Automatic Facial Expression Recognition System Based on Geometric and Appearance Features

Aliaa A. A. Youssif

Computer engineering department, College of engineering, Arab Academy for Science

Technology & Maritime Transport(AASTMT), Cairo, Egypt

(On leave from Faculty of Computers & Information, Helwan University, Helwan, Egypt.)

E-mail: aliaay@helwan.edu.eg

Wesam A. A. Asker

Arab Academy for Science, Technology & Maritime Transport (AASTMT)

College of engineering Computer engineering department, Cairo, Egypt

E-mail: wesamaaa@gmail.com

Abstract

This paper presents a computer vision system for automatic facial expression recognition (AFER). The robust AFER system can be applied in many areas such as emotion science, clinical psychology and pain assessment it includes facial feature extraction and pattern recognition phases that discriminates among different facial expressions. In feature extraction phase a combination between holistic and analytic approaches is presented to extract 83 facial expression features. Expression recognition is performed by using radial basis function based artificial neural network to recognize the six basic emotions (anger, fear, disgust, joy, surprise, sadness). The experimental results show that 96% recognition rate can be achieved when applying the proposed system on person-dependent database and 93.5% when applying on person-independent one.

Keywords: Human computer interaction, Facial expression recognition, Utilizing facial features, Shape and appearance modeling, Expression recognition, Emotion analysis

1. Introduction

The interface and interaction between humans and computers has received much attention nowadays with a target focused on development a natural human interaction with the computers based on the normal human to human behaviour interaction. It is well recognized that facial expressions are the most expressive indicators that humans display emotions. In a natural interpretation, the humans can easily detect the expressions the indication of faces and facial expressions. However, in the world of computers and recognition, it is not easy to develop an automatic system to recognize the indication of faces and facial expressions. This is mainly duo to many difficulties such as: i) face detection and segmentation in the captured image, ii) extracting the facial expression information, and iii) the process of classification of the face expression in emotional state.

Therefore, the real challenge is to develop a real time system to solve these difficulties with accurate operations to perform human-like interaction between the man and the machine. Based on the work published (Pantic & Rothkrantz, 2000), the essential role in interpersonal communication is usually clarified by the terms "face-to-face" and "interface". In human behavior, the face is the main element to identify other members of the species, it has the means of lip-reading to interpret what has been said, and gives indication of the emotional state based on his/her facial expression.

Also, in another published work (Mehrabian, 1968), it was confirmed that only 7% of the effect of the message is contributed to the verbal part; the vocal part (e.g., voice intonation) has an effect of 38 percent, while 55% goes to the facial expression of the speaker on the effect of the spoken message. That means, the facial expressions has a great major modality in human communication. Facial expressions provide the building blocks which help in process of understanding emotions. In order to effectively use facial expressions, it is necessary to understand how to interpret expressions, and it is also important to study what others have done in the past.

The development of automatic system for facial expression recognition, therefore, would be a new modality for the man-machine interaction at high efficiency and tighter interaction. It will also serve many applications in practice such as research in behavioral science, medicine and security. The research work made in reference (Batty & Taylor, 2003), is a substantial activities in this area. The work succeeded to identify six basic emotions

(anger, fear, disgust, joy, surprise, sadness) that appear to be universal across humanity as shown in Figure 1.

Generally there are three major steps in AFER. The first step is to detect the face in the scene. The second step is to extract the facial expression information (facial features) that convening the facial expression and the third step is to classify the facial display conveyed by the face. To start solving the problem of facial feature extraction from input images three questions are needed to be answered: 1) Are the features holistic (spanning the whole face) or analytic (spanning subparts of the face)?; 2) Is the information spatial or spatio-temporal (static images or video)?; 3) Are the features comes from 2D images or 3D model?. The usually extracted facial features are either geometric features such as the shapes of the facial components (eyes, mouth, etc.) and the locations of facial characteristic points (corners of the eyes, mouth, etc.), or appearance features representing the texture of the facial skin in specific facial areas including wrinkles, bulges, and furrows. Appearance-based features include learned image filters from Principal Component Analysis (PCA), Gabor filters, features based on edge-orientation histograms, etc. Several efforts have also been reported which use both geometric and appearance features. These approaches to AFER are referred to as hybrid methods. The facial expression classification systems are divided into two categories; person dependent and person independent systems. In the first category the system trains on facial expressions of certain persons and recognize the facial expressions for the same persons but in the second category the persons in training phase defer from the testing one.

