Facial Expression Recognition using Hybrid Transform

Anmar. A. Razzak, PhD
Professor Assistant at the University of AL-Mustansyria College of Education /Computer Department, Baghdad, Iraq

Ahmed Rafid Hashem
Postgraduate student at the University of AL-Mustansyria College of Education /Computer Department, Baghdad, Iraq

ABSTRACT
Automatic analysis of facial expressions is rapidly becoming an area of intense interest in computer vision and artificial intelligence research communities. In this paper an approach is presented for facial expression recognition of the six basic prototype expressions (i.e., happy, surprise, anger, sadness, fear, and disgust) based on Facial Action Coding System (FACS). The approach utilizes the hybrid transform in which consists of two transforms; the Wavelet transform and the Discrete Cosine Transform (DCT). The approach suggested includes many steps such as preprocessing, feature extraction, clustering and recognition. In feature extraction phase the Wavelet transform and the Discrete Cosine Transform (DCT) were implemented, in the clustering phase the Self Organizing Feature Map produced by Kohonen was implanted. Topological ordering patterns produced by Kohonen Self Organizing Map, in which implemented on feature extracted for each prototype facial expression was used to classify the six basic expressions. The map will compute the topological relationship between the particular expressions featured. While in recognition phase Euclidean distance measure had been used. The method tested using FACS-Coded expressions database of basic emotions: "Cohn-Kanade Database". An average recognition rate of 92.2% was achieved for six basic expressions.

Keywords
DCT, wavelet, facial expression recognition, SOM, Euclidean distance.

1. INTRODUCTION
Human beings are capable of communicating with each other in many different ways. The most common methods exploited for this include the use of words, gestures and facial expressions, either individually or in some combination. Using facial expressions has to be one of the most complicated forms, if not the most complicated form, much can be communicated by even a single facial expression, and hence these expressions have become a very important aspect of our communication. Consequently, it is only natural that the demand for a computer system that can recognize human facial expressions has arisen.

It is argued that to truly achieve effective human-computer intelligent interaction (HCII), there is a need for the computer to be able to interact naturally with the user, similar to the way human-human interaction takes place. Humans interact with each other mainly through speech, but also through body gestures, to emphasize a certain part of the speech and display of emotions. One of the important ways humans display emotions is through facial expressions. Automatic facial expression analysis is a complex task as the topology of faces varies from one individual to another quite considerably due to different age, ethnicity, gender, facial hair, cosmetic and occluding objects such as glasses and hair. Further, faces appear disparate because of pose and lighting changes. Variation such as these have to be addressed at different stages of an automatic facial expression analysis system, such as normalization task including (pose and illumination), and face segmentation task including (background and facial feature separation).

Automatic analysis of facial expressions is rapidly becoming an area of intense interest in computer vision and artificial intelligence research communities. Automated systems that sense, process, and interpret human facial expressions have important commercial potential; they seem to have a natural place in commercial products such as computer systems for video conferencing, video telephony, video surveillance, video indexing ,robotics as well as virtual reality, image understanding [1],, psychological studies, facial nerve grading in medicine, face image compression and synthetic face animation .

Facial expression intensities may measured by determining either geometric deformations of facial features or the density of wrinkles appearing in certain face regions.

There is one main methodological approach of how to measure the previously mentioned characteristics of facial expressions, this is the FACS (Facial Action Coding System), which was developed by (Ekman and Friesen [2]) and has been considered as a foundation for describing facial expressions.

There are many studies that interest in facial expression recognition can show this survey of studies in [3]. The approaches to facial expression recognition divided into two classes, which are geometrical feature-based approaches and appearance – based approaches [4]. The geometrical features, features which existing the shape and location of facial components such as eye, eyebrow, nose, canthus, and mouth , etc. There are many of studies work on geometry like , Suwa et al[5] , Pantic and Rothkrank[6], Edwards et al[7], Zeng et al[8], Bartlett et al[9] , Wan et al [10], Tian et al [11], Black and Yacoob[12], Littlework et al[13]. With regard to as for the appearance –based approaches , the whole face or specific regions in a face are used for the feature extraction by using optical flow (Yacoob and Davis)[14] and there are anther studies like used PCA (Feng et al)[15] ,ICA, LDA (Jun et al) [16] , Gabor filter (Zhang Z) [17], and other studies [2],[18],[19] .

