training folder (Fig. 3)

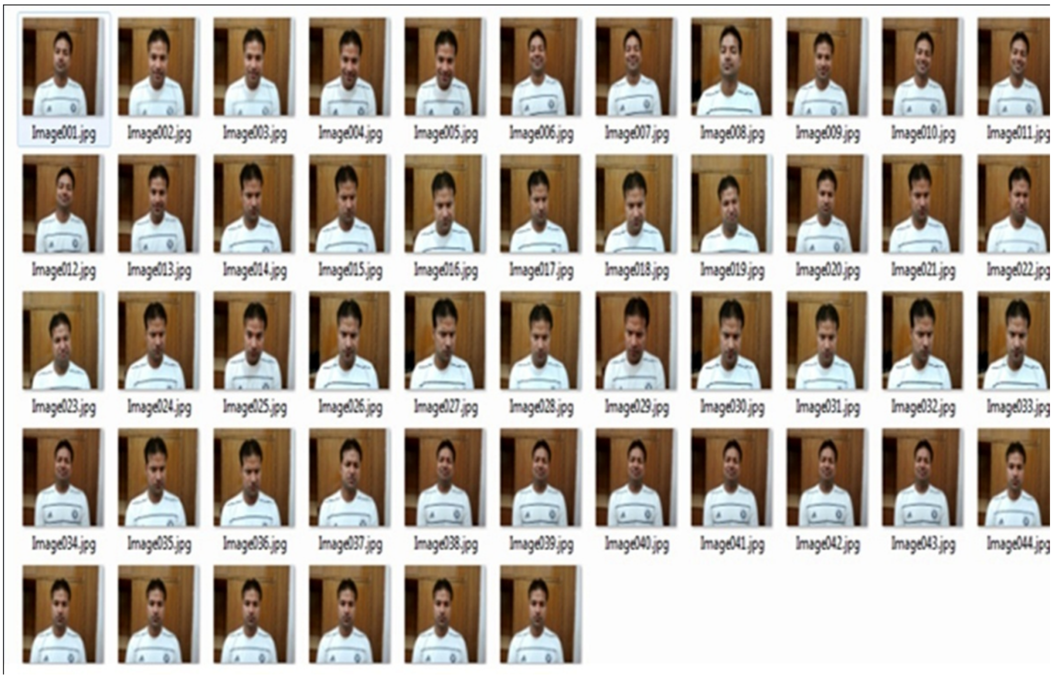

Fig. 3 Training Folder containing Expression Database

To check the accuracy and validity of our system we have created a scrambled version of images and were put in a folder called Test folder. One is to one mapping was done for each image with respect to the expression of the subject in the photograph. This mapping was stored in a text file and loaded in the workspace of Matlab while executing the program. In an iterative process each trained image, test image and expression is read and loaded in memory for execution. Once the loading completes the test data matrix or the big T-matrix or the database is now ready for dimension reduction using ICA and FLD.

A projected image is created after multiplying the ICA, FLD and mean of the images. Once the projected image is ready then each test image is compared with the trained image using a distance classifier. If the distance is equal to zero this means a perfect match else the index of the closest image is extracted. Simultaneously with the respect to this image expression is extracted from the expression database and finally it is displayed as a result. In case it fails to find an image within a threshold the message is displayed that the image is not recognized.

The implementation of the algorithm is done in Matlab 7.1 environment and Image Processing toolbox in Matlab is used. The database used for Facial Expression System is Real time database. For the implementation of mood detection a real time database data is used. Face database contains 50 colored face images of individual. There are 5 images per subject, and these 5 images are, respectively, under the following different facial expressions or configuration. In this implementation, all images are resized to a uniform dimension of 256 x 256. The samples of expression database is shown in **Figure 4**