2. Related Work

The pain expression is one of the important facial expressions. Md. Maruf Monwar et al use location and shape features to represent the pain information (Monwar & Rezaei, 2006). These features are used as inputs to the standard back-propagation in the form of a three-layer neural network with one hidden layer for classification of painful and painless faces. They achieve 91.67% recognition rate using 10 hidden layer units.

The Facial Action Coding System (FACS) (Ryan, Cohn, Lucey, Saragih, Lucey, Torre, & Rossi, 2009) and (Lucey, Cohn, Lucey, Matthews, Sridharan & Prkachin, 2009) is currently the most widely used method in recognizing facial expressions. FACS encodes the contraction of each facial muscle (stand alone as well as in combination with other muscles) changes the appearance of face. It has been used widely for the measurement of shown emotions.

Gizatdinova and Surakka used feature-based method for detecting landmarks from facial images (Gizatdinova, & Surakka, 2006). The method was based on extracting oriented edges and constructing edge maps at two resolution levels. Edge regions with characteristic edge pattern formed landmark candidates.

An optical-flow based approach (Cohn, Zlochower, Lien, & Kanade, 1998) is sensitive to subtle changes in facial expression. Action unit (AU) (Valstar & Pantic, 2006) and (Valstar, Patras & Pantic, 2004) combinations in the brow and mouth regions were selected for analysis if they occurred at a minimum of twenty-five times in the database. In these studies, the hierarchical algorithm for estimating optical flow was adopted to automatically trace the facial features and the image sequences were randomly divided into training and test sets.

Cootes et. al and Ratliff et. al developed the Active Appearance Model AAM that shown strong potential in a variety of facial recognition technologies (Ratliff, Patterson, 2008) and (Ratliff, & Patterson, 2008). It has the ability to aid in initial face-search algorithms and in extracting important information from both the shape and texture (wrinkles, nasio-labial lines, etc.) of the face that may be useful for communicating emotion.

Zhang et al. (Zhang, Lyons, Schuster & Akamatsu, 1998) made a comparison between geometry-based and gabor-wavelets-based facial expression recognition system without dealing with facial expression information extraction in an automatic way. They use 34 facial points for which a set of Gabor wavelet coefficients is extracted. Wavelets of three spatial frequencies and six orientations have been utilized. Zhang et al. deal only with 256×256 pixels frontal view images of nine female Japanese subjects, manually normalized so that the distance between the eyes is 60 pixels. Their system achieved a generalized recognition rate of 73.3% with geometric positions alone, 92.2% with Gabor wavelet coefficients alone, and 92.3% with the combined information.

3. Proposed System

This paper propose a system for classifying the six basic emotions (anger, disgust, fear, happy, sad, surprise) in addition to the neutral one using two types of features. It generally enter the image that contains the face under check into face detection process to segment the face image. Then the feature extraction process will applied on the face image to produce a feature vector that consists of two types of features: geometric features and appearance features which represents a pattern for facial expression classes. Finally; this feature vector used as an input into the radial basis function artificial neural network to recognize the facial expressions. The block

diagram of the proposed system is shown in Figure 2.

3.1 Face Detection

The first step in this work is the face detection and segmentation from the whole scene. In recent years a lot of researches have been developed many systems to produce robust techniques for face detections. This paper use the open source code library (OpenCV) that employs a face detection algorithm based on Viola & Jones features (Sreekar 2010) and (Viola & Jones 2001). The resulted face image was normalized to a uniform size of 300×300 pixels.

