A Neural Network Based Classifier for a Segmented Facial Expression Recognition System Based on Haar Wavelet Transform

By Ongalo P. N. Fedha, Huang Dong Jun & Richard Rimiru

Central South University, Changsha, China

Abstract - Automatic recognition of facial expressions is a vital component of natural human-machine interfaces. Facial expressions convey information about one's emotional state and helps regulate our social norms by helping detect and interpret a scene. In this paper, we propose a novel face expression recognition scheme based on Haar discrete wavelet transform and a neural network classifier. First, the sample image undergoes preprocessing where noise is removed using binary image processing techniques. Then feature vectors are extracted using DWT from corresponding pixels in the image. The extracted image pixel data are used as the input to the neural network. We demonstrate experimentally that when wavelet coefficients are fed into a back-propagation based neural network for classification, a high recognition rate can be achieved by using a very small proportion of transform coefficients. Based on our experimental results, the proposed scheme gives satisfactory results.

A NEURAL NETWORK BASED CLASSIFIER FOR A SEGMENTED FACIAL EXPRESSION RECOGNITION SYSTEM BASED ON HAAR WAVELET TRANSFORM

Strictly as per the compliance and regulations of:

A Neural Network Based Classifier for a Segmented Facial Expression Recognition System Based on Haar Wavelet Transform

Ongalo P. N. Fedha [α], Huang Dong Jun [σ] & Richard Rimiru [ρ]

Abstract - Automatic recognition of facial expressions is a vital component of natural human-machine interfaces. Facial expressions convey information about one's emotional state and helps regulate our social norms by helping detect and interpret a scene. In this paper, we propose a novel face expression recognition scheme based on Haar discrete wavelet transform and a neural network classifier. First, the sample image undergoes preprocessing where noise is removed using binary image processing techniques. Then feature vectors are extracted using DWT from corresponding pixels in the image. The extracted image pixel data are used as the input to the neural network. We demonstrate experimentally that when wavelet coefficients are fed into a back-propagation based neural network for classification, a high recognition rate can be achieved by using a very small proportion of transform coefficients. Based on our experimental results, the proposed scheme gives satisfactory results.

Keywords : *Discrete wavelet transform, Neural network classifier, Facial expression recognition.*

I. INTRODUCTION

The human face is undoubtedly the most common characteristic used by humans to recognize and reflect face expressions with speed and accuracy in a passive and non intrusive manner. Facial expressions deals with the classification of facial muscles motion and facial feature deformation into abstract classes that are purely based on visual information while human emotions result from other different factors and their state might or might not be revealed through a number of communication channels such as emotional voice, pose, gestures, gaze direction and facial expressions [1]. Consider a scenario where a person tries downloading one of his favorite movies but becomes frustrated with the system's inability to load the program due to bandwidth or some other reason. In this case the person is likely to express some sort of emotional dissatisfaction well captured via face expressions either intentionally or unintentionally. There are existing works like Chen et al. [2] and De Silva et al.[3] that studied combined detection of facial and vocal expressions of emotion. However, majority of the studies treat the various human communication channels separately.

An automatic recognition of face expression through facial Action Unit (AU) has attracted much attention in the recent years due to its potential applications in behavioral science, medicine, security and human machine interaction. The Facial Action Coding System (FACS) developed by Ekman and Friesen [4] is the most commonly used system for facial behavior analysis. Based on image data, face expressions can be categorized based on whether they are static images or dynamic image sequence. Static images analyze single, still image, based on the spatial information of a frame and the face geometry. It has less computation and is more suitable for real-time facial expression recognition. Dynamic image sequences on the other hand take the motion information of expression images, the expression changes in time and space information together into account, so the recognition rate is very high, but the amount of calculation is also so large. Currently, international standardized face expression classification includes seven classes neutral, anger, happiness, sadness, surprise, disgust and fear [5, 6].

