Facial expression based emotion detection,A Review

Minakshee Sarma[1], Kaustubh Bhattacharyya[2]
Don Bosco College of Engineering and Technology, Assam Don Bosco University
Airport Road, Azara, Guwahati - 781017, Assam. INDIA.
[1]sarma.momi213@gmail.com
[2]kaustubh.bhattacharyya@dbuniversity.ac.in

Abstract: Emotion detection is the task of recognizing a person's emotion state. Understanding facial expression accurately is one of the challenging task for interpersonal relationship. Automatic emotion detection using facial expressions recognition is now a main area of interest within various fields such as computer vision, medicine and psychology. Various feature extraction techniques have been developed for recognition of expression from static images as well as real time videos. Artificial Neural Network (ANN) based detection for emotion like anger, confusion, happy, sad, annoyed, stressed etc. is now a days, gathering more popularity among the researchers as it provides better results. Human emotion can be detected image through digital processing. Few work done on emotion detection, those are published recently reviewed are summarizes here briefly.

Keywords—Emotion, interpersonal, facial expression, ANN, Recognition, extraction, digital processing.

1. INTRODUCTION

Human are often more emotional, than we wish to be, and our feelings influence the way we work, play and interact with computers. Effective computing is a domain that focuses on user emotions while interacting with computers for such applications [1, 2]. As emotional state of a person may influence concentration, task solving and decision making skills, effective computing vision is to make systems able to recognize and influence human emotions in order to enhance productivity and effectiveness of working with computers [3, 4].

Emotion recognition and effective intervention are nowadays well recognized desired features of intelligent tutoring systems [5], with primary focus on such learner effective states as flow, boredom or frustration. Another areas of effective computing methods applications include: testing driver stress, psychological diagnosis and training, neuro-biofeedback etc.

Recognition of facial expression results in identifying the basic human emotion like anger, fear, disgust, sadness, happiness and surprise. These expressions can vary in every individual. Researchers [1,6] have found that 7% of message is conveyed by spoken words, 38% by voice intonation while 55% of message is conveyed by facial expression. Facial expressions [5] are produced by movement of facial features.

The basic mechanisms of the emotion detection system are as follows [7]. First step is face detection. First the machine takes an image, then by skin color segmentation it detects human skin color and then it detects human face. Then background need to be separated so as to obtain the region of interest of the captured image. So second is normalization phasethat remove the noise and normalize the face against brightness and pixel position. In third phase features are extracted and irrelevant feature are eliminated. In the final step basic expressions are classified into six basic emotions like anger, fear, disgust, sadness, happiness and surprise.

Emovoice [8] is a comprehensive framework for real time recognition of emotion from acoustic properties of speech. It has been recently integrated as toolbox into the social signal Interpretation (SSI) framework which is also from the Lab for multimedia concept and application.

There are several types of classifiers such as ANN [9][10] Hidden Markov models (HMM) [11], statistical tools etc. HMM have been widely used for many classification and modeling problems. One of the main advantages of HMMs is their ability to model non-stationary signals or events. Dynamic programming methods allow one to align the signals so as to account for the non-stationarity[9]. However, the main disadvantage of this approach is that it is very time-consuming since all of the stored sequences are used to find the best match of emotion expression, the signal is the measurements of the facial motion. Again the efficiency of a statistical classifier is poor as it cannot classify samples with minor differences. However researchers have found the ANN as one of efficient classifier for stationary as well as non-stationary signals. ANNs are non- parametric prediction tools that can be used for a host of pattern classification and speech recognition purposes. It is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information [9]. The fundamental information-processing unit of the operation of the ANN is the McCulloch-Pitts Neuron (1943).