3.2 Facial Features Extraction

Generally the most important step in the field of facial expression recognition is the facial feature extraction which based on finding a set of features that conveying the facial expression information. This problem can be viewed as a dimensionality reduction problem (transforming the input data into a reduced representation set of features which encode the relevant information from the input data). This paper use two methods to extract the facial features: geometric features and appearance features.

3.2.1 Geometric Features Extraction

In this step 19 features are extracted from the face image. First; the segmentation process is performed to divide the face image into three regions of interest: mouth, nose and two eyes and two eyebrows as shown in Figure 3. This is done by taking into consideration that the detected face is frontal or near frontal and assuming certain geometric constraints such as: position inside the face, size and symmetry to the facial symmetry axis.

Second; the facial characteristic points (FCPs) are located in each face component using mouth, nose, eyes and eyebrows FCPs extraction techniques. Finally; certain lengths between FCPs are calculated using Euclidean distance D:

$$D = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$$
(1)

Where (x1, y1) and (x2,y2) are the coordinates of any two FCPs P1(x1,y1) and P2(x2,y2) respectively. Also two angles in the mouth area are computed to represent with geometric lengths the geometric features, typically results are in Figure 4 (a, b).

The problem now is how to detect the FCPs. Due to the great evolution in image processing researches; many approaches in facial feature detection (Yang, Stiefelhagen, Meier & Waibe 1998) and approaches in eye detection (Rajpathaka, Kumarb, & Schwartzb (2009) have get encouraging results.

3.2.1.1 Mouth FCPs Extraction:

As illustrated before the first process after face detection is the segmentation process. The mouth area chosen by taking the portion of the face image from 25% of the face image width after the left boarder to 25% of the face image width before the right edge and from 66.67% of the face image height after the top boarder to 5% of the face image height before the bottom boarder. The next step is detecting the FCPs inside the mouth region. This is done by using two transformations of the mouth image after applying "Unsharp" filter to sharpen the mouth image for more clarification of mouth details. The first one is transform the mouth image from gray scale into binary image using certain threshold τ which chosen as a function of the entire mouth image threshold T and factor f as the following equation:

$$\tau = \frac{T}{f}$$
(2)

Where the value of f determined by manual iterations (empirically). Due to the large diversity in mouth shape as shown in Figure 5. It is recoded that best results of $f = 3$. Binary dilate morphological operation with linear 1×5 structural element and erode morphological operation with linear 5×1 structural element are applied after the first transformation.

The second transformation is converting the mouth image into binary using the canny edge detection with also static threshold value which after manual iterations chosen to be 0.5. Also close morphological operation with structural element 5×1 is applied after the second transformation. The two results are added together to get the final binary image of the mouth as shown in Figure 6:

The previous transformations have overcome the problem of mouth shape diversity. In case of angry, disgust, normal, sad and surprise classes the first transformation can get the mouth region in good way because of the

mouth often represented by the two lips only unlike in the fear and happy classes where the teeth often appears and results to generate many objects after the transformations. So using canny edge detection is more accurate in the last category of classes. Since the first transformation is robust for the first category of classes and the second for the second; the proposed approach here is using the both to get an object which represents the mouth region in the mouth image. After that the labeling operation is performing then using the blobs analysis to determine the area and "Bounding Box" properties of each object in the image. The object with maximum area selected to be the mouth rather than other candidate objects. Using the bounding box property of the mouth object will be easy to determine each of the four mouth FCPs as a mean point of non zero values of the corresponding boarder (left, right top and bottom).

3.2.1.2 Nose FCPs Extraction

The nose FCPs are mainly used to measure the distance from them to eyes centers and mouth FCPs. Also the first step is the segmentation process with the values 40%, 40%, 45% and 30% for the left, right, top and bottom boarders respectively to get the nose image Figure 7(a). Unlike the mouth in varying their shapes with respect to facial expression classes; the nose hole shape often have the same shape as circle (hole). So an iteration process is performed to get at least two holes from the binary image and the maximum two object are selected after labeling and blob analysis operations Figure 7(b). After that the "Centroid" property of the two holes are determined two locate the corresponding two nose FCPs. Figure 7(c).

