



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Realising Human Empathy Using Artificial Neural Network

Rhea Mahajan and Sandeep Sharma
Department of Computer Science and Engineering,
Guru Nanak Dev University Amritsar, Punjab, India

Abstract: Human facial expression is controlled by neurons and their neural network. It is also dependent on the muscles of the skin that enables the user to convey the behavior and carry inter human communication. Facial expression recognition is a challenging problem in computer vision. In this paper, we have proposed a novel approach to recognize facial expressions from static images, via extraction and selection of salient features points and finding the corresponding feature vectors. We have also constructed an authentic emotion databases by combination of the available databases and the real time images that we have used in our experimental work. Finally, we perform several experiments and show that our neural network approach can be successfully used for emotion recognition. Our main aim is to work upon five different emotions which include neutral, happy, sad, surprised and angry. Using only two-thirds of the total features, our approach achieves a classification rate (CR) which is higher than the CR obtained using all features. Our system also outperforms several existing methods, evaluated on the combination of existing and self-generated databases.

Keywords: emotion recognition, classification rate, self generated database

I. INTRODUCTION

Human beings possess and express emotions in everyday interactions with others. Emotions are often reflected on the face, in hand and body gestures, in the voice, to express our feelings or liking. Recent psychology research has shown that the most expressive way humans display emotions is through facial expressions. Mehrabian[1] indicated that the verbal part of a message contributes on for 7% to the effect of the message as a whole, the vocal part (e.g. voice intonation) for 38%, while facial expressions contributes for 55% to the effect of the speaker's message. In addition to providing information about the affective state, facial expressions also provide information about cognitive state, such as interest, boredom, confusion, and stress, and conversational signals with information about speech emphasis and syntax.

While a precise, generally agreed definition of emotion does not exist, it is undeniable that emotions are an integral part of our existence, as one smiles to show greeting, frowns when confused, or raises one's voice when enraged. The fact that we understand emotions and know how to react to other people's expressions greatly enriches the interaction. There is a growing amount of evidence showing that emotional skills are part of what is called "intelligence". Computers today, on the other hand, are still quite "emotionally challenged." They neither recognize the user's emotions nor possess emotions of their own [2]. In the past 20 years there has been much research on recognizing emotion through facial expressions. This research was pioneered by Paul Ekman [6] who started his work from the psychology perspective. In the early 1990s the engineering community started to use these results to construct automatic methods of recognizing emotions from facial expressions in images or video based on various techniques of tracking [3].

An important problem in the emotion recognition field is the lack of agreed upon benchmark database and methods to compare existing methods' performance. The various approaches [4][5] are:

A. The static approach: In this approach the classifier classifies each frame in the video to one of the facial expression categories based on the tracking results of that frame. Bayesian network classifiers were commonly used in this approach. While Naive-Bayes classifiers were often successful in practice, they use a very strict and often unrealistic assumption that the features are independent given the class. Therefore, another approach using Gaussian TAN classifiers have the advantage of modelling dependencies between the features without much added complexity compared to the Naive-Bayes classifiers. TAN classifiers have an additional advantage in that the dependencies between the features, modelled as a tree structure, are efficiently learned from data and the resultant tree structure is assured to maximize the likelihood function.

B. The dynamic approach: These classifiers take into account the temporal pattern in displaying facial expression. Hidden Markov model (HMM) based classifiers for facial expression recognition has been previously used in recent works. Cohen and Sebe [4] further advanced this line of research and proposed a multi-level HMM classifier, combining the temporal information, which allowed not only to perform the classification of a video segment to the corresponding facial expression, as in the previous works on HMM based classifiers, but also to automatically segment an arbitrary long video sequence to the different expressions segments without resorting to empirical methods of segmentation.