2. EMOTION DETECTION SYSTEM

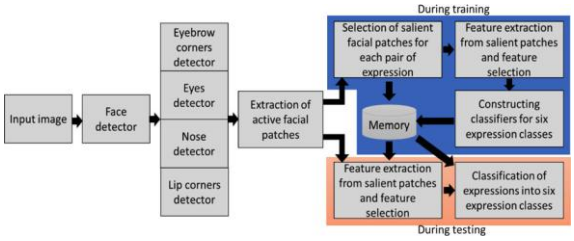

Figure. 1. Block diagram of emotion detection [12]

The overview of the method is shown in Figure. 1 [12]. Observations from Figure. 1 suggests that accurate facial landmark detection and extraction of appearance features from active face regions improve the performance of expression recognition. Therefore, the first step is to localize the face followed by detection of the landmarks [12]. A learning-free approach is proposed [12] in which the eyes and nose are detected in the face image and a coarse region

of interest (ROI) is marked around each. The lip and eyebrow corners are detected from respective ROIs. Locations of active patches are defined with respect to the location of landmarks. Figure shows the steps involved in automated facial landmark detection and active patch extraction.

Basically any face recognition system can be depicted by the following block steps [8,3]:

1. Pre-processing unit: In the initial phase, the color may be converted to gray scale image, raised to the required image, noise is removed, illumination normalization is carried out by Histogram Equalization and expression normalized is carried out using higher decomposition of Discrete Wavelet Transform.
2. Feature Extraction: In this phase facial features are extracted using edge detection technique, Principle Component Analysis (PCA) technique, Discrete Cosine Transform coefficients etc.
3. Training and testing: Here Euclidian Distance, Hamming Distance, Support Vector machine, neural network and Random Forest may be used.

3. PREPROCESSING

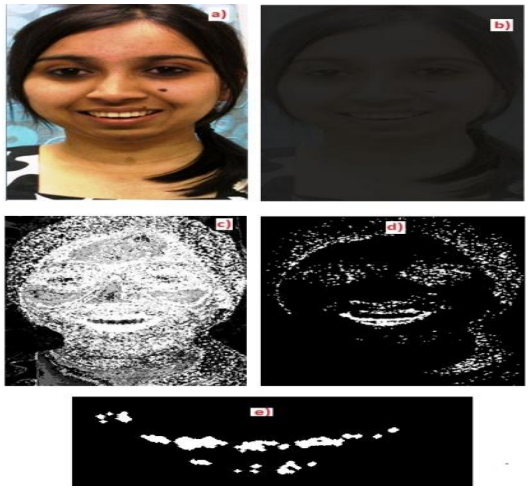

Figure. 2. A few preprocessing steps for cropping lip from original image [12]

a) Original image b) Normalized image c) Hartley transformed image

d) Gaussian filtered image. e) Cropped dilated image

The preprocessing is an important step in recognition system design [13]. The efficiency of the system is based on the outcome of the preprocessing steps. Different types of preprocessing steps are a) Original image. b) Normalized image. c) Hartley transformed image d) Gaussian filtered image. e) Cropped dilated image Figure 2 shows some preprocessing steps and the resultant outcome for different emotion state is shown in figure 3. The final outcome of the preprocessing steps is then fed to feature extraction block. In some cases the preprocessed image can be directly fed to the classifier for detection purpose. Either the extracted features or the raw preprocessed images are used to make the database to be used for training.

.