In this paper an approach is presented for facial expression recognition of the six basic prototype expressions (i.e., happy, surprise, anger, sadness, fear, and disgust) based on Facial

Action Coding System (FACS). The approach utilizes the Wavelet transform and the Discrete Cosines Transform (DCT). The method tested using a database publicly available, FACS-Coded expressions of basic emotions: "Cohn-Kanade [AU-Coded] Facial Expression DataBase" .

2. PREPROCESSING STEPS

The method tested using FACS-Coded expressions database of basic emotions: "Cohn-Kanade Database".Where 123 images had been used with mixed ethnicity, age and gender.

2.1 Image Cropping

In this stage cropping face image only to get rid of image background to increase the efficiency of the facial expression.

2.2 Images subtraction

The subtraction is performed on two images, the normal expression image and the peaked expression image. This process had been implemented for all the images sequences for all the six basic expressions. The subtraction done by subtracting the normal image from the peak image. This process was important to obtain only the motion effect for each expression by moving from normal expression to peak expression, in which reflect only the moving parts of the face for any particular expression and neutralized all other parts of the image in which does not moved (see Figure 1).

Figure (1) subtract operation for happy facial expression.

2.3 Image sharpening

Image sharpening is a powerful tool for emphasizing texture and draw viewer focused and emphasize edge in image. The sharpening is applied on the result of images subtract to show more details especially the edges. (See Figure 2)

Figure (2) sharpening effect

2.4 Histogram Equalization

Histogram equalization technique equalizing the level of intensities such that the output image contains a uniform distribution of intensities, its applied on the sharpened images of all the images used in facial expression recognition process (see Figure 3).

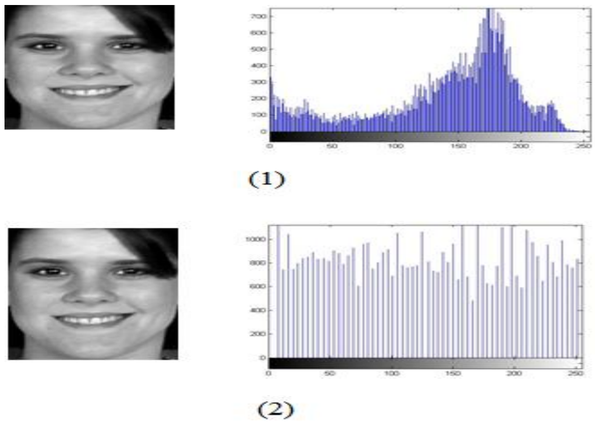


Figure (3)
(1) Shows the image before using histogram equalization (HE) technique. (2) Shows the image after using the histogram equalization technique.

2.5 Images Resize

After the completion of the preprocessing stage, an image resize operation had been performed on all the images to be of equal size. The new resized image was (128*128) pixel. These equal size images will be entered into the next phase, the feature extraction phase.

3. FEATURES EXTRACTION STEPS

Features extraction phase considered as the most important phase in the proposed system, a used hybrid transform in which consist of Wavelet and DCT.

3.1 Wavelet Transform

Wavelet is a waveform of effectively limited duration that has an average value of zero. In some time we need to know frequency and temporal information at the same time, so wavelet can present this benefit. The wavelet transform has many properties such as multiresoultion, locality, sparsity and decorrelation [20] . The wavelet transform is divided into two classes, Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) [21]. In this paper (Haar filter) had been used to get rid of redundant of CWT, the 2D-DWT decompose the signal as following [22] (see Figure 4)

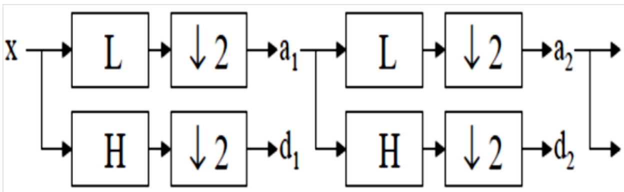
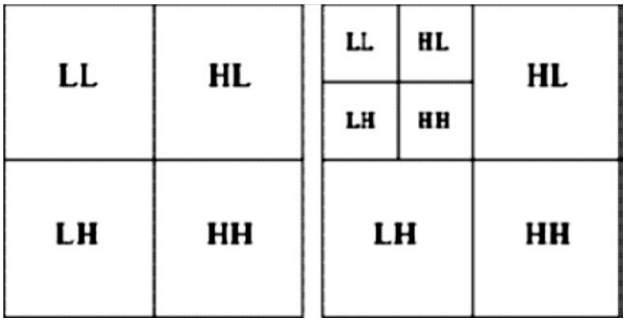


Figure (4) DWT tree.

