Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP 2007)

Cognitive Model – Based Emotion Recognition From Facial Expressions For Live Human Computer Interaction

Maringanti Hima Bindu, Priya Gupta, and U.S.Tiwary, *Senior Member, IEEE*

Abstract—**Human emotions are deeply intertwined with cognition. Emotions direct cognitive processes and processing strategies of humans. The goal of this work is to design a model with the capability of classifying the uncertainty, contradiction and the cognitive nature of the emotions. For achieving this, 3D cognitive model is designed. This model enhances our vision of classification of emotions produced by reinforcing stimuli. In this model the dimensions represent the positive reinforcers, the negative reinforcers and the emotion content present. The positive reinforcer increases the probability of emission of a response on which it is contingent, whereas the negative reinforcer increases the probability of emission of a response that causes the reinforcer to be omitted. This model increases the number of emotions, that can be classified. Presently this model can classify 22 emotions subject to the presence of a facial expression database. It has the flexibility to increase upon the number of emotions. For emotion (pattern) identification, the pose and illumination factor are removed using Gabor wavelet transforms and the size is reduced by finding its principle components (PCA). This component vector is used for training the neural network. The test result shows the recognition accuracy of 85.7% on The Cohn-Kanade Action Unit Coded Facial Expression Database. The real time processing for identification, aids in applying emotions to real time audio player. An environment, that is all pervasive or ubiquitous, that would sense one's mental state and play the appropriate musical track to maintain the positive emotional state or ease from a negative emotional state.**

Index Terms— **Cognitive model, Emotions, 3D architecture, reinforcing stimuli.**

Manuscript received November 9, 2006.

Maringanti Hima Bindu is an Assistant Professor with the Indian Institute of Information Technology, Allahabad, India 211011 (phone: +91-532-2922096,+91-9335070621; fax: +91-532-2430006; e-mail: mhimabindu@iiita.ac.in).

Priya Gupta was with Indian Institute of Information Technology, Allahabad, India 211011 as a post graduate student.

Prof.U.S.Tiwary is with the Indian Institute of Information Technology, Allahabad, India 211011, also as a Dean of Academic Affairs. (e-mail: ust@iiita.ac.in).

I. INTRODUCTION

Cognitive Science is an interdisciplinary science which includes the mental states and processes such as thinking, remembering, perception, learning, consciousness, emotions etc. Of all the kinds of cognition, emotion holds the key for human social behavior. Identification and classification of emotions by computers has been a research area since Charles Darwin's age. These actions could be performed with the help of facial images, blood pressure measurement, pupillary dilation, facial expressions and many more quantifiable attributes of humans [1].

Facial expression recognition is an area which poses great challenges for the researchers. It is an area where a lot has been done and a lot more can be done. Facial expression recognition is not a theoretical field but finds practical applications in many fields. Coupled with human psychology and neuroscience it can come up as an area which can bridge the divide between the more abstract area of psychology and the more crisp area of computations.

The characteristic feature points [10] of a face are located at eyebrows, eyelids, cheeks, lips, chin and forehead. The feature points after being extracted from these regions help in recognizing the various expressions of a face. The first and the most important step in feature detection is to track the position of the eyes. Thereafter, the symmetry property of the face with respect to the eyes is used for tracking rest of the features like eyebrows, lips, chin, cheeks and forehead. Splitting face into two halves eases the process further. This paper uses Discrete Hopfield Networks as the basis for pre-processing of the Face-Signal and then Feature Extraction.

A number of techniques have been proposed in this field and are being used which include Bayesian Classification [11],Gabor Wavelet Transform [12], Principle Component Analysis, HMM [13], Line-based Caricatures [14], Method of Optical Flow Analysis [15] etc. But they have an inherent complexity which makes them opaque and are computationally expensive.