Many researchers have proposed various methods to detect and recognize face expressions. In general face expression representation could be categorized either as holistic, analytic or hybrid methods [1][7]. In holistic approach, the whole face region is taken as data input into the face expression recognition system. Examples of holistic methods are eigenfaces, probabilistic eigenfaces, fisherfaces, support vector machines, nearest feature lines (NFL) and independent-component analysis approaches. In analytic-based approaches, local feature points on face such as the nose, the mouth, the eyes are segmented and then used as input data to the classifier while hybrid separately extracts both local and global features and combines them for recognition. However, this field still remains very challenging especially in real life applications.

Recently, Feng in 2004[8] used Local binary parameters to extract face appearance features. They used a two stage classifier, at the first stage, two expression candidates from initial seven were selected.

Author α σ ρ *: School of Information Science and Engineering, Central South University, Changsha, 410083, China.*
E-mail α *: fedhap@yahoo.com*
E-mail σ *: djhuang@mail.csu.edu.cn*
E-mail ρ *: rimirurm@yahoo.com*

At the second stage, one of the two candidate classes was verified as final expression class. In 2006, Tsai and Jan [9] used subspace model analysis to analyze the data and to recognize facial expressions. They also did some research on facial deformation problems e.g. pose or illumination variations. Nan and Youwei [10] used five classifiers and then used Dempster-Shafer (DS) classifier combination approach. They reportedly achieved a maximum accuracy of 95.7%. Wallhoff et al. in [11] discussed innovative holistic and self organizing approaches for efficient facial expression analysis. Their experiments were based on publicly available FEEDTUM database. They achieved accuracy of 61.67% by using macro motion blocks and SVM-SFFS as feature extraction and feature classification respectively. In 2008, Kotsia et al. [12] did an analysis of the effect of partial occlusion on facial expression recognition, using Gabor wavelets, Discriminant Non-negative Matrix Factorization and a shape-based method as feature extraction techniques. Whitehill et al. [13] explored an idea for recognition of facial expression in relation with intelligent tutoring system. Their idea was to automatically estimate the difficulty level of the lecture as perceived by the student as well as to determine the preferred viewing speed of the student. In 2009, Tai and Huang [14] proposed a method for facial expression recognition in video sequences. They performed noise reduction using median filter and then used a cross-correlation of optical flow and mathematical models from the facial points. Finally the features were fed to an ELMAN neural network for expression classification.

Even though many researchers have used various methods to recognize facial expressions from images and videos, utilizing multi-resolution technique to process image pixel value for recognizing the facial expressions has not been exhausted. One of the most popular multi-resolution analysis technique is the wavelet transform. Wavelet transform can be performed for every scale and translation, resulting in Continuous Wavelet Transform. Wavelet transform can also be performed at multiples of scale and translation intervals, resulting in Discrete Wavelet Transform (DWT). Since CWT provides redundant information and involves more computational effort, normally DWT is preferred [15]. In [16], Haar-like technique has been used to extract the features. Six statistical features namely variance, standard deviation, mean, power, energy and entropy were derived from the approximation coefficients of Haar-like decomposition. These statistical features were used as an input to the neural network for classifying 8 facial expressions.

Preprocessing is regarded as an important step in image processing as it helps improve the image quality by removing noise, highlighting features of interest and separating the object of interest from the background. In this work, we use morphological opening and closing operators to eliminate noise and its effects from the input image before using Haar discrete wavelet transform to extract features for the neural network classifier.

a) Proposed Method

In this paper, a face represents a set of connected regions of similar texture and intensity levels that are combined to form objects. Some of these objects are small and low in contrast while others are large and high in contrast, presenting the need to analyze them using multi-resolution processing. In order to address the curse of dimension, we map the face image model into a low level dimension system that reflects the dynamics of the human facial expression system, then use binary image processing techniques to reject noise introduced by cropping facial images through low resolution disparity calculations that consequently lead to gain of better results. Next the discrete wavelet transform is used to extract features for a neural network classifier.