H and L denote high and low-pass filters respectively, **↓ 2** denotes sub sampling. In wavelet can make decompose in wavelet using DWT for many times (see Figure 5, Figure 6)



(1) (2)

Figure (5) Wavelet decomposition: (1) Single level decomposition (2) Two levels decomposition.

Where LL contain low frequency that contain energy, where HL, LH, HH contain high frequency.

(1) (2) (3)

Figure (6) Wavelet decomposition for image: (1) original image (2) Single level decomposition (3) Two levels decomposition

3.2 Discrete Cosine Transform (DCT)

The DCT is regarded as a discrete-time version of the Fourier-cosine series. Therefore, it is considered as a Fourier-related transform similar to the Discrete Fourier Transform (DFT), using only real numbers. Since DCT is real-valued, it provides a better approximation of a signal with fewer coefficients. The DCT possess some fine properties such as de-correlation ,energy compaction , separability ,symmetry , compression , and orthogonality .in this paper 2D-DCT transform had been used because the work done on image

$$C(u,v) = \alpha(u)\,\alpha(v) \sum_{x=0}^{n-1} \sum_{y=0}^{n-1} f(x,y) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right]$$

M*N dimension. In general the DCT work efficient on image with 8*8 dimension, the 2D-DCT of an M*N image is defined as following.

for u ,v = 0,1,2,…,N −1 and α(u) and α(v) are defined as:

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } u = 0 \\ \quad \vdots & \\ \sqrt{\frac{2}{N}} & \text{for } u \neq 0 \end{cases}$$

$$C(u=0) = \sqrt{\frac{1}{N}} \sum_{x=0}^{N-1} f(x)$$

When apply 2D-DCT on 8*8 images the result coefficient matrix is M*N matrix. The coefficient matrix 8*8 contain low frequency at upper left corner (one pixel) called DC, and other 63 pixel called AC, the DC pixel contain the most energy so the DC is important than 63 pixel AC[23] (see Figure 7).

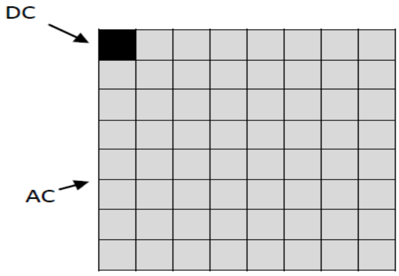


Figure (7) 2D-DCT (8*8) coefficient matrix.

4. CLUSTERING STATE

When solving a pattern recognition problem, the ultimate objective is to design a recognition system which will classify unknown patterns with the lowest possible probability of misrecognition. It is well known that, the complexity of a classifier grows rapidly with the number of dimensions of the pattern space. Thus, the problem is how to face the dimensionality reduction? An efficient way of reducing dimensionality and clustering expression information is to compute the topological relationship between these feature vectors. Thus, Self Organizing Feature Map neural network is used, the dimensionality is reduced to two dimensions instead of 256. Also the discriminating information regarding the face expression is presented on the Self Organizing Feature map in terms of the topological relationship.

Thus, SOM are responsible for visualizing low-dimensional views of high dimensional data, having similar character to multidimensional scaling. SOM learn to recognize groups of similar input vectors in such a way that neurons physically near each other in the neuron layer respond to similar input vectors. One of the most interesting aspects of SOM is that they learn to classify data without supervision [24]. During the learning the neurons having weight closest with the input vector declare as winner [25]. Based on winning neuron weights of all neighborhood neurons are adjusted by an amount inversely proportional to the Euclidean distance. The learning algorithm is summarized as follows:

1-Initialization: Choose random values for the initial weight vectors $w_{j(0)}$, the weight vector being different for j = 1, 2,...,l where l is the total number of neurons.

$Wi=[Wi1, Wi2, \ldots, Wil]^T \in R^n$

2. Sampling: Draw a sample x from the input space with a certain probability.

$X = [x1, x2, \ldots, xl]^T \in R^n$

3. Similarity Matching: Find the best matching (winning) neuron i(x) at time t, $0 < t \leq n$ by using the minimum distance Euclidean criterion:

$i(x) = \arg\min_j \| x(n) - w_j \|, \; j = 1,2,\ldots,l$

4. Updating: Adjust the synaptic weight vector of all neurons by using the update formula:

$W_j(n+1) = W_j(n) + \eta(n) h_{j,i(x)}(n)(X(n) - W_j(n))$

Where η (n) is learning rate parameter, and $h_{j,\ ix}(n)$ is the neighborhood function centered around the winning neuron. Both η (n) and are $h_{j,\ ix}(n)$ varied dynamically during learning for best results.