Apart from emotion classification and identification, the response from the computers is also being generated for making the human-computer interaction livelier. The field of affective computing has emerged for this kind of developments [2].Most of the work in the field of emotions has been performed on six basic emotions of humans. The architectures [9] defined in this paper by the authors for representing all the six emotions are 2D in nature. These do not provide space for expressing the contradicting or mixed emotions at the same instant of time. For this constraint, an architecture has been developed, which overcomes this limitation by defining the emotions in terms of the positive and negative reinforcers, which are mutually exclusive, along with the emotion content present. It is 3D in nature, which has a reward or positive reinforcers on x-axis, punishment or negative reinforcers on y- axis and the emotion content on the z-axis. This architecture, instead of classifying to a crisp value of the emotion, fuzzifies the emotion by depicting the positive and the negative reinforcers present in the input for a particular emotion content. The architecture is described in the next section, followed by Model Design in section II. Section III describes the basic building blocks for the model. Section IV and V explain the experiment performed and the results obtained. The paper ends with discussion and conclusion in sections VI and VII.

II. Cognitive Model of Emotions

A. Description

To simulate emotions on machines, a mathematical model is required. Emotions are the states produced by the reinforcing stimuli, both the positive and the negative stimuli. On the basis of this reinforcement, emotions are classified into two categories. The emotions produced after achieving a reward or the omission of punishment are termed as positive emotions and the emotions produced after getting a punishment or the omission of a reward are termed as the negative emotions [3]. For example, happy, pride, enthusiasm etc are the first category emotions and sad, anger, shame etc are the second category emotions. For modeling this nature of emotions, a 3-dimensional model (Figure 1) is developed. The dimensions in the model have immense importance, the first two dimensions represent the amount of reward and punishment present in the emotion and the other represents the position of emotion in the 3D emotional space. With the help of this model , emotions can be expressed as a vector V(R, P, E) and for each emotion, this model has unique values. The value of R and P is directly proportional to the content of reward and punishment in the emotion. The value of E represents the position of emotion on the third axis in a 3D space.

B.Fuzzy Rules

For writing down the rules for this architecture, the manhanttan distance between the desired value, that is a theoretical or ideal state (expected subjectively) of emotion vectors and the actual value of emotion vector is calculated from the architecture. On the basis of this difference, rules are written for the classification of facial image to an emotion class. The authors have also experimented with Euclidian distance, which is giving slightly better results. The fuzzy membership function for an emotion vector A and desired value X is,

If $|X - A| \leq 0.03$

Then A belongs to emotion X i.e. present (X) = 1

If $|X - A| > 0.03$ & If $|X- A| \leq 0.05$

Then A belongs to the same class with lesser emotion content.

$$\text{If } |X-A| > 0.05 \tag{1}$$

Then A does not belong to the emotion category X.

The rules with Euclidian distance as the basis are respectively, Eucldist(X,A) $\leq$ 0.04 ; inbetween 0.04 and 0.07 and greater than 0.07.

III. System Architecture

For the architecture defined in section II, we develop a system model for fuzzily classifying emotions and playing music as an application on the basis of the emotion identified. The block diagram for it is shown in Figure 2.

For the classification and identification, the input used is a facial image. In this image, initially the face[10] region is found and its intransient features extracted, i.e. eyes, eyebrows and mouth. These extracted facial features are used for finding out the facial expression vectors by calculating their difference from the average image of the facial features. Facial expressions are enhanced by removing the pose and illumination constraint and reducing the dimensionality to only the important components in the vectors produced. These vectors are used for learning the emotion patterns. The vectors produced are used for identifying emotions corresponding to the architecture described above. The fuzzy rules are applied corresponding to the different values obtained in three dimensions. On the basis of the fuzzy classification, a smart audio application is played corresponding to the emotion present on the face. For a happy mood, some hearty songs are played to maintain it and for a sad mood some rejuvenating ones are played.

Figure 3 shows the flow of facial expressions' extraction. It takes the raw image as input and then finds the face region in it by simple edge detection and histogram approach. The intransient features are extracted from the face region using the Hough transform for detecting eyes in the gray scale images and face geometry.

The features extracted are cleaned and enhanced. From the different vectors produced, the principal components are taken and are fed as input in Figure 4. The architecture defined in Figure 4 is used for learning emotional patterns by the system.

