Emotion Recognition from Facial Expressions: A Target Oriented Approach Using Neural Network

Sreevatsan. A.N, Sathish Kumar. K.G, Rakeshsharma. S and Mohd. Mansoor Roomi
Dept. of Electronics and Communication Engineering,
Thiagarajar College of Engineering,
Madurai – 625015, INDIA.
{sreevats_raji@yahoo.com, sathish_mmw@yahoo.com, s_rakesh145@yahoo.co.in, smroomi@yahoo.com}

Abstract

Effective Human Computer Intelligent Interaction (HCII) requires the information about the user's identity, state and intent which can be extracted from images, so that computers can then react accordingly, e.g. systems behaving according to the emotional state of the person. The most expressive way humans display emotions is through facial expressions. Here, we propose an efficient method for emotion recognition from facial expressions in static color images containing the frontal view of the human face. Our goal is to categorize the facial expression in the given image into six basic emotional states – Happy, Sad, Anger, Fear, Disgust and Surprise. Our method consists of three steps, namely face detection and localization, facial feature extraction and emotion recognition. First, face detection is performed using a novel skin-color based segmentation and connected component analysis which is then followed by the exact face localization by using a knowledge based approach. Next, the extraction of facial features such as the eye and the mouth is performed by employing an iterative search algorithm, on the edge information of the localized face region in gray scale. Finally, emotion recognition is performed by giving the extracted eye and mouth blocks as inputs to a feed-forward neural network trained by back-propagation.

1. Introduction

Computer vision is the branch of artificial intelligence that focuses on making computers to emulate human vision, including learning, making inferences and performing cognitive actions based on visual inputs, i.e. Images. Computer vision also plays a major role in Human Computer Intelligent Interaction (HCII) which provides natural ways for humans to use computer as aids. With the ubiquity of new information technology and media, more effective methods for HCII are being developed which rely on higher level image analysis techniques whose recent applications in HCII include automatic interactive tutoring, multimedia, process control and user authentication. For these tasks, the required information about the identity, state and intent of the user can be extracted from images and make the computers to react accordingly, e.g. by observing a person's facial expressions. Face perception is a very important component of human cognition. Faces are rich in information about individual identity, and also about mood and mental state, being accessible windows into the mechanisms governing our emotions. It is argued that for the computer to be able to interact with humans, it needs to have the communication skills of humans. One of these skills is the ability to understand the emotional state of the person. The most expressive way humans display emotions is through facial expressions. Recognition of emotions from facial expressions involves the task of categorizing active and spontaneous facial expressions so as to extract information about the underlying emotional states. Building an automatic system for emotion recognition from facial expressions is also useful for designing new interactive devices offering the p ossibility of new ways for humans to interact with computer systems. Approaches for the recognition of emotions from facial exp ressions can be are divided into two main categories [7]: target oriented and gesture oriented. In the former, recognition of a facial expression is performed using a single image of a face at the apex of the expression. Gesture-oriented approaches extract facial temporal information from a sequence of images. Transitional approaches were also developed that use two images, representing a face in its neutral condition and at the apex of the expression. We employ a target-oriented approach and our proposed work consists of three steps, namely face detection and localization, facial feature extraction and emotion recognition. The existing methods for face detection and feature extraction are knowledge based approaches, feature invariant techniques, template matching and appearance based methods [1]. Next, for emotion recognition, the existing methods are template based, neural network based and rule based approaches [7]. We have adopted a method, which is robust for varying skin colors using a combination of feature invariant [3] and knowledge based approaches for face detection and localization from a color image containing the frontal view of the human face. After localizing the face region we perform the extraction of facial features namely the eye and mouth

blocks by an iterative search algorithm making use of the edge information of the cropped face region in gray scale. Finally, we perform the recognition of six basic emotions [9] – happy, sad, anger, fear, disgust, surprise, by giving the extracted blocks of eye and mouth to a feed forward neural network trained by back-propagation [6]. Section 2 describes the proposed algorithm for emotion recognition. Section 3 presents the results of our algorithm on several face databases. Conclusions and future work are described in Section 4

2. Proposed Algorithm for Emotion Recognition

An overview of the proposed work contains three major modules: (i) face detection and localization, (ii) facial feature extraction and (iii) emotion recognition from the extracted features. The first step, namely face detection is performed on a color image containing the frontal view of a human subject.

2.1 Face detection and Localization

The block diagram for face detection and localization is given in Figure 1. A feature-invariant technique namely skin color based segmentation and connected component analysis is performed for face detection which is followed by a knowledge-based approach for exact face localization.

2.1.1 Skin-color based segmentation

The human skin color is distributed in the RGB color space. But, in the chromatic color space, the color distribution of skin colors of different people is found to be clustered in a small area.

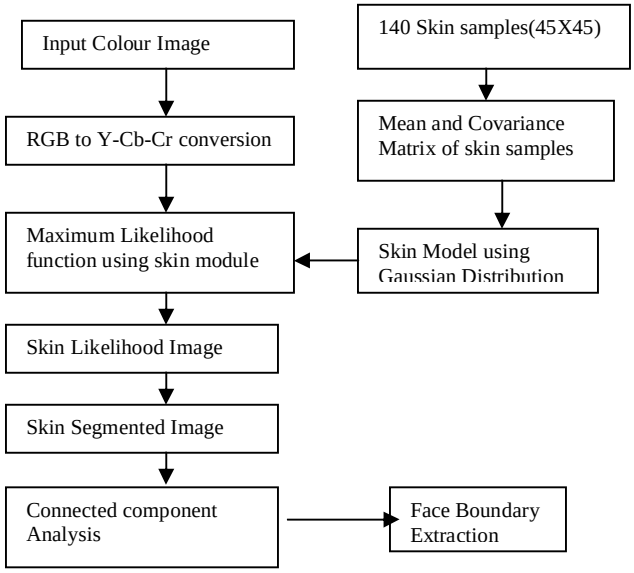


Figure1:FaceDetection_and_Localization

Although skin colors of different people appear to vary over a wide range, they differ much less in color than in brightness. With this finding, we proceed to develop a skin-color model in the chromatic color space. Here we adopt the YCbCr color space [10] since it is perceptually uniform, is widely used in video compression standards (e.g., MPEG and JPEG), and it is similar to the TSL (Tint, Saturation and Luma) space in terms of the separation of luminance and chrominance as well as the compactness of the skin cluster. In order to segment human skin regions from non-skin regions based on color, we need a reliable skin color model [3] that is adaptable to people of different skin colors and to different lighting conditions. We generate a statistical skin-color model by means of a supervised training, using a set of 140 skin-color regions of size 45 x 45, obtained from a color face database. Such images were obtained from people of different races, ages and gender. Thus, a skin color distribution can be represented by a Gaussian model N (m, C), [3], where distribution of skin colors of different people is found to be clustered in a small area. Although skin colors of different people appear to vary over a wide range, they differ much less in color than in brightness. With this finding, we proceed to develop a skin-color model in the chromatic color space. Here we adopt the YCbCr color space [10] since it is perceptually uniform, is widely used in video compression standards (e.g., MPEG and JPEG), and it is similar to the TSL (Tint, Saturation and Luma) space in terms of the separation of luminance and chrominance as well as the compactness of the skin cluster. In order to segment human skin regions from non-skin regions based on color, we