II. AN IDEAL SYSTEM FOR FACIAL EXPRESSION RECOGNITION: CURRENT PROBLEMS

There are three main factors to construct a Facial Expression Recognition system, namely face detection, facial feature extraction and emotion classification. An ideal emotion analyzer should recognize the subjects regardless of gender, age and any ethnicity. The system should be invariant to different lightening conditions and distraction as glasses, changes in hair style, facial hair, moustache, beard etc. It should also be able to "fill in" missing parts of the face and construct a whole face. It should also perform robust facial expression analysis despite large changes in viewing condition, rigid movement, etc. A good reference system is the human visual system [4]. The current systems are far from ideal and they have a long way to achieve these goals.

A. FACE DETECTION AND FEATURE EXTRACTION

By using FACS[12], expression is described as a combination of Action Units(AUs) which are anatomically independent and are visually distinguishable. Facial Action Cording System (FACS) is an expression descriptive method which was proposed by Ekman and Friesen, and is widely applied to the researches about facial expressions. There are more than 40 AUs[9][10] and it is said that every sort of facial expression can be described by the AUs. For example, the combination of nation of AU6 (rise cheeks) and AU12 (pull up edgesof mouth) can express "smile".

B. FACIAL MUSCLE AND THREE REGIONS OF FACE AREA

The variations of expressions are caused by contractions of facial muscles. There are over 20 facial muscles, and the combination of their contractions and relaxation cause various complex expressions. The face area can be divided into the following three regions which can move independently.

Upper region: Eyebrows and forehead,

Middle region: Eyes and base of nose and

Lower region: Mouth and cheeks

The basic expressions are are "anger", "neutral", "suprise", "happiness", "sadness" and "surprise fig 2.. Each expression has distinctive actions in the three face regions.

In this paper, we mainly focus on the emotion classification part, not much on face detection or on facial feature extraction. The feature value extraction part calculates the feature values from dynamic facial images captured by a color camera. The procedure is that 18 feature points set on the facial organs are first extracted, then 16 feature values are calculated from the positions of the feature points.

The feature value extraction part consists of the following 4 steps.

- 1.Extraction of face region,
- 2.Extraction of facial organs,
- 3.Extraction of feature points and
- 4.Calculation of feature values

Taking into consideration of well-description of AUs and easiness for image processing, the feature values were decided as shown in Figure 1. As the feature points to calculate the feature values, 3 points on both eyebrows, 4 points on both eyes and mouth, totally 18 points were selected. The feature values calculated from these points are input to the classifier.

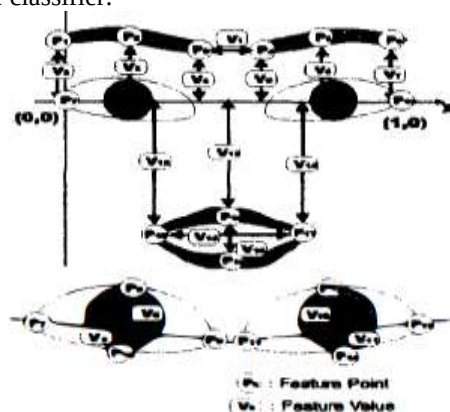


Fig. 1 : Feature point and values



Fig. 2 : Basic emotions

III. CLASSIFICATION TECHNIQUE

The neural network used in our classification method is a feed forward multi-layer perceptron which follows back propagation algorithm using sigmoid function and three hidden layer. Other techniques used are SOM[15]. SOM, known also as Kohonen's self-organizing map, which is a clustering technique that can be used to automatically provide insight into the nature of data without supervision. We can transform this unsupervised neural network into a supervised LVQ (Learning Vector Quantum) neural network, which is another clustering technique. But it was seen when handling noise and multiple inputs of data, back propagation performs better than SOM. LVQ is excellent for classification, but when handling noise it is a little bit worse than back propagation. This is special reason that makes back propagation better than the numerous other neural network models.