5. Continue with step 2 until no noticeable changes in the feature map are observed.

5. EXPERIMENTS AND RESULTS

The experiment is performed as follows:

Step1: Preprocessing, Features Extraction And Clustering Part.

1- In preprocessing steps, had been applied the cropping image, subtracting images, sharpening images , histogram equalization and image resize on original image and get image which size (128 *128) .

2- The final image from preprocessing is used with size (128*128), then DWT transform is applied on the images for one level of decomposition to get four parts (LL , HL ,LH , HH) and the energy part is used only (LL that size (64*64) pixel) .

3- After apply the first transform of the hybrid transform (wavelet), then the second transform (DCT) is implemented on the result of wavelet transform the LL part, the DCT used to compact the energy and to compress the features extracted without losing the features characteristics.

4- The SOFM is applied to the final features extracted from step 3. The clustering of the features extracted, such that similar features are grouped together and dissimilar features are grouped into other clusters. These clusters on the map are distinguished to represent the six basic facial expressions (happy, sad, disgust, fear, surprise, anger). The approach suggested is implemented on 123 facial expression images taken from Cohn-Kande database for six facial expressions. Table (1) shows the number of images taken for each facial expression for the six basic facial expressions.

Anger	Disgust	Fear	Happy	Sad	Surprise	Total
20	20	20	23	20	20	123

Table (1) Number of images for each facial expression that used

The Table (2) show the SOFM parameters used.

Number Of Patterns	Map size	Accepted Error	Number of iteration
123	21*21	1.E-5	100000

Table (2) The SOM parameters

The clustering efficiency using SOFM can be calculated by using the following equation:

$$\frac{\boldsymbol{no.of\ expression\ node - no.of\ non\ cluster\ node}}{\boldsymbol{no.of\ cluster\ node}}$$

The efficiency obtained for 123 images for six basic expressions was **91.869%**. Figures (8,9) show the distribution of the images' features extracted regarding their similarities and dissimilarities grouped as clusters on the map of 21x21 elements.

Step2: Recognition Part

1- Preprocessing And Features Extracted For Test Data.

In recognition step 45 new test images had been used that representing the six basic expressions, these new images were subjected to preprocessing and features extraction phases as mentioned before, and then the features extracted were saved for the next step.

2- Euclidean distance for recognizing the facial expression.

3- The Euclidean distance is a method used to obtain the distance between two vectors, (see equation 1, where x and y are two vectors of the same length n) [26].

$$d\ x, y = \sqrt{\sum_{j=1}^{j} \left(x_j - y_j \right)^2} \qquad \text{----------- (1)}$$

Euclidean distance had been used to measure the distance between two vectors, the test images features vectors 45 images for the six basic facial expression and the actual 123 feature vectors also for the six basic facial expression images obtained in step 1.

So that Euclidean distance used to find the similarity between the features vectors that want to be recognized (45 images) and others features vectors in which are already known(123 images). The test result shows a recognition rate of 92.2%.

The recognition rate for each expression is shown in the Table (3) bellow.