IV. Expression Representation and Extraction

Expressions of a human face can be captured through facial features. There are two types of facial expression features, transient (wrinkles and bulges) and intransient (mouth, eyes and eyebrows). Expressions are extracted by capturing the changes in shape and position of the intransient features from the neutral face. Maximum expressions appear in the vicinity of intransient features [4].

For capturing only the changes, the emotion content areas are subtracted form their average neutral image. Then for expression enhancement, Gabor Wavelet (GW) based filters are applied. Gabor filters are capable of detecting line endings and edge borders over multiple scales and with different orientations [8]. These features show a lot about facial expressions, as both transient and intransient facial features often give rise to furrows, wrinkles and contrast changes. GW Transform is defined as a convolution,

$$Jj(x) = \int I(x)^* \Psi j\ (k,x)dx \qquad (2)$$

where I(x) is the gray scale image and Ψj is the j-th kernel function and x = (x,y) represents the coordinates of a given pixel. The results of transformation are described by complex values, so we can calculate their amplitudes as the final transformation results [Movellan, 2002].
A 2D Gabor wavelet (GW) kernel function is defined as:

$$\Psi j\ (\mathbf{k},x) = (|\mathbf{k}j| \div \sigma^2\ ) \times \exp(-(\ |\mathbf{k}j^2|\ \ |x|^2\ ) \div 2\sigma^2\ )[\exp(i\mathbf{k}j \times x) - \exp(-\sigma^2 \div 2)] \qquad (3)$$

Each Ψ_j is a plane wave characterized by the vector k_j and enveloped by a Gaussian function, where σ is the standard deviation of this Gaussian.

The center frequency of j^{th} filter is given by the characteristic wave vector and different GW functions are defined by selecting different values of k_j, they are defined as:

$$k_j = k_y\ e_i\ \Phi_w \qquad (4)$$
$$k_i = k_y\ [\cos(\Phi_w) + i \sin(\Phi_w)]$$

where, $k_y = \pi/2[(v+2)/2]$, which represents the frequencies of a kernel wavelet and Φ_w represents the orientation of GW. After testing with different combinations of v and Φ values, the values of v = 2 and $\Phi w = \pi/6,\ 2\pi\ /6,\ 3\pi/6,\ 4\pi/6,\ 5\pi/6,\ \pi$ were set for achieving the best results.
The image(figure 5) is of M*N (640 * 480) pixels and after applying the GW filters, we have M*N*6. For achieving the maximum information, the average(mean) value is taken out from all the six filtered images. Then this average value is subtracted from each filtered image. Now the image is converted into a column vector of Px6 (P =M*N). PCA is applied to reduce the dimensionality without loosing the information content. The input vector I of size $P \times 1$ is constructed .

V. EXPERIMENT

The steps included in experimentation are the same as explained in Figure 3 and 4. The database used for experimentation is the Cohn-Kanade Action Unit Coded Facial Expression Database [Donato et al., 1999]. From this database, images corresponding to eight emotions are unambiguously identified. Initially the image of 640*480 is taken as input and the face region is found using histogram method. From the face, the intransient features are extracted. The eyes and mouth are used for computation and their dimensions are 100*200 and 150*80. The average neutral image for eyes and mouth is calculated. For the input image, the difference of the eyes and mouth with their average image is taken. Six Gabor filters of frequency $k_v = \pi$ and phase difference $\Phi_w = w\ (\pi/6)$, where w = 1,2,….,6 are used. The six filters obtained are used for computing PCA. For each facial image, the PCA vector is 1280 * 1 dimensionality. PCA obtained is given as input to the neural network for learning. It has two hidden layers of 69 and 9 computing neurons. It uses gradient based Back-propagation algorithm for learning.