3.2.1.3 Eyes and Eyebrows FCPs Extraction

In eyes and eyebrows the proposed approach is to deal with each eye-eyebrow separately. Also starting with the segmentation process with the values 20%, 20%, 20% and 50% for the left, right, top and bottom boarders respectively is the first step to obtain eyes and eyebrows area. Taking into account the symmetry of the frontal face image the left eye-eyebrow pair is separated from the right one by taking the left half of the segmented area for it and the right one for the other. The next step is to separate each eye-eyebrow to deal with it singly. This done by finding a separated border between them. The integral projection method is used to determine the horizontal borders of the line between the lips. The boarder finding algorithm consists of five steps:

1- Apply "Prewitt" filter that emphasizes horizontal edges by approximating a vertical gradient.

2- Compute the vertical integral projection (Zhou & Geng, 2004) $s(x)$ for $n \times m$ eye-eyebrow image by using equation (3):

$$s(x) = \sum_{y=1}^{m} I(x, y) \tag{3}$$

Where $x \in \{1, 2, \ldots, n\}$ is the row number.

3- Smooth the results using a moving average filter to minimize the number of peaks around global maxima.

4- Find global maxima

5- If the number of maxima is greater than two peaks; the border is computed as the average position of the maximum two peaks position. Else it is one peak and then the border will be the position of this peak.

Figure 8(a-e) shows a typical result from the previous algorithm steps.

The next step after determining the eye-eyebrow border is to separate the eye and eyebrow using this border. Then the FCPs of the eye and eyebrow can be determined using the transformation to binary with threshold τ technique. Also some image enhancements are applied using Gaussian filter and image histogram equalization before getting the binary image. And also "Close" morphological operation (with 5×5 "Disk" structural element in eye image and 3×12 "rectangle" structural element in brow image) and labeling operation are applied in eye image. Finally, by using blobs analysis the area property is calculated to choose the object with the maximum area that corresponds to the eye. Then locate the eye center from the "Centroid" property in eye image. After that the "Bounding Box" property is used to determine the object with maximum area that corresponding to the eyebrow. Then left, top and right FCPs in eyebrow image can located.

The final step after locating all FCPs is using equation (1) to calculate the Euclidean distance between certain

FCPs to be the extracted geometric features. These lengths with the two mouth angles were chosen because they are the most lengths that convey facial expression information as shown in Figure 4.

3.2.2 Appearance Features Extraction

The appearance features represent an important part of the facial expression feature. That is because of its holistic nature that deals with the whole face image. As mentioned before many approaches can be used to extract facial expression features such as edge orientation histograms that proposed in this paper. In this major step the normalized image (300×300 pixels) is reduced to 250×200 pixels Figure 9(a) by removing a 50 pixels from left, right and top to focus on the face without the hair. Then the "Canny" edge detection is applied to get an edge map Figure 9(b) after some image enhancements (histogram equalization and Gaussian filter). The edge map is divided into 4×4 zones. The coarsely quantized edge directions are represented as local appearance features and more global appearance features are presented as histograms of local appearance (edge directions). The edge directions are quantized into 4 angular segments Figure 9(c). An example of the histogram of edge directions is shown in Figure 9(d).

Finally, the face map is represented as feature vector of 64 components (4 histograms of 16 zones).

3.3 Facial Expression Classification

The last stage in this system is the classification of the facial expressions. The facial feature vector of 83 feature elements (19 geometric features plus 64 appearance features) forms the input of radial basis function RBF neural network. As shown in Figure 10 the RBF neural network has one hidden layer uses neurons with RBF activation functions ($\varphi_1 .. \varphi_{m1}$) for describing the receptors. The following equation is used to combine linearly the outputs of the hidden neurons:

$$y = w_1 \varphi_1(\| x - t_1 \|) + \cdots + w_{m1} \varphi_{m1}(\| x - t_{m1} \|) \tag{4}$$

Where $\| x - t \|$ is the distance of the input vector $x = (x_1, ..., x_m)$ from the vector $t = (t_1, ..., t_{m1})$ which called the center for m inputs (receptors).