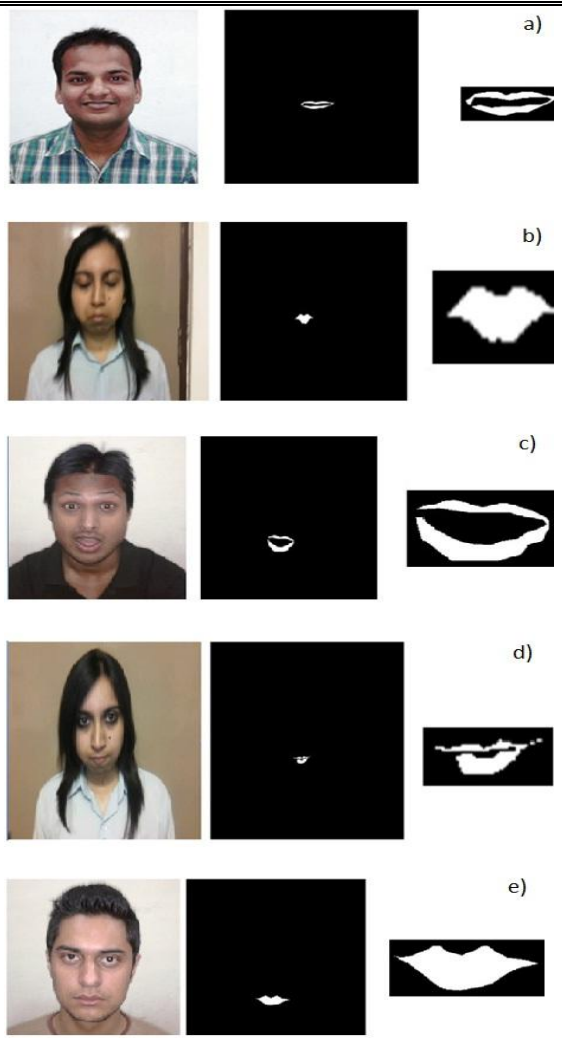

Figure3: Final outcome with the original image for different emotion states.[11]

a) Happy
b) Sad
c) Surprised
d) Angry
e) Natural

The preprocessing is an important step in recognition system design [13]. The efficiency of the system is based on the outcome of the preprocessing steps. Different types of preprocessing steps are a) Original image. b) Normalized image. c) Hartley transformed image d) Gaussian filtered image. e) Cropped dilated image Figure 2 shows some preprocessing steps and the resultant outcome for different emotion state is shown in figure 3. The final outcome of the preprocessing steps is then fed to feature extraction block. In some cases the preprocessed image can be directly fed to the classifier for detection purpose. Either the extracted features or the raw preprocessed images are used to make the database to be used for training.

4. FEATURE EXTRACTION

So far, numerous research projects have been done on recognizing emotion from face. Many methods are used to implement their systems. Facial expressions provide the building blocks with which to understand emotion. In order to effectively use facial expressions, it is necessary to understand how to interpret expressions, and it is also

important to study what others have done in the past. For self-organized learning method, principle component analysis
(PCA) is widely used in the field of data compression and feature extraction. There are two basic approaches to the computation of principal components: batch and adaptive methods. The batch methods include the method of Eigen decomposition and the method of singular value decomposition (SVD), while the adaptive methods are mainly done by neural networks. The main target of PCA is to explain the variance–covariance structure of the data through a few linear combinations of the original variables. The main concerning thing about PCA is that it utilize only the global information of face images, this method is not very effective for different facial expressions.
As mentioned earlier several classifier are generally used for image detection, however, the ANN is one of the best bio – inspired based classifier as discussed by researchers [9-12]. Lineardiscriminant analysis (LDA)method is used in statistics and pattern recognition to find a linear combination of features. The resulting combination may be used as a linear classifier or, more commonly, for dimensionality reduction before later classification. 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.

5. ANN based classification for emotion detection

Figure. 4 show a generic neuron structure. The block diagram of Figure. 5 show the model of a neuron, which forms the basis for designing ANN. From this model the interval activity of a neuron k can be shown to be:

$$u_k = \sum_{j=1}^{m} w_{kj}\, x_j$$

and

$$y_k = \varphi(u_k + b_k)$$

Where, x_1, x_2 … are the input signals, w_{k1}, w_{k2}, ...w_{km} are the synaptic weights of neuron k, b_k is the bias, $\phi(.)$ is the activation function and y_k is the output signal of the neuron.

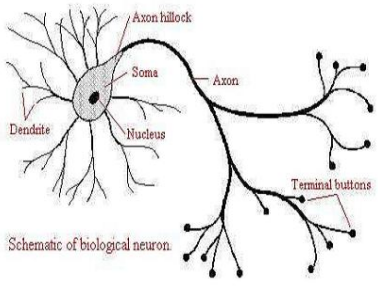

Figure4:A generic biological neuron[10]

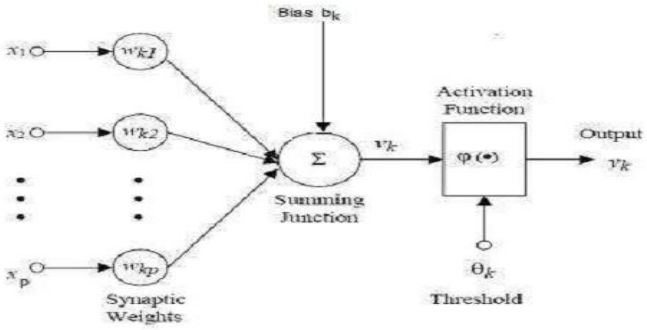

Figure5: A Nonlinear model of a neuron[10]