VI. RESULTS

The results with the elementary processing abilities form the building blocks, hence the results obtained are promising. With the training performed without momentum, the Back-propagation algorithm requires 21,000 epochs for reducing the mean square error to 0.001. The learning graph for the system architecture is shown in Figure 5. Corresponding to the fuzzy rules developed, the classification performed is satisfactory. The accuracy in classifying eight emotions (happy, anger, disgust, fear, neutral, sad, joy and surprise) of emotion classification using manhanttan distance method is 85.7%, while the Euclidian distance method with the parametric values changed, is giving us an efficiency of 87.5%. The latest experiment was totally conducted with a training set of 80 and tested with a set of 40, of which 35 facial expressions were rightly recognized. The learning graph is shown in figure 6 and the results in the Table 1 below it.

VII. DISCUSSION

Understanding human emotions has various applications and can be used universally. This is another form of language used by living beings and this is universally spoken and understood. It has the same building blocks all over the world. These are subjective in nature and are ambiguous in nature. The crisp boundaries do not exist between the emotions. For these characteristic of emotions, we have developed this architecture which has fuzzy rules to classify emotions. This classification architecture is based on the reinforcement learning concept. The likely situations are reinforced positively and the unlikely ones are negatively reinforced. Using this mechanism and the emotion content in a facial expression, this 3D architecture is designed. 22 emotions have been classified with this architecture as evident from[fig.1]. This classification capability is extendable. The constraint with the architecture is the presence of database corresponding to the emotions. Most of the data prepared by researchers is for the six basic emotions. The technique of Gabor wavelet transform is implemented for the reduction of noise, pose effect and illumination effect from the images. This cleans up the facial image for further processing. PCA of the different Gabor filters produced, corresponding to a facial image is taken to reduce the dimensionality to only the important features of the facial image. Gabor Filters are found to have better performance in noise and other reductions as compared to independent component analysis (ICA) [5] [6] [8]. The learning process is performed in batch mode, so that the system learns the emotion pattern instead of face of the subject. Manhanttan distance metric is used for distance calculation as it finds the correlation between the axis independently.

VIII. CONCLUSION

Emotions being subjective in nature could be best represented using fuzzy rules and their expression on the basis of the reward and punishment quotient makes them closer to the cognitive mechanism for classification. This model provides space for quantifying emotions in terms of positive quantifiers and negative quantifiers along with the amount of emotion present in the facial images. On the third axis[fig.1], where the emotion content has been obtained, there is a congestion observed , both at the top(positive) and the bottom(negative). These are the emotions, which have very subtle differences or very nearly the same and hence very meager. The steps for enhancement are performed with an aim to reduce the effect of noise in data collection. The research for the enhancement step has been focused on the collection of ideas for the processing to be performed optimally.

REFERENCES

[1] B. Rosalind W. Picard, Elias Vyzas, and Jennifer Healey.Toward, " Machine Emotional Intelligence: Analysis of Affective Physiological State", *IEEE Transactions on pattern analysis and machine intelligence*, volume 23, number10, October 2003.

[2] D. A. Norman, A. Ortony, D. M. Russell, "Affect and machine design: Lessons for the development of autonomous machines", *IBM Systems Journal,* volume 42, number 1, 2003.

[3] Edmund T Rolls. *Emotion Explained* Oxford University Press, 2005.

[4] Jingfu Ye, Yongzhao Zhan, Shunlin Song," Facial expression features extraction based on Gabor wavelet transformation", *IEEE International Conference on Systems, Man and Cybernetics*, volume 3, page(s): 2215-2219, October 2004.

[5] B. Fasel,.,"Automatic Facial Expression Analysis: A Survey", *Pattern Recognition,* pages 259-275, November 2002.

[6] Y Tian, T Kanade, J.F.Cohn," Recognizing action units for facial expression analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* volume 23, issue 2, page(s):97-115, February 2001.

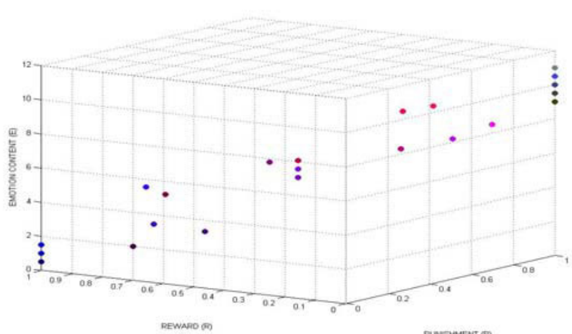


Figure 1: 3D Cognitive Model and the Matrix (beside)

[7] Yuwen Wu, Hong Liu, H. Zha," Modeling Facial expression space for recognition. *IEEE International Conference on Intelligent Robots and Systems*, page(s): 1968-1973, August 2005.