The proposed system use the Gaussian function described by the equation 5:

$$\varphi(x) = e^{\left(-\frac{x^2}{2\sigma^2}\right)} \tag{5}$$

Where σ is the spread parameter of the Gaussian function. Table 1 shows the neural network topologies (number of neurons and RBF spread value) used in person-dependent and person-independent datasets.

4. Experimental Results

4.1 Datasets

One of the very famous facial expression databases is the Cohn-Kanad facial expression database (Anitha, Venkatesha, & Adiga 2010). It includes approximately 2000 image sequences with 640×480 pixel arrays and 8-bit precision for grayscale values from over 200 subjects. Subjects in the available portion of the database were 97 university students. They ranged in age from 18 to 30 years. 65% were female, 15% were African-American, and 3% were Asian or Latino.

Two datasets were prepared from the available portion to examine the proposed system with different two manners. In the first manner the proposed system trained on each subject (person) different facial expression. So; all subjects with their facial expressions were exist in both training and testing datasets which called person-dependent dataset. In this dataset the last five images from each subject facial expression were taken and used the odds (60%) in training and the evens (40%) in testing. In the second manner the system trained in the facial expressions independent on the subjects. The person-independent dataset were prepared by selecting one image that convey facial expression information from each subject facial expressions images to form the seven facial expression classes. Then a 60% of the images were selected for training phase and the rest (40%) for testing. In both datasets the normal class was prepared by picking up the first image from each expression for subject.

4.2 Results

The analysis of the results in tables (2, 3) show that the proposed system can classify the facial expressions correctly with recognition rates between 90% and 99% using person-dependent dataset and between 83% and

100% using person- independent dataset. The results prove that the proposed system could classify anger, disgust, fear, happy and surprise in maximum rates but the sad in minimum rate in person-dependent dataset. In person-independent dataset the disgust, happy and surprise classes could be recognized at maximum rates but angry and sad classes registered the minimum recognition rate. The system is implemented in MATLAB 7.6.0 on a core 2 due 1.83 GHz Windows Vista workstation.

5. Conclusions

This paper presented an automatic system for facial expression recognition. It addressed the problems of how to detect a human face in static images and how to represent and recognize facial expressions presented in those faces. A hybrid approach is used for facial features extraction as a combination between geometric and appearance facial features. RBF based neural network is used for facial expression recognition. The proposed system tested on Cohn-Kanad database and acquired results of 96 % recognition rate was achieved in person-dependent dataset while 93.5% in person-independent one.

6. Challenges and Future Work

In this work the proposed system deals with the static images. The future work will be the dealing with other **types** of images such as image sequence and 3D images. Dealing with image sequence may require approaches with very low execution time. Also dealing with 3D images will need to start with creating with 3D face models.

Acknowledgment

References

Anitha, M. K. Venkatesha, B. Adiga. (2010). A survey on facial expression databases. *International Journal of Engineering Science and Technology, Vol. 2(10), 5158-5174.*

Batty, M. & Taylor, M. (2003). Early processing of the six basic facial emotional expressions, *Elsevier B.V.*

Cohn, J. F., Zlochower, A. J., Lien, J.J., & Kanade, T. (1998). Feature-Point Tracking by Optical Flow Discriminates subtle Difference in Facial Expression., *Proc. Int'l Conf. Automatic Face and Gesture Recognition, Nara, Japan, pp3396, April 14-16.*

Edwards, G.J., Cootes, T.F. & Taylor, C.J. (1998). Face recognition using active appearance models. *In proceedings of the European conference on computer vision.*

Gizatdinova, Y. & Surakka, V. (2006) .Feature-Based Detection of Facial Landmarks from Neutral and Expressive Facial Images, *IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28, no. 1, pp.135-139.*

Lucey, P., Cohn, J., Lucey, S., Matthews, I. Sridharan, S. & Prkachin, K. M. (2009). Automatically detecting pain using facial actions, *IEEE.*