Anger	Disgust	Fear	Happy	Sad	Surprise	Total
80%	100%	85.7%	100%	87.5%	100%	92.2%

Table (3) Recognition rate for each expression

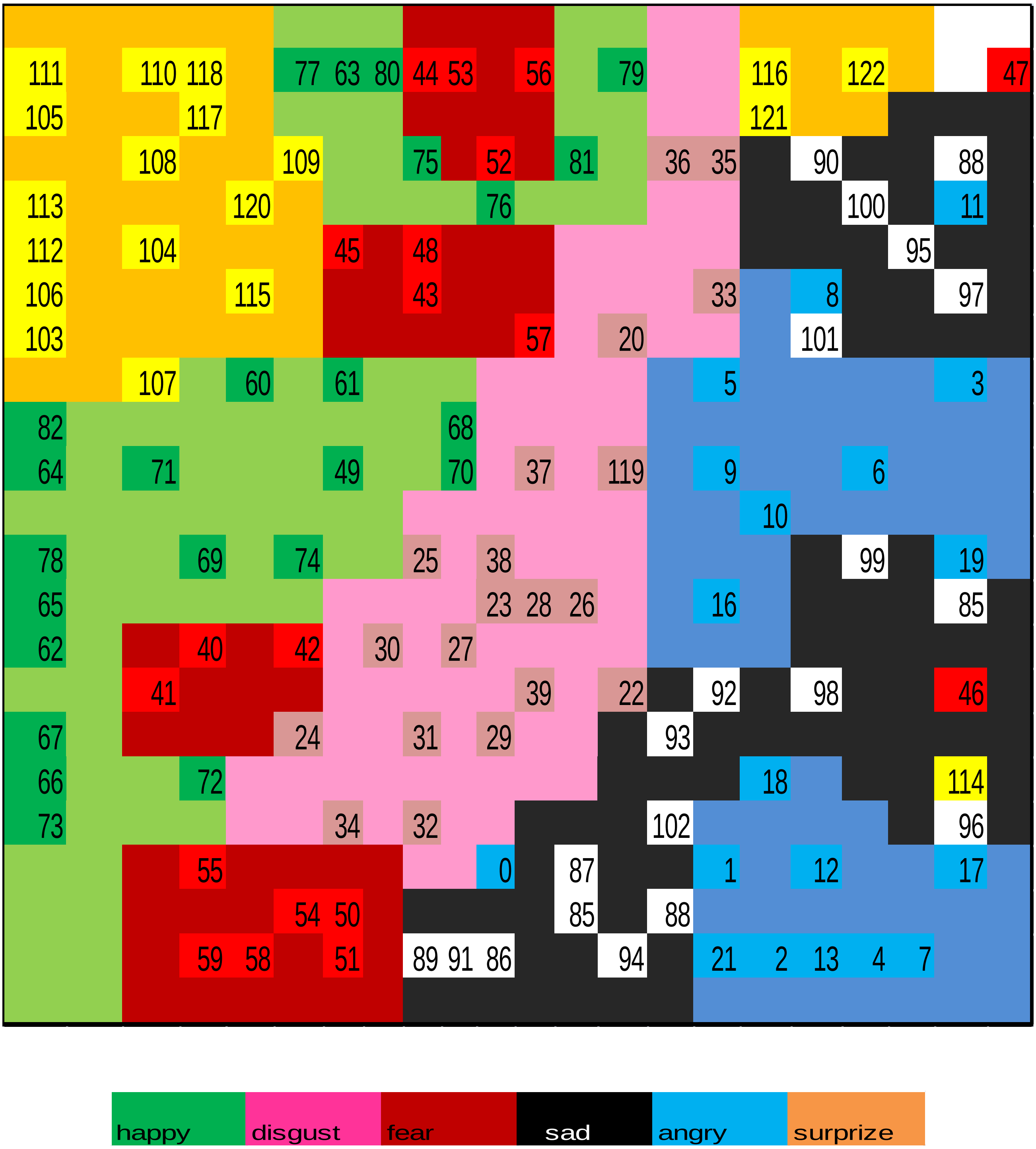


Figure (8) shows the topological relationships between the feature coefficient vectors.