[8] I. Buciu, C. kotropoulos, I. Pitas," ICA and Gabor representation for facial expression recognition", *International Conference on Image Processing*, volume 2, page(s): 855-858, September 2003.

[9] Hima Bindu Maringanti, Ujjwal Mishra, Prashant Manglik, Prashant, " Facial Expression Recognition ",", *IEEE InternationalConference on Systems, Man and Cybernetics*, at Netherlands , UK.

[10] Rohit dhawan, Dhruv Kumar, M.Hima Bindu, "Face Recognition using Neural Networks", *International Conference on Signal Processing and Robotics,ISPR-2004, IIIT, Allahabad.*

[11] C. Cirelo_, Thomas S. Huang : "Learning Bayesian Network Classifiers for Facial Expression Recognition both Labeled and Unlabeled Data" *proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'03) 1063-6919/03 IEEE.

[12] Buciu, C.Kotrropoulus and I.Pitas : "ICA and Gabor Representation for Facial Expression Recognition" ,page 803 – 7750 IEEE 2003.

[13] Jenn-Jier James Lien: "Automatic Recognition of Facial Expressions Using Hidden Markov Models and Estimation of Expression Intensity" , IEEE 1999.

[14] Yongsheng Gao, Maylor K. H. Leung, Siu Cheung Hui, and Mario W. Tananda : "Facial Expression Recognition From Line-Based Caricatures", IEEE 2003.

[15] Yuri Ivano , Masahide Yoneyama and Katsuhiko Shirai : "Recognition of facial expressions using Associative Memory", 1996 *IEEE Digital Signal processing Workshop Proceedings*, P.243-246, (Leon, September,1996).

[16] Berthold K.P.Horn and Brian G.Schunck : "Determining optical flow" , Artificial Intelligence Laboratory, MIT, Cambridge, 1981 pp. 185-203.

[17] Zhengyou Zhang, Lyons M, Schuster M., Akamatsu S., "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron", Proceedings. Third IEEE International Conference on Automatic Face and Gesture Recognition, Page(s):454 - 459, April 1998.

	Reward (R)	Punishment (P)	Emotion (E)	Name of Emotion
1	1	0	0.5	Happy
2	1	0	1.0	Pride
3	1	0	1.5	Enthusiasm
4	0.7	0	2.0	Joy
5	0.6	0.2	2.5	Love
6	0.7	0.1	3.0	Tenderness
7	1	0.5	3.5	Ecstasy
8	0.8	0.3	4.0	Lust
9	0.8	0.8	4.5	Surprise
10	0.5	0.5	5.0	Conformity
11	0.5	0.5	5.5	Boredom
12	0.5	0.5	6.0	Indifference
13	0.3	0.7	6.5	Disgust
14	0.2	0.8	7.0	Fear
15	0.5	1	7.5	Revenge
16	0.4	1	8.0	Rage
17	0	0.7	8.5	Sadness
18	0	1	9.0	Hate
19	0	1	9.5	Grief
20	0	1	10.0	Shame
21	0	1	10.5	Sorrow
22	0	1	11.0	Anger