The rest of the paper is organized as follows: Section 2 gives the image preprocessing, Section 3 gives detailed description on discrete wavelet transform. Section 4 gives the details of a neural network 5 presents the experiment results and analysis. Finally, we conclude in Section 6.

II. Image Pre-Processing

The first step is to acquire images from the sensor or from the database. In our experiment we use static images from JAFEE database. Image preprocessing is a significant step it transforms the image data until regions of interest, better suited for analysis are found. First we crop the facial part of the image in order to remove hair, the neck and other background details that are not central to face expression followed by histogram equalization to enhance the quality of the image. Next we use morphological operator opening and closing to eliminate noise and its effects which may have arisen from image acquisition while distorting the image as little as possible. Morphological processing compares pixels to those pixels surrounding it. They change the shape of particles by processing each pixel based on its number of neighbors and the values of those neighbors. A neighbor is a pixel whose value affects the values of nearby pixels during certain image processing functions. Morphological transformations use a 2D binary mask to define the size and effect of the neighborhood on each pixel, controlling the effect of the binary morphological functions on the shape and the boundary of a particle. Morphological opening defined as the opening of an image X by structuring element B is the erosion of X by B followed by the dilation of the result by B.

$$X \circ B = (X \ominus B) \oplus B \qquad (1)$$

Similarly closing of an image X by structuring element B is dilation followed by erosion.

$$X \bullet B = (X \oplus B) \ominus B \qquad (2)$$

Erosion stage eliminates the noise from the background but increases the size of noise elements (dark spots) contained in the image because they are inner boundaries that increase in size as objects are eroded. The enlargement is countered by performing dilation on the resulting face image. Morphological opening creates some gaps within the image this is

fixed by performing a closing operation on the opening. The results are given in figure1. (d) Closing has the overall effect of smoothening the image and eliminating small holes.

Figure1 (a) shows the original face, Figure1 (b) original cropped image. Figure1 (c) gives the result of opening image in (b) with a structuring element. Figure1 (d) is the result of performing a closing on results of figure1(c), it gives the net result a smoothened image after noise elimination both in the background and in the face image.
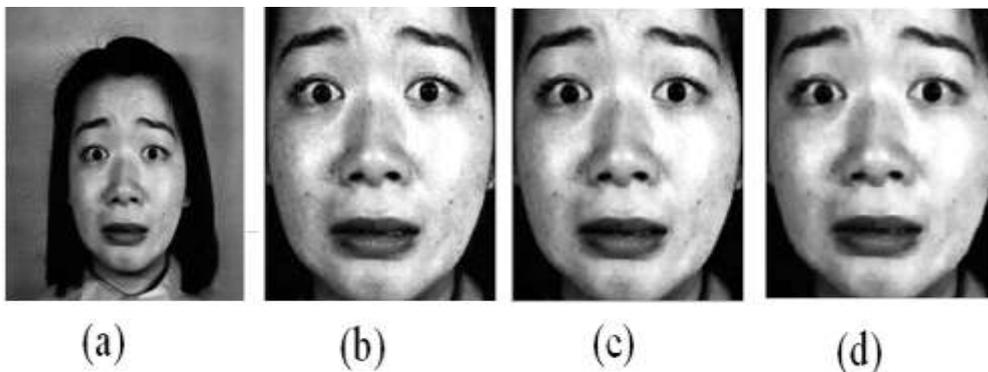

Figure 1. (a) Original test image (b) A cropped test image (c) morphological opened test image (d) morphological closing

III. DISCRETE WAVELET TRANSFORM

Wavelet transform has good time-frequency localization property and better energy concentration performance in addition to benefits of dimension reduction that leads to less computation complexity, multi-resolution data approximation and insensitive feature extraction. Discrete wavelet transform operates by convolving a target function with wavelet kernels, to obtain wavelet coefficients representing the contributions of wavelets in the function at different scales and orientations. DWT can be mathematically expressed by Equation 1 as:

$$DWT_{x(n)} = \begin{cases} d_j,k = \sum x(n) h_j^*(n-2^j k) \\ a_j,k = \sum x(n) g_j^*(n-2^j k) \end{cases} \qquad (3)$$