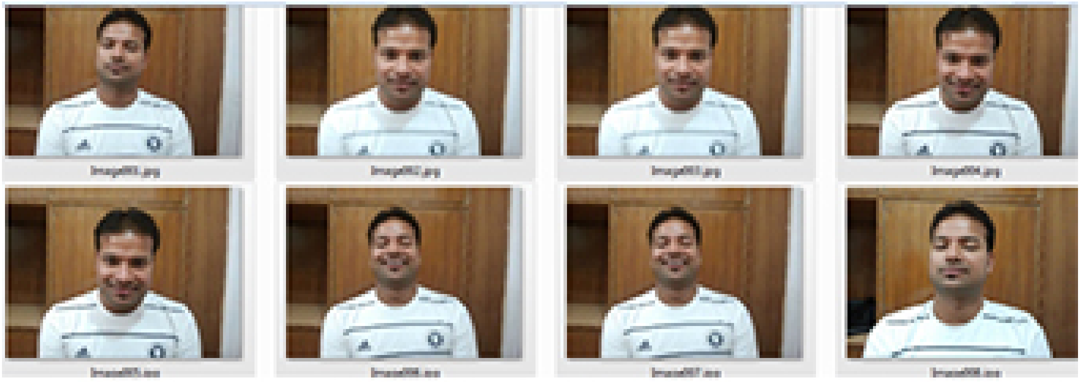

Fig. 4 Samples of database used for training the recognition system

IV METHODOLOGY FOR SPEECH BASED EMOTION DETECTION

The block diagram of the emotion recognition system through speech considered in this study is illustrated in **Figure 5**. Emotion recognition system through speech is similar to the typical pattern recognition system. An important issue in evaluation of Emotion recognition system through speech is the degree of naturalness of the database used. Proposed system is based on prosodic and spectral features of speech. It consists of the emotional speech as input, feature extraction, classification of Emotional state using SVM classifier and detection of emotion as the output.

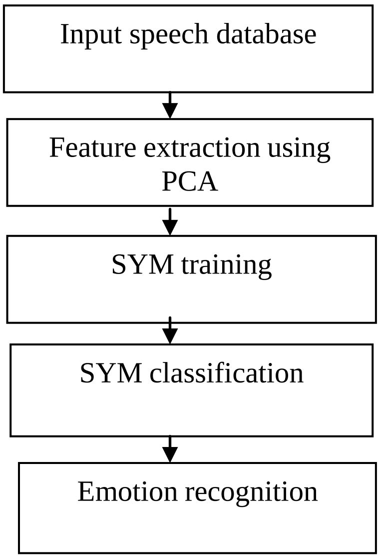

Fig.5 Emotion recognition through speech

A) SPEECH DATABASE

The emotional speech system may contains the collection of the acted speech data the real world speech data. After collection of the database containing short Utterances of emotional speech sample which was considered as the training samples, proper and necessary features such as prosodic and spectral features were extracted from the speech signal. These feature values were provided to the SVM for training of the classifiers. Then recorded emotional speech samples presented to the classifier as a test input. Then classifier classifies the test sample into one of the emotion from the mentioned five emotions and gives output as recognized emotion.

B) PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT), the Hostelling transform or proper orthogonal decomposition (POD).

PCA was invented in 1901 by Karl Pearson. Now a days used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute. The results of a PCA are usually discussed in terms of component scores (the transformed variable values corresponding to a particular case in the data) and loadings (the weight by which each standardized original variable should be multiplied to get the component score).

PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a "shadow" of this object when viewed from its (in some sense) most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.

PCA is closely related to factor analysis; indeed, some statistical packages (such as Stata) deliberately conflate the two techniques. True factor analysis makes different assumptions about the underlying structure and solves eigenvectors of a slightly different matrix.

C) SUPPORT VECTOR MACHINE (SVM)

Support Vector Machines (SVM) are one of the most useful techniques in classification problems. One clear example is face recognition. However, SVM cannot be applied when the feature vectors defining samples have missing entries. A classification algorithm that has successfully been used in this framework is the all-known Support Vector Machines (SVM), which can be applied to the original appearance space or a subspace of it obtained after applying a feature extraction method . The advantage of SVM classifier over traditional neural network is that SVMs can achieve better

generalization performance.