Mehrabian. (1968). Communication without Words, *Psychology Today, vol. 2, no. 4, pp. 53-56.*

Monwar, M. & Rezaei, S. (2006). Pain recognition using artificial neural network, *Signal Processing and Information Technology, 2006 IEEE International Symposium on.*

Pantic, M. & Rothkrantz, L. (2000). Automatic analysis of facial expressions: the state of the art, *IEEE transactions on pattern analysis and machine intelligence, vol. 22.*

Rajpathaka, T., Kumarb R. & Schwartzb. E. (2009). Eye Detection Using Morphological and Color Image Processing, *Florida Conference on Recent Advances in Robotics, FCRAR.*

Ratliff, M. S., and Patterson, E. (2008). Emotion recognition using facial expressions with active appearance models. *International association of science and technology for development.*

Ryan, A. Cohn, J. F., Lucey, S., Saragih, J. Lucey, P. Torre, F. & Rossi, A. (2009). Automated Facial Expression Recognition System, *IEEE.*

Sreekar Krishna. (2010). Open CV Viola-Jones Face Detection in Matlab. [Online] Available: http://www.mathworks.com/matlabcentral/fileexchange/19912 (June 6, 2010).

Valstar, M. Patras, I. & Pantic, M. (2004). Facial action unit recognition using temporal templates, *IEEE International Workshop on Robot and Human Interactive Communication Kurashiki, Okayama Japan.*

Valstar, M. & Pantic, M. (2006). Fully automatic facial action unit detection and temporal analysis. *Computer Vision and Pattern Recognition Workshop.*

Viola, P. & M. Jones. (2001). Robust real-time object detection. *Second international workshop on statistical and computational theories of vision- modeling, learning, computing, and sampling.*

Yang, J., Stiefelhagen, R., Meier U. & Waibe A. (1998). Real-time face and facial feature tracking and applications, *in Proceedings of the Auditory-visual Speech Processing (AVSP 98), NSW, Australia.*

Zhang, Z., Lyons, M., Schuster, M., & Akamatsu S. (1998). Comparison between geometry-based and gabor wavelets-based facial expression recognition using multi-layer perceptron. *Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 454-459.*

Zhou, Z. & Geng. X. (2004). Projection functions for eye detection, *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China, pattern recognition, Volume 37.*

Table 1. Radial basis function neural network topologies

Dataset	Neurons	Spread
Person-dependent	250	250
Person-independent	80	300

This table contains different topologies of the neural network.

Table 2. Confusion matrix for facial expression recognition in person-dependent dataset

%	Anger	Disgust	Fear	Happy	Normal	Sad	Surprise
Anger	**95.83**	0	0	0	3.13	1.04	0
Disgust	0	**96.59**	0	0	3.41	0	0
Fear	0	0	**99.17**	0	.83	0	0
Happy	0	0	0.58	**97.70**	1.72	0	0
Normal	2.23	0.89	1.79	2.23	**89.73**	2.78	0.45
Sad	1.61	0	1.61	0	4.84	**91.94**	0
Surprise	0	0	0	0	1.27	0	**98.73**

This table contains the experiential results of facial expression recognition in person-dependent dataset.

Table 3. Confusion matrix for facial expression recognition in person-independent dataset

%	Anger	Disgust	Fear	Happy	Normal	Sad	Surprise
Anger	**83.33**	5.56	5.56	0	0	5.56	0
Disgust	0	**100**	0	0	0	0	0
Fear	0	0	**91.68**	0	4.16	4.16	0
Happy	0	0	0	**100**	0	0	0
Normal	0	0	0	0	**94.6**	5.4	0
Sad	0	0	0	0	12	**88**	0
Surprise	0	0	3.13	0	0	0	**96.87**

This table contains the experiential results of facial expression recognition in person-independent dataset.