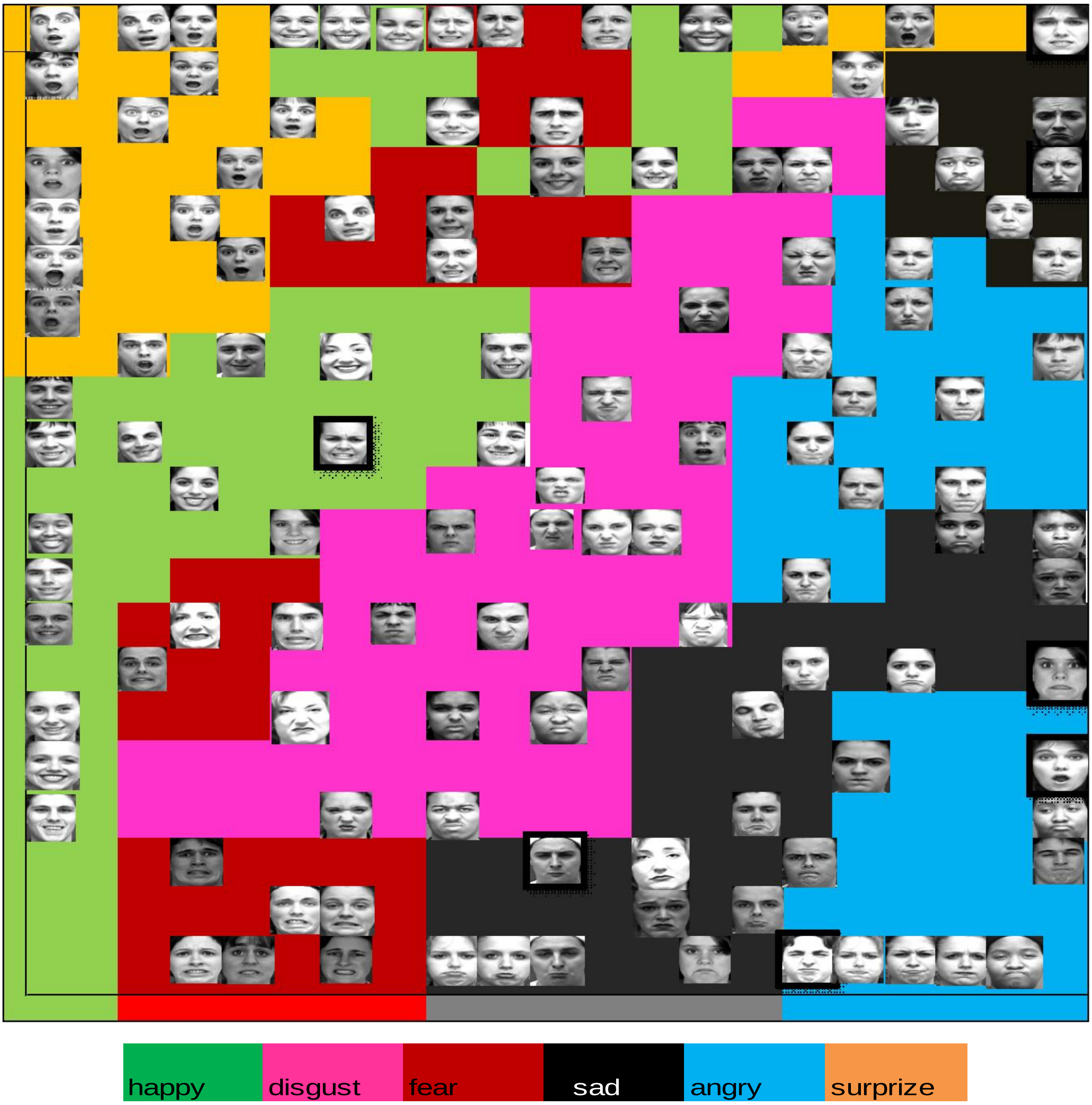


Figure (9) the topological relationships between the actual facial expression images corresponding to their feature coefficient vectors extracted by the hybrid and distributed by SOFM

6. CONCLUSIONS

This paper addressed the problem of automatic facial expression recognition based on FACS. A computer vision system is developed that automatically recognizes the six basic prototypes facial expressions. An approach is presented where wavelet and DCT transforms are used to analyzing gray scales image in order to compute and capture the facial expression features.

The topological ordering (topological relationships) between the facial feature extracted for each basic prototype expression were represented on the SOFM map. To recognize the basic expressions, Euclidean distance used to recognize the test images of six basic facial expressions.

Facial expressions of different types, intensities and durations have been tested. The result shows that the system has a very good accuracy in facial expression recognition, an average recognition rate of 92.2% was achieved for six basic expressions. This system can be implemented without performing image normalization such as rotation, scaling and translation in which considered manually tedious work before the facial feature extraction step took place. This research is of interest in computer vision, since a new model for human facial action in video stream can be found. In future work on preprocessing phase more effective techniques are needed to be implemented to improve the performance of feature detection process. Also for more efficient classification and recognition, learning vector quantization methods can be suggested for further studies.

7. REFERENCES

[1] Vinay Bettadapura ," Face Expression Recognition and Analysis: The State of the Art", College of Computing, Georgia Institute of Technology.

[2] P. Ekman and W. V. Friesen. Constants Across Cultures in the Face and Emotion. Journal of Personality and Social Psychology, 17(2):124-129, 1971.

[3] B.Fasel ,Juergen Luettin ,"Automatic Facial Expression analysis: a survey",pattern recognition ,2003(36).