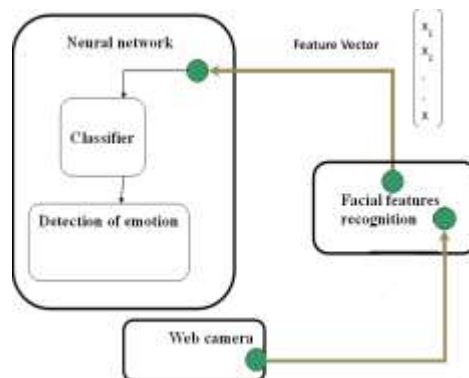


Fig. 3: Proposed Methodology

In this paper we are using the Artificial Neural Net technique to increase the generalization accuracy in face recognition and it's a powerful technique to predict not only for the seen data but also for the unseen data. However, Linear Discriminant Analysis is another powerful technique but it can predict only for the seen data and not the unseen data and also when data set is high dimensional the performance highly degrades.

Factors for classification

The back propagation technique actually depends on the three main factors:

No. of Iterations, No. of local maxima and local minima Error difference rate

Our designed application would be able to judge the efficiency on all the parameters but the most efficient one is the error difference rate. This parameter would just help to provide a stopping condition or a stopping parameter to our algorithm.

According to this parameter as the in one reading of a face the value is suppose 76.37 while matching with a neutral face and in next reading for the same expression reading is 76.3705 the algorithm would stop and say that the result is obtained for the neutral expression i.e. it will match the expression as neutral

The technique of the back propagation used forms a six layered network with one input layer, two output layers and three hidden layers. It actually contains 16 inputs for the 16 input feature values in input layer, then it contains 5 nodes in the first hidden layer 40 units in the second and third hidden layer and finally 5 in the first output layer for the five expressions and 2 in the second output layer for detecting the male and female facial organs. This classification technique is using the log sigmoid function as the activation function of the neural network.

IV. IMPLEMENTATION AND RESULTS

An ideal way to get a natural emotion data is using a secret camera, and thus not telling the person in advance that he is being filmed. However, this type of experiments may cause some ethical problems, so we only tested with our friends and colleagues. They were told that we are doing a psychology test or a test with the image quality after decompression.

To find enough persons to spend time giving different expressions, sit at a proper distance and pose correctly to allow the camera to catch their faces is a very tedious and difficult task. Therefore, we decided to use static images, even if they are not desirable to build up most of the emotions, such as sad, anger or even surprise. It was also difficult and time consuming to find suitable situations that could arouse people's emotions immediately, i.e., it is tricky to capture the angry emotion without annoying them.

Despite all these difficulties, in total we tested 28 people. However, some people did not always sit in a proper position, or they were talking and moving. Others didn't even show any emotion or their expressions could not be differentiated.

In the end, we selected the data from 9 persons, including 5 females and 4 males. We did not put any constraint on their age. We did this test mostly with short clips, thus it was more likely to get only some particular expressions such as neutral, joy, and sometimes surprise.

The girls participating in the whole experiment showed several emotions such as joy, anger sad and surprise. Some of them confessed they were very tensioned and scared but they didn't show the obvious emotions as we expected. They only looked very serious, while others showed surprise but blended with fear. Taking all these into account, we decided to select the data of neutral, joy, sad, angry and surprise.