The activation function acts as a squashing function, such that the output of a neuron in ANN is between certain values (usually 0 and 1, or -1 and 1). Three basic types of activation functions [11] are Threshold Function, Piecewise-Linear function and sigmoid function. Learning algorithms define an architectural-dependent procedure to encode pattern information into weights to generate these internal models. Learning proceeds by modifying connection strengths [10]. There are two types of learning algorithm- Supervised Learning and Unsupervised Learning. Supervised learning employs the teaching input D_k, the associates of X_k to reduce the error (D_k, S_k) in the response the system. Input-output sample pairs are employed to train the network through a simple form of error correction learning or gradient descent weight adaptation. These procedures are based on global error measures derived from the difference between the desired (D_k) and actual (S_k) output of the network. In Unsupervised learning the alternate way of learning is to simply provide the system with an input X_k, and allow it to self-organize its parameters, which are the weight of the network to generate internal prototypes of sample vector. An unsupervised learning system attempts to represent the entire data set by employing a small number of prototypical vectors enough to allow the system to retain a desired level of discrimination between samples.

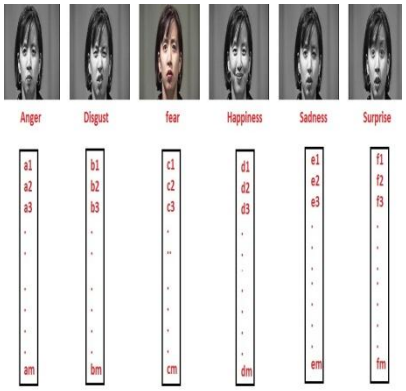

Figure6: The database creation method for ANN training [15]

As new samples continuously buffer the system, the prototypes will be in a state of constant flux. There is no teaching input. This kind of learning is often called adaptive vector quantization and is essentially unsupervised in nature. The ANN based emotion detection can be nicely described with figure. 7 to figure. 9. First as shown in figure. 6, a data base is formed with features of emotion images. These data base are now used to train the ANN classifier.

As shown in figure 8, the ANN is built with 3 layers of neurons namely input layer, hidden layer and output layer. The number of neurons [10] of input layers can be any number, generally depends on system complexity. The number of hidden layers should be around 1.5 times the input neurons. The number of output layers depends on the number of different input samples that has to be trained.

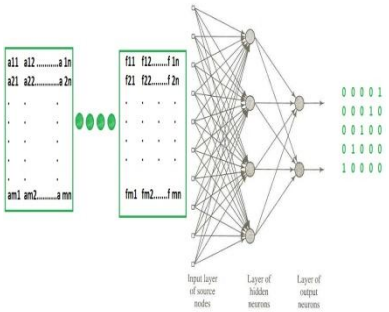

Figure7:Training of ANN with n samples for same emotion state.[14]

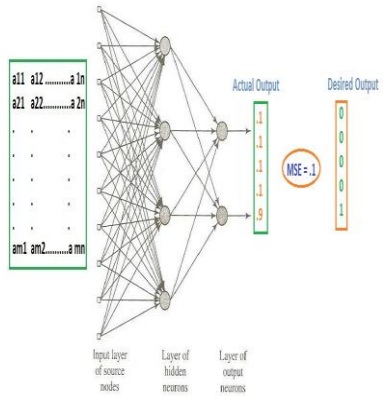

Figure8:Testing the trained ANN with some sample vector.[14]

As shown in figure. 7, the trained ANN is tested with some sample. The desired output of the input sample is column matrix of [0 0 0 0 1]', however the mean square error is taken as .1, so the actual output of [.1 .1 .1 .1 .9]' is taken as matched condition. Another Approach is to make Appearance based Models. These rely on techniques from statistical analysis and machine learning to find the relevant characteristics of face images. Some appearance-based methods work in a probabilistic network. An image or feature vector is a random variable with some probability of belonging to a face or not.