[4] Y. l. Tian, T. Kanade, and J. F. Cohn, "Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity," in Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 229-234, 2002.

[5] Kaimin Yu ," TOWARDS REALISTIC FACIAL EXPRESSION RECOGNITION" , School of Information Technologies at The University of Sydney, Thesis, PHD, October 2013.

[6] M. Pantic and L. Rothkrantz, "Expert System for Automatic Analysis of Facial Expression", Image and Vision Computing J., Vol. 18, No. 11, p. 881-905, 2000.

[7] H. Rowley, S. Baluja, T. Kanade, "Neural Network-Based Face Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, p. 23 – 38,1998.

[8] Z. Zeng, Y. Fu, G. I. Roisman, Z. Wen, Y. Hu and T. S. Huang, "Spontaneous Emotional Facial Expression Detection", Journal of Multimedia, Vol. 1, No. 5, p. 1-8, 2006.

[9] Bartlett, M. S., G. Littlewort, I. Fasel, J. R. Movellan 'Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction' IEEE Workshop on Face Processing in Video, Washington, 2004.

[10] Kwok-Wai Wan, Kin-Man Lam,Kit-Chong Ng,"An accurate active shape model for facial feature extraction", Pattern Recognition Letters ,vol.26,pp. 2409–2423,May 2005.

[11] Y. Tian, T. Kanade, and J. Cohn" Recognizing action units for facial expression analysis" IEEE Transaction on Pattern Analysis and Machine Intelligence, 2001.

[12] M.J. Black and Y. Yacoob, "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion," lnt'l J. Computer Vision, Vol. 25, No. 1, p. 23-48, 1997.

[13] G.C. Littlewort, M.S. Bartlett, J. Chenu, I. Fasel, T. Kanda, H. Ishiguro, J.R. Movellan,"Towards social robots: Automatic evaluation of human-robot interaction by face detection and expression classification", Advances in Neural Information Processing Systems, Vol 16, p. 1563-1570, 2004.

[14] Yacoob Y., L. Davis. 'Computing spatio-temporal representation of human faces' In CVPR, pages 70–75, Seattle, WA, June 1994.

[15] Feng, G. C., P. C. Yuen, D. Q. Dai, "Human face recognition using PCA on wavelet subband", Journal of Electronic Imaging -- April 2000 -- Volume 9, Issue 2, pp. 226-233, 2000.

[16] Jun, S., Z. Qing, W. Wenyuan, "A improved facial recognition system method", ISBN: 978-3-540-41180-2, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, vol.1948, pp. 212-221, 2000.

[17] Zhang, Z., M. Lyons. M. Schuster, S. Akamatsu, 'Comparison between geometry based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron', in Proc. IEEE 3rd Int'l Conf. on Automatic Face and Gesture Recognition, Nara, Japan, April 1998.

[18] Stathopoulou, I. O., G. A. Tsihrintzis, 'An improved neural network-based face detection and facial expression classification system', Proceedings of IEEE SMC, pp. 666–671, 2004.

[19] Cohen, I., N. Sebe, A. Garg, M. S.Lew, T. S. Huang 'Facial expression recognition from video sequences' Computer Vision and Image Understanding, Volume 91, pp 160 – 187 ISSN: 1077-3142 2003.

[20] Dao-Qing Dai and Hong Yan, "Wavelets and Face Recognition ",*Sun Yat-Sen (Zhongshan) University and City University of Hong KongChina.*

[21] Wavelet transform. From Wikipedia: http://en.wikipedia.org/wiki/wavelet transform

[22] S. Mallat: "A wavelet Tour of Signal Processing", Academic Press, San Diego 1998.

[23] W. B. Pennebaker and J. L. Mitchell, "JPEG – Still Image Data Compression Standard," Newyork: International Thomsan Publishing, 1993.

[24] JAWAD NAGI," Pattern Recognition Of Simple Shapes In A Matlab / Simulink Environment: Design And Development Of An Efficient High-Speed Face Recognition System", College Of Engineering University Tenaga Nasional , Thesis, M.Sc, 2007,pp (97).

[25] Keerti Keshav Kanchi," Facial Expression Recognition using Image Processing and Neural Network", International Journal of Computer Science & Engineering Technology (IJCSET), ISSN: 2229-3345 Vol. 4 No. 05 May 2013.

[26] Euclidean distance. From Wikipedia: http://en.wikipedia.org/wiki/Euclidean_distance.