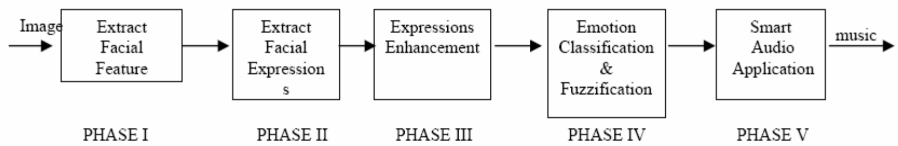


Figure 2. Block diagram for the system architecture

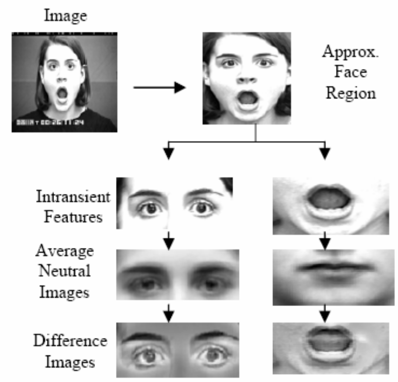


Figure 3: Facial Feature Extraction

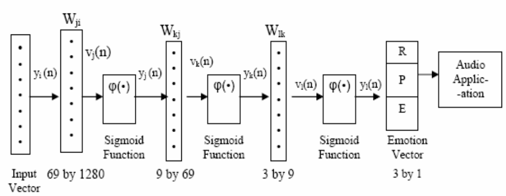


Figure 4. Neural network architecture for emotion pattern learning.

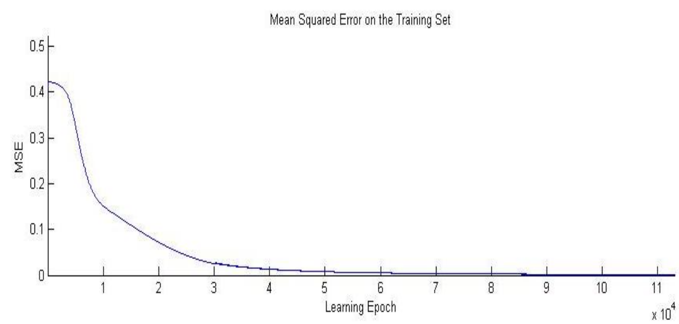


Figure 6: Learning graph for 113150 epochs with error 0.001380

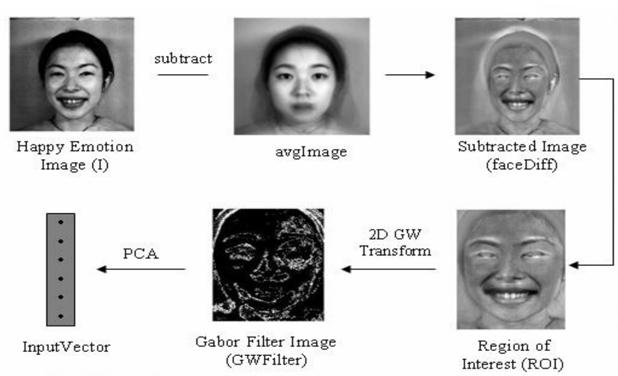
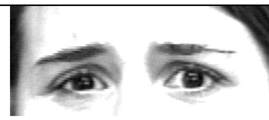
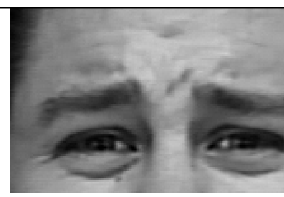
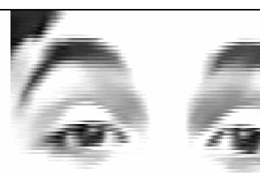
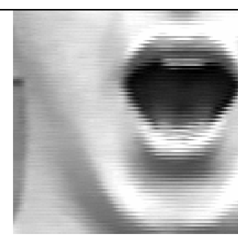


Figure 5: Steps in Image Pre-Processimg

<u>Correct Classification of emotions</u>:

TABLE 1 : RESULTS

EMOTION	IMAGE	EYES	MOUTH	DESIRED	ACTUAL		
ANGER				0.0000 1.0000 0.1000	0.0166 0.9939 0.1335		
FEAR				0.2000 0.8000 0.3286	0.3564 0.6836 0.4110		
DISGUST				0.3000 0.7000 0.2143	0.3344 0.7278 0.1824		
SURPRISE				0.8000 0.8000 0.7857	0.7893 0.8174 0.7467		