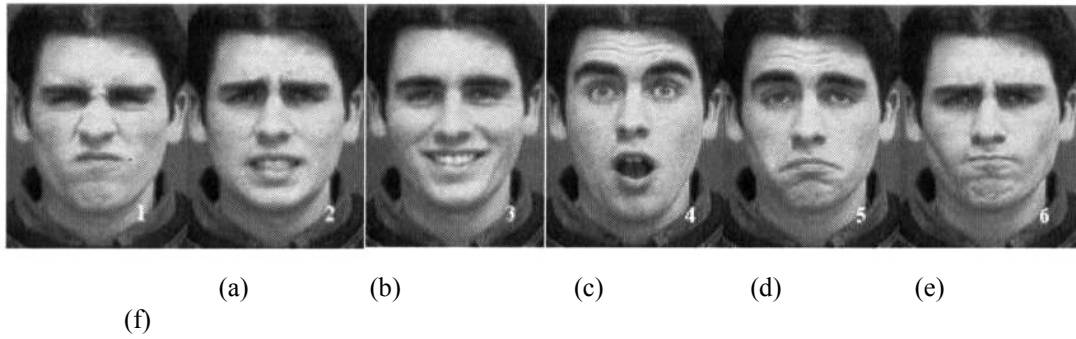(a)   (b)   (c)   (d)   (e)
(f)

Figure 1. Basic emotion :(a) disgust, (b) fear, (c) joy, (d) surprise, (e) sadness, (f) anger
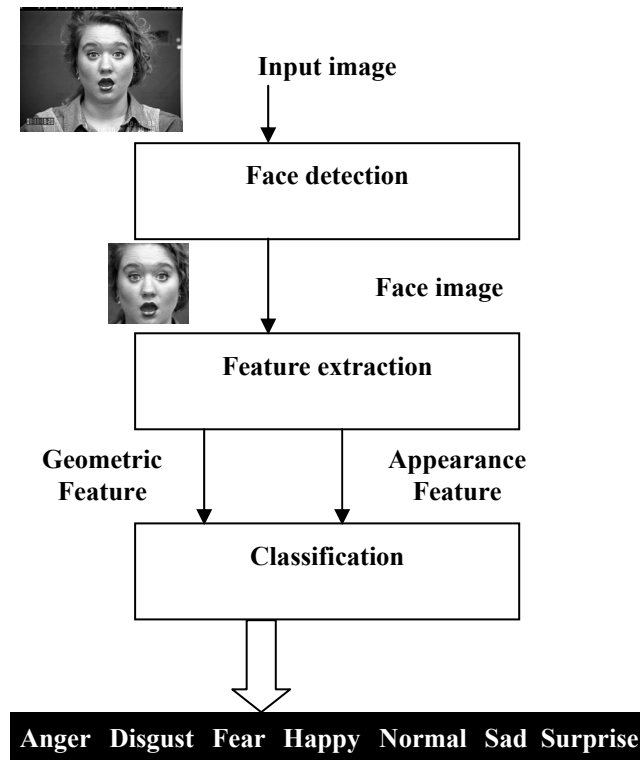

Figure 2. Block diagram of the facial expression recognition system

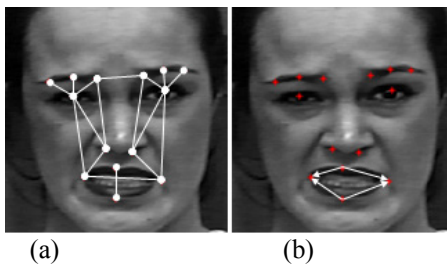

Figure 3. Segmentation process

(a)   (b)