The coefficients $d_{j,k}$ refer to the detail components in signal $x(n)$ and correspond to the wavelet function, whereas $a_{j,k}$ refer to the approximation components in the signal. The functions $h(n)$ and $g(n)$ in the equation represent the coefficients of the high-pass and low-pass filters respectively, whilst parameters j and k refer to wavelet scale and translation factors.

By applying 2-D DWT to an image we decompose it into four equal sub-bands each a fourth of the original image as shown in figure 2(a). The LL, LH, HL and HH sub-bands corresponding to approximate, horizontal, vertical and diagonal matrices respectively.

The LL corresponds to low frequency information from both the horizontal and vertical directions; LH corresponds to the low frequency components in the horizontal direction and high frequency in the vertical direction. HL corresponds to the high frequency components in the horizontal direction and low frequency in the vertical direction. And HH corresponds to the high frequency components in both the horizontal and the vertical direction. Taking the LL sub-band to represent the image we compress the original image into quarter its dimension, which leads to less computation complexity and recognition time. Further wavelet decomposition on the LL image generates lower dimensional multi resolution facial image. If the decomposed components continue to be decomposed, it forms a pyramid-like wavelet decomposition tree structure that might be beneficial for further analysis.

It is worth noting that the LL sub-band of an image carries the general and most important features of the face which is necessary for recognition while the high frequency carries most detailed information that happen when a person is smiling, sleeping, annoyed It and so on which are key to our face expression recognition. From figure 2(b) it is clear that the eyebrows, eyes and mouth contour in the HL are very clear. To be able to verify the impact of high frequency components on the expression recognition, we extract the low-frequency components LL and add high-frequency components HL, LH and HH. The steps are as follows:

Step 1: Two-level wavelet decomposition of training set with haar wavelet packet. Organize the low-frequency LL into column vector, denoted by X;

Step 2 : Change high-frequency components HL, LH and HH into a column vector, then respectively add them to column vector X to form feature vector X';

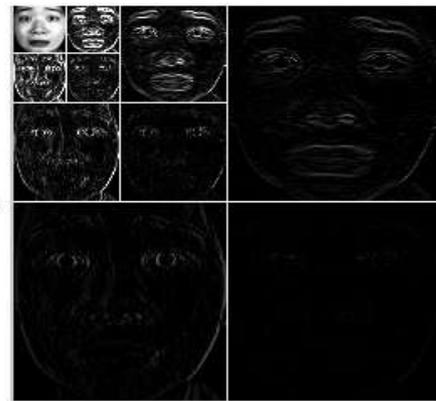**Step 3 :** Feature vectors obtained from Step2 form the input to our neural network for classification.

a(n)	h(n)	h(n-1)	h(n-2)
v(n)	d(n)		
v(n-1)		d(n-1)	
v(n-2)			d(n-2)

Figure 2(b) A processed image

3rd level Haar wavelet decomposition

Figure (2a) 3rd level DWT decomposition structure, n denotes the decomposition scale and a, h, v, d are approximation, horizontal, vertical and diagonal detail coefficients, respectively.

The sub image in the upper left corner is the approximation image that results from final decomposition step surrounded in a clockwise manner by the horizontal, diagonal and vertical detail coefficients that were generated during the same decomposition.