V) DATABASE AND BIMODAL FUSION

Based on facial expressions, I have used my own Real Time database which includes 100 videos of eight subjects from the RML Emotion database and 100 videos from Real Time database. This database is modified by using face localization and then used for our experiment. Total 50 images are used as a database. For each emotion 8 images are used. **(Table 1 and 2)**

S.NO.	EMOTIONS	RATE IN %
1.	HAPPY	70
2.	SAD	57
3.	NUETRAL	48
4.	SURPRISE	40
5.	FEAR	58

Table 1. Expected result classification rate for SVM classifier

S.NO.	EMOTIONS	RATE IN %
1.	HAPPY	98
2.	SAD	95
3.	NUETRAL	94
4.	SURPRISE	93
5.	FEAR	82

Table 2. Expected result of Bimodal fusion.

VI) CONCLUSION

The aim of this project was to explore the area of audio visual mood detection. A wide variety of image processing techniques was developed to meet the mood recognition system requirements. However, there are still many challenges and problems to solve in such systems, especially in the area of their performance and applicability improvement. In this project we proposed PCA, FLD, ICA and Speech Database methods for dimension reduction of different types of facial moods. The proposed algorithm is successfully implemented on Real time database. Experiments results show that algorithm can effectively recognize different emotions by indentifying different feature.

It is also concluded that FLD, ICA is still required to be more optimized to reduce computational time and combine usage of audio and visual database is more effective than using single database alone. Also we observed that there are still some setbacks in our project where efficiency needs to be considered. As we know that our methodology has higher mood detection rate but the algorithm still needs to be more optimized so as to reduce the computational time and sometimes it is inevitable to trade of between accuracy and speed

REFERENCES

1. Anil K. Jain, Rober P.W.Duin and Jiancchang Mao,(2000)" Statistical Pattern Recognition: A Review", *IEEE Trans. Pattern Analysis and Machine Intelligence,*Vol.22,No.1.pp. 4-37

2. Fadi Dornaika and Franck Davoine,(2005)” Facial Expression Recognition In Continuous Videos Using Linear Discriminant Analysis" *IAPR Conference on Machine Vision Applications*”, pp. 277-28.
3. Hyung-Soo Lee, Daijin Kim,(2006),” Facial Expression Transformations for Expression-Invariant Face Recognition”, *Springer-Verlag Berlin Heidelberg*, pp. 323–333.
4. James Jenn-Jier Lien, Takeo Kanade, Jeffrey F. Cohn, Ching-Chung,(2000),” Detection, tracking, and classification of action units in facial expression”, *Robotics and Autonomous Systems*”,pp. 131-146
5. Kailash J. Karande, Sanjay N.Talbar,‖ Face Recognition under Variation of Pose and Illumination using Independent Component Analysis‖, *ICGST-GVIP,ISSN* 1687-398X, Volume (8), Issue (IV), December 2008.
6. Michel F. Valstar, Bihan Jiang, Marc Mehu, Maja Pantic,(2011),” The First Facial Expression Recognition and Analysis Challenge”, *Imperial College London, UK.*
7. Rajneesh Singla (2015). Mood detection system via using various dimension reduction techniques.(2015) ISBN 978-3-659-79560-2.
8. Peter N. Belhumeur, Joao P. Hespanha, David J. Kriegman,(1997), Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. Pattern Analysis and Machine Intelligence* , VOL. 19, NO. 7,pp.711-72
9. P. Ekman,(1992),”Facial Expression and motion”,*Am.psychologist*,vol.48, pp.384-39.
10. Rajneesh Singla, Simpel Jindal, (2011). “Comparitive study of dimension reduction techniques for mood detection”, *Research Cell: An International Journal of Engineering Sciences* Vol.4, pp.45-61.
11. Simpel Jindal and Rajneesh Singla, (2011). “A new approach for mood detection via using principal component analysis and fisherface algorithm” *Journal of Global Research in Computer Science*, Vol. 2, No. 7, pp. 88-92.
12. Yi Huang and Yun Tie (2013) ” Human Emotion Recognition using the Adaptive Sub-Layer-Compensation Based Facial Edge Detection,” IEEE.