Another Approach is to make Appearance based Models. These rely on techniques from statistical analysis and machine learning to find the relevant characteristics of face images. Some appearance-based methods work in a probabilistic network. An image or feature vector is a random variable with some probability of belonging to a face or not.

Neural Networksis another key tool of extracting feature from an image. Many pattern recognition problems like object recognition, character recognition, etc. have been faced successfully by neural networks. These systems can be used in face detection in different ways.

6. SVM BASED EMOTION DETECTION

Support Vector Machines (SVM)s [16] are linear classifiers that maximize the margin between the decision hyper plane and the examples in the training set. So, an optimal hyper plane should minimize the classification error of the unseen test patterns. SVM classify data with a set of support vectors by minimizing the structural risk, the average error between input and their target vectors is reduced[17]. This classifier was first applied to face detection by Osuna et al. With the use of the SVM package LIBSVM, Dumas [18] has proposed an emotion classification approach. The LIBSVM has being developed by ChihChung Chang and Chih-Jen Lin. The objective of this study was to determine the highest possible accuracy attainable with SVM to classify the Pictures of Facial Affect (POFA) dataset [17]. The POFA dataset is a static human facial image dataset.

7. HMM BASED EMOTION DETECTION

Hidden Markov Model [19]is another statistical model that has been used for face detection. This classification is done at an image sequence level. Each frame of an image sequence is represented by a feature vector, which is mapped to one of the words from the dictionary generated using K-means. Latent Dirichlet allocation [20] then models each image sequence as a set of topics. We further know the order of topics for image sequence from the order of words which we use for classification in the next step, done by training a HMM for each emotion. The challenge is to build a proper HMM, so that the output Probability can be trusted. The states of the model would be the facial features which are often defined as strips of pixels. The probabilistic transition between states is usually the boundaries between these pixel strips. As in the case of Bayesians, HMMs are commonly used along with other methods to build detection algorithms. The accuracy obtained on the HMM technique is 80.77%. The use of word sequence is found to give better result in general.

8. Conclusion

As we know that we can recognize human emotion using facial expressions without any effort or delay but reliable facial expression recognition by computer interface is still a challenge. An ideal emotion detection system should recognize expressions regardless of gender, age and any ethnicity. Such a system should also be invariant to different distraction like glasses, different hair styles, facial hairs and different lightening conditions. It should also be able to construct a wholeface if there are some missing parts of the face due to these distraction. It should also be good facial expression analysis regardless of large change in viewing conditions and rigid movements. Achieving optimal feature extraction and classification is a key challenge in this field because we have a huge variability in the input data. For better recognition rate more current facial expression recognition methods require some work to control imagining conditions like position and orientation of the face with respect to the camera as it can result in wide variability of image views. More research work is needed for transformation invariant expression recognition.

Emotion detector can be useful in the field of video gaming, medicine and marketing and also in driver safety as in addition to fatigue. Although several ways of emotion recognition are there, the facial expression recognition can be an effective tool with greater efficiency. ANN based emotion detector could be able to work as a good human computer interaction tool. However precise feature extraction method has to be adopted in order to have reliable and flaw less system.

ACKNOWLEDGEMENT: The authors would like to acknowledge Assam Don Bosco University, Guwahati for providing the resources necessary for compiling this review article.