IV. BACK PROPAGATION NEURAL NETWORK (BPNN)

Back propagation method enables the network to learn a predefined set of input-output example pairs by using a two-phase propagate-adapt cycle. First an input pattern is applied as a stimulus to the first layer of network units, which is propagated through each hidden layer until an output is generated. The actual network outputs are subtracted from the desired outputs and an error signal is produced. This error signal is the basis for the back propagation step, whereby the errors are passed back through the network by computing the contribution of each hidden processing layer and deriving the corresponding adjustment needed to produce the correct output. This process repeats, layer by layer, until each node in the network receives an error signal that describes its relative contribution to the total error. Based on the error signal received, connection weights are then updated by each unit to cause the network to converge toward a state that allows all the training patterns to be encoded.

The back propagation network used consists of input layer, hidden and output layer. Before the training process the weights are initialized to small random numbers to ensure that the network is not saturated by large values of the weights and to prevent other training pathologies in addition to making sure that the network learns. The number of neurons in the hidden layer were varied between 6 and 15.In our analysis we found the system to perform well with 10 neurons. The number of neurons in the output layer equals to 6, the number of classes. The stopping criteria used the sum of squared error (SSE) (1.0) and a maximum number of epochs equals to (10000). The basic procedure for training back propagation network is embodied in the description given below [17].

Algorithm :

i. Initialize the network weights and biases
ii. Select the training pair from the training set apply input vector to the network input
iii. Sum the weighted inputs and apply activation function to compute output signal.

$$X_{pj}^{h} = \sum_{i=1}^{N} w_{ji}^{h} x_{pi} + b_{j}^{h} \Rightarrow i_{pj=} f_{j}^{h}(X_{pj}^{h}) \quad (4)$$

where w_{ji}^{h} is the weight connection from the ith input unit and b_{j}^{h} is the bias. $'h'$ superscript refers to quantities of the hidden layer.

iv.   Calculate the output of the network

$$y^o_{pk} = \sum_{j=1}^{L} w^o_{kj} i_{pj} + b^o_k \Rightarrow \mathbf{o_{pk}} = \mathbf{f^o_k(y^o_{pk})} \quad (5)$$

where superscript 'o′' refers to quantities at the output layer.

v.   Calculate the error terms for the output units:

$$\delta^o_{pk} = (\mathbf{y_{pk}} - \mathbf{o_{pk}}) \mathbf{f^{o'}_k(y^0_{pk})} \quad (6)$$

y_{pk} is the desired output value, and o_{pk} is the actual output Followed by the error terms for the hidden units:

$$\delta^h_{pk} = \mathbf{f^{h'}_j(X^h_{pj})} \sum \delta^o_{pk} w^o_{kj} \quad (7)$$

Update weights on the output layer:

$$w''_{ji}(t+1) = w^h_{ji}t + \eta \delta^n_{pj} x_i \quad (8)$$

Update the weights on the hidden layer:

$$w''_{ji}(t+1) = w^h_{ji}t + \eta \delta^n_{pj} x_i \quad (9)$$

Repeat the above steps with all the training vectors until the error for all vectors in the training set is reduced to an acceptable value.
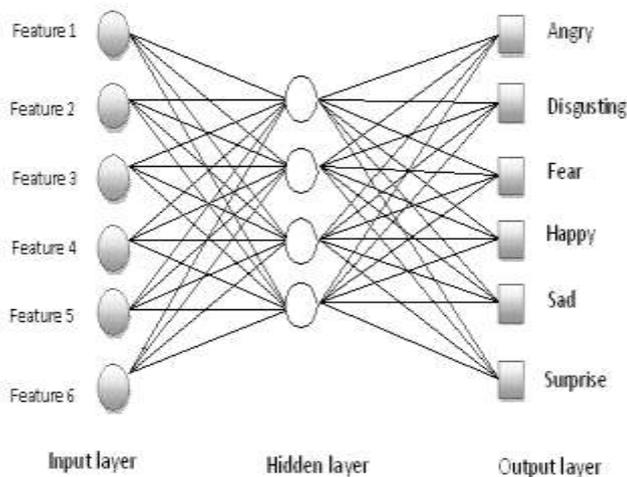