Figure 4. Geometric features: (a) geometric lengths, (b) mouth angles
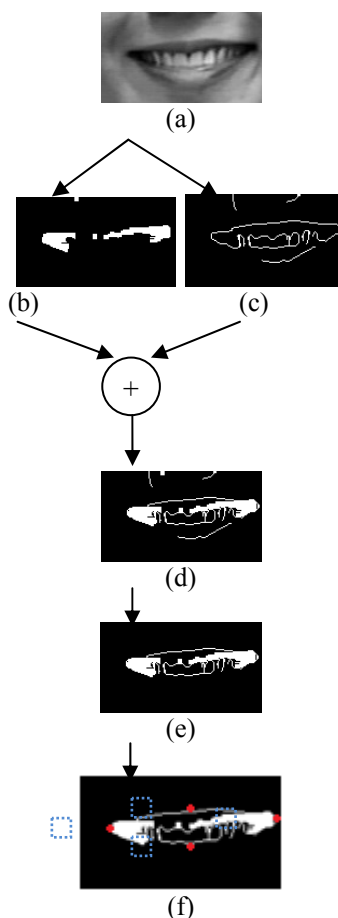

Figure 5. Different mouth shapes

(a)

(b)   (c)

(d)

(e)

(f)

Figure 6. Mouth object extraction: (a) mouth image, (b) transformation to binary using threshold τ, (c) canny edge detection, (d) summing result, (e) mouth object, (f) mouth FCPs.
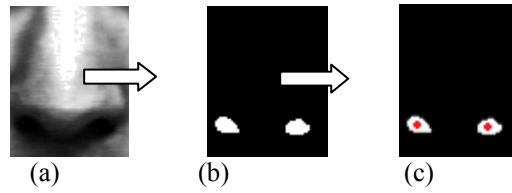
(a)   (b)   (c)

Figure 7. Nose FCPs: (a) nose image, (b) binary nose image, (c) FCPs
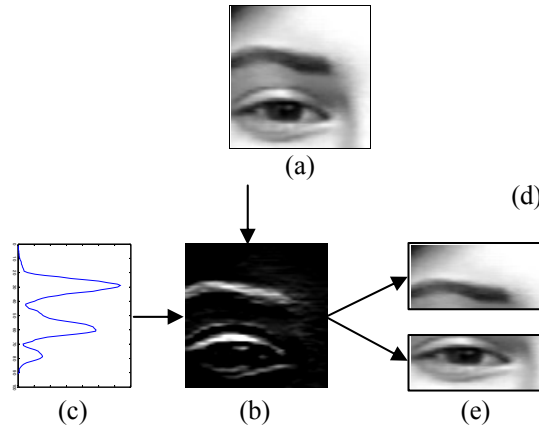

(a)

(d)

(c)   (b)   (e)

Figure 8. Eye-eyebrow separation: (a) eye and eyebrow image, (b) binary image,
(c) integral projection, (d) two eye and eyebrow images
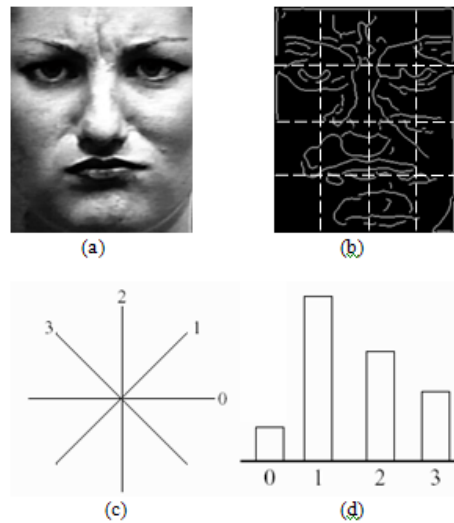

(a)   (b)

(c)   (d)

Figure 9. Zonal-histogram features. (a) Normalized face (b) zones of the edge map,

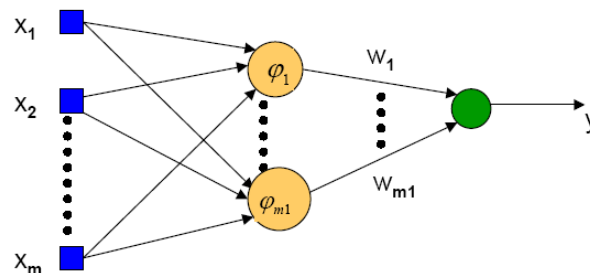(c) four quantization levels for calculating histogram features, (d) histogram of one zone.

Figure 10. Radial basis function neural network architecture