References

[1] Picard R,"Affective computing: From laughter to IEEE", IEEE Transactions on Affective Computing,Vol.1, No. 1,pp.11-17, 2010.
[2] DamirFilko, GoranMartinovic, "Emotion Recognition System by a Neural Network Based Facial Expression Analysis", ATKAFF, Vol. 54, No. 2, pp. 263–272, 2013.
[3] Shrutika C. Dhargave ,ApurvaSonak ,VandanaJagtap M.E. Student, CE Dept, 3Assistant Professor, CE Dept. MAEER'S MIT, Pune "Survey Paper on Methods of Emotion Detection from Still Images"
[4] Zeng, Z. et State College,Pennsylvania, USA, "Bimodal HCI-related Affect Recognition. ICMI", October pp.13–15, 2004.
[5] Picard R,"Affective computing: challenges". International Journal of Human-Computer Studies, Vol. 59 No. 1, pp.55-64,2003
[6] Healey J, Picard R,"Detecting stress during real-world driving tasks using physiological sensors". IEEE Trans. on Intelligent Transportation Systems, Vol 6, No. 2, pp.156-166,2005.
[7] Hone K "Empathic agents to reduce user frustration: The effects of varying agent characteristics. Interacting with Computers",Vol. 18 No. 2, pp.227–245,2006
[8] P. M. Chavan, Manan C .Jadhav, Jinal B. Mashruwala, Aditi K. Nehete, Pooja A. Panjari,"Real Time Emotion Recognition through Facial Expressions for Desktop Devices", International Journal of Emerging Science and Engineering (IJESE), ISSN: 2319–6378, Vol.1, No. 7, pp.104 -108,2013
[9] Christos Stergiou and DimitriosSiganos: "Neural Networks", Vol. 4, 1996.
[10] Simon Haykin, "Neural Networks A Comprehensive Foundation", Pearson Education, 2nd edition, 2003.
[11] I. Cohen, A. Garg and T. S. Huang," Emotion Recognition from Facial Expressions using Multilevel HMM",2015.
[12] IEEE Transaction on Affective Computing, Vol. 6, No. 1, January-march 2015
[13] Pushpaja V. Saudagare, D.S. Chaudhari "Facial Expression Recognition using Neural Network –An Overview" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Vol. 2, No. 1, March 2012
[14] DipjyotiSarma and Kandarpa Kumar Sarma, "Multicore Parallel Processing Architecture for ANN based Speech and Image Processing Applications", in the Journal of Instrument Society of India, IISC, Bangalore, India, Vol. 42, No. 2, pp. 92-94, 2012.
[15] Kenny Hong, Member, IEEE, Stephan K. Chalup, Senior Member, IEEE, and Robert A.R. King, "Affective Visual Perception Using Machine Pareidolia of Facial Expressions",Vol. 5, No.4,October-december 2014
[16] WathsalaNayomiWidanagamaachchi,"Facial Emotion Recognition with a Neural Network Approach",Registration Number : 2005CS155
[17] M. Dumas, "Emotional expression recognition using support vector machines." Technicalreport, Machine Perception Lab, Univeristy of California, 2001.
[18] International Journal of Computer Science & Engineering Survey (IJCSES) Vol.5, No.2, April 2014
[19] Akhil Bansal, SantanuChudhary and Sumantra Dutta Roy, "A Novel LDA and HMM Based Technique for Emotion Recognition from Facial Expression". IIT, Delhi, India
[20] DamirFilko, GoranMartinovi´c "Emotion Recognition System by a Neural Network Based Facial Expression Analysis". pp.263–272(2013)
[21] Nermine Ahmed Hendy and HaniaFarag. World Academy of Science, Engineering and Technology, "Emotion Recognition Using Neural Network: A Comparative Study",Vol. 7, March 2013
[22] SakibHussain,Abu Mohammad Hammad Ali, "Thesis Report Emotion Detection from Frontal Facial Image",
[23] RohiniPatil, C.G.PatilSinhgad academy of engineering, Kondhwa, Pune(India)"Automatic face emotion recognition and classification using Genetic Algorithm" IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) e-ISSN: 2278-1676 ,p-ISSN: 2320-3331, Vol. 9, No. 5 Ver. II pp. 63-68.Sep – Oct. 2014

Authors' Profile

Minakshee Sarma has completed her B.Tech degree from the Department of electronics and telecommunication engineering Of Institution of electronics and telecommunication engineers (IETE) India in 2015. She is currently pursuing her M.Tech degree in the department of ECE with specialization in Digital Signal Processing in School Of Technology, ADBU, Guwahati, India. Her area of interest is Image Processing and Communications.

Kaustubh Bhattacharyya is currently serving as an Assistant Professor in the Electronics and Communication Engineering Department of Assam Don Bosco University, Assam, India. Prior to this he held research positions at Gauhati University, Assam, India. He completed his Bachelor of Science in Electronics, Master of Science in Electronics Science, Master of Philosophy in Electronics Science and Master of Technology in Electronics and communication Technology from Gauhati University, Assam, India in 2004, 2007, 2009 and 2012 respectively. His research interest is in high frequency communication and soft computing.