Figure 3. Illustration of the neural network training

V.   EXPERIMENT RESULTS AND ANALYSIS

To assess the validity and efficiency of our approach, the experiments are conducted on the Japanese Female Facial Expression (JAFFE) database, which contains 213 images of 7 facial expressions posed by 10 Japanese female models. Ten expressers posed 3 or 4 examples of each of six facial basic expressions (happiness, sadness, surprise, anger, disgust, and fear) and a neutral expression. Images

from the JAFFE database are cropped to eliminate most of the background and some parts of the hair and chin. The sizes of the images are 256×256 in tiff format. Figure 4 shows examples of the original images in the JAFFE database. In the real-world environments, the rotation of the camera axis and the head pose variations often exist. JAFFE database presented includes these images with minor rotation of camera axis and variations in head poses. As a result, the robustness of the proposed method is evaluated.

The images depicting six different facial expressions: fear, disgusting, happiness, sadness, anger and surprise were used. The feature vectors are extracted from second level Haar discrete wavelet decomposition of the corresponding image. Beyond this level we noted that the size of images become unreasonably small and no valuable information could be extracted. The extracted image data are divided into testing and training data. In training phase two feature vectors per class were used and in the testing phase the remaining feature vector from each class was used. The images used in the testing set were not included in the training set. We use six binary NN classifiers, the data is divided into six blocks according to six expression classes, and each classifier is trained for a particular expression using one-against-all approach. The output of these binary classifiers gives the probabilities to which extent the input image belongs or does not belong to the class for which the particular classifier has been trained. After the training, we use the output generated as a feature map that provides an indication of the presence or absence of many face expression feature combinations at the input. We repeated the training procedure 30 times and got average results for the trials, with a percentage accuracy of 81%.
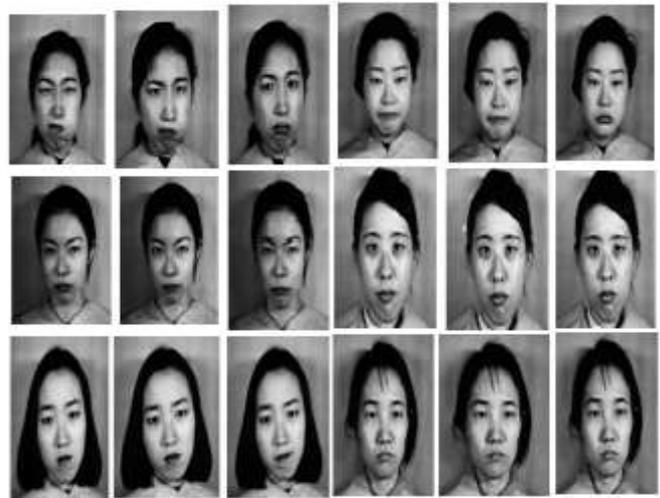

Figure 4: A sample of angry faces from JAFEE database

The results are illustrated using the confusions matrix table 1. The row corresponds to the six face expressions while the columns correspond to the

predicted emotions. The diagonals entries show the percentage of correct classification for each class and off diagonal entries show misclassification. Accuracy is computed as follows

$$\%accuracy = \frac{number\ of\ correct\ classification\ classification}{number\ of\ all\ classification} \times 100$$

Table 1. Confusion matrix of six basic face expressions

	angry	disgusting	fear	happy	sad	surprise	unclassified
angry	100						
disgusting		100					
fear			57				43
happy				100			
sad					27		73
surprise						100	

VI. Conclusion

In this work the high accuracy recognition system based on machine learning with a reasonable number of samples are introduced. First the input image is preprocessed, morphological operators are used to remove noise and to smoothen it the resulting binary data is used as input to discrete wavelet transform. Second level Haar wavelet is computed and the resulting feature vectors are used as input to the back propagation neural network classifier. Experiments for evaluation were carried out on JAFEE database presenting the six facial expressions, 'angry', 'disgusting', 'fear', 'happy', 'sad', 'surprise' and the results have shown that the proposed method can perform at 81% accuracy. The low results arise from unclassified data of fear class (43%)and sad class (73%). In these analysis the simplicity and robustness of the system is significant. For our future work, we plan to look into the facial expression recognition of subjects in real-time videos.