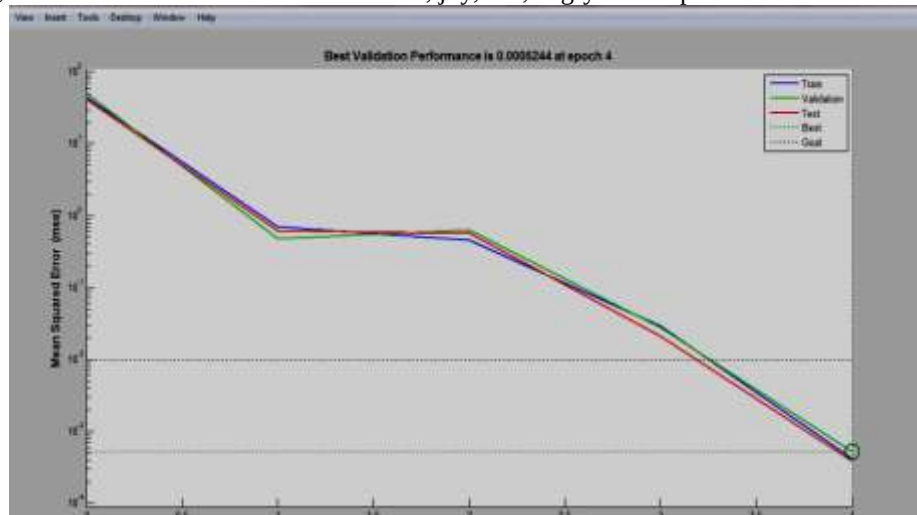


Fig. 6 : Performance plot

V. DISCUSSION AND CONCLUSION

After training and testing various images, performance plot was generated[fig.6] which showed that mean square error was reduced to a great extent and on the basis of that recognition rate was calculated which was around 60-70%.. Apart from emotions our algorithm could also recognize gender which is one of the improvements over previous static and dynamic approaches. It was also seen when handling noise and multiple inputs of data, our method performs better than SOM and LVQ. We can further make our system efficient by making it able to handle blended emotions or subtle. i.e. surprise with fear or surprise with joy. Increasing the number of output layers can be a feasible solution to this problem. This method can further be improved by using the improved techniques for the thresholding and edge detection in the image. Recognizing the facial expressions can also help us to study other moods psychologically

REFERENCES

- [1] A.Mehrabian."Communication Without Words." Psychology Today, 2,(1968)
- [2] Lyons, M., &Bartneck, C. "*HCI and the Face*". Proceedings of the Conference on Human Factors in Computing Systems (CHI 2006), Montreal, pp 1671 – 1674.
- [3] HuaGuGuangda and Su Cheng Du, "feature points extraction from faces", image and vision computing: Palmerton North, November 2003
- [4] Nicu Sebe , Michael S. Lew , Ira Cohen , Ashutosh Garg , Thomas S.Huang Emotion Recognition Using a Cauchy Naive Bayes Classifier ICPR, 2002.
- [5] Ira Cohen Nicu Sebe , Larry Chen, AshutoshGarg , Thomas Huang , Facial Expression Recognition from Video Sequences: Temporal and Static Modeling Computer Vision and Image Understanding (CVIU) special issue on Face recognition
- [6] P. Ekman and W.V. Friesen.Facial Action Coding System: Investigator's Guide. Consulting Psychologists Press, Palo Alto, CA, 1978.
- [7] K. Scherer and Ekman.Handbook of methods in nonverbal behaviour research.Cambridge University Press, Cambridge, UK, 1982.
- [8] James J. Lien, Jeffrey F. Cohn, Takeo Kanade and Ching-Chung Li, "automated facial expression recognition based on face action units",IEEE (1998).
- [9] Ying-li-Tian , Takeo Kanade and Jeffrey F. Cohn. "recognizing lower face action units for facial expression analysis", IEEE (2000).
- [10] Ying-li-Tian , Takeo Kanade and Jeffrey F. Cohn. "recognizing upper face action units for facial expression analysis", IEEE (2000).
- [11] Ying-li-Tian , Takeo Kanade and Jeffrey F. Cohn. "recognizing action units for facial expression analysis", IEEE (2001)
- [12] H.ShimadaT.Kunihiro,D.Yang,H.Yoshikawa, "design of affective interface for realizing human-machine empathy", IEEE (2000).
- [13] Hong-Bo Deng ,Lian – Wen Jin ,Li-Xin Zhen, Jian –Cheng Huang, "a new facial expression recognition method based on local gabor filter bank and pca plus lda" . International Journal of Information Technology Vol. 11 No. 11 (2005)
- [14] MariofannaMilanoval, Nikolay Sirakov2 , "recognition of emotional states in natural human-computer interaction", IEEE (2008).
- [15] Byoung-Jun Park, Eun-Hye Jang, Sang-Hyeob Kim, Chul Huh, Myoung-Ae Chung and Jin-Hun Sohn ,"emotion recognition using autonomic nervous system responses",The Sixth International Conference on Advances in Computer-Human Interactions (2013).