VII. Acknowledgments

References Références Referencias

1. B. Fasel and J.Luettin, "Automatic facial expression analysis": a survey, *Pattern Recognition*, 2003, 36(1): 259∼ 275
2. L. Chen, H. Tao, T. Huang, T. Miyasato, and R. Nakatsu, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 83–88, 1998.
3. L. C. De Silva and P. C. Ng, "Bimodal emotion recognition," in Proc.4th IEEE Int. Conf. on Automatic Face and Gesture Recognition,France, Mar. 2000, pp. 332–335.
4. Ekman, P. and Friesen W, "Facial action coding system: a technique for the measurement of facial movement", Palo Alto: Consulting Psychologists Press, 1978.
5. S. Dongcheng and JIANG Jieqing,"The Method of Facial Expression Recognition Based on DWT-PCA/LDA", 2010 3rd International Congress on Image and Signal Processing.
6. WANG Zhiliang, LIU Fang,WANG Li. "Survey of Facial Expression Recognition Based on Computer Vision[J]", Computer Engineering, 2006, 32(11),231∼33.
7. Elham Bagherian, Rahmita Wirza O.K. Rahmat*, "Facial feature extraction for face recognition: a review",Information Technology, 2008. ITSim 2008. International Symposium* on 09/*2008*; 2:*1-9.*
8. Xiaoyi Feng, "Facial expression recognition based on local binary patterns and coarse-to-fine classification", in Proceedings of the The Fourth International Conference on Computer and Information Technology (CIT'04). Pages: 178 – 183. Year: 2004.
9. Tsai, P.H. Jan, T, "Expression-invariant face recognition system using subspace model analysis", in IEEE International Conference on Systems, Man and Cybernetics, 2005.
10. Zhang Nan, Zhang Youwei, "Inducement analysis in facial expression recognition". In The 8th International Conference on Signal Processing 2006, Volume III. Year: 2006.
11. F. Wallhoff, B. Schuller, M. Hawellek and G. Rigoll, "Efficient recognition of authentic dynamic facial expressions on FEEDTUM database". In IEEE Conference on Multimedia and Expo (ICME'06), Pages: 493-496, Year: 2006.

12. Irene Kotsia, Ioan Buciu, Ioannis Pitas, "An analysis of facial expression recognition under partial facial image occlusion", In transanctions on Image and Vision Computing, Volume 26, Issue 7, Pages: 1052-1067, Year 2008.

13. Whitehill, J.; Bartlett, M.; Movellan, J., "Automatic facial expression recognition for intelligent tutoring systems", In proceedings of IEEE computer scociety workshop on Computer Vision and Pattern Recognition, Pages: 1-6, Year 2008.

14. Shenchuan Tai, Hungfu Huang, "Facial expression recognition in video sequences", In Proceedings of the 6th International Symposium on Neural Networks: Advances in Neural Networks - Part III, Pages 1026- 1033, Year 2009.

15. Murugappan, M., Rizon, M., Nagarajan, R., and Yaacob, S., FCM clustering of human emotions using wavelet based features from EEG. Trans. Biomed. Soft Comput. Hum. Sci. IJBSCHS 14(2):35–40, 2009.

16. Satiyan, M., and Nagarajan, R., Recognition of facial expression using Haar-like feature extraction method. Proceedings of the 3rd IEEE International Conference on Intelligent and Advanced Systems (ICIAS), Kuala Lumpur, Malaysia, pages 1-4, Year 2010.

17. James A. Freeman. David M. Sukapra." Neural Networks Algorithms, Applications and Programming Techniques", Addison-Wesley Publishing Company, New Yolk.1991.

14

This page is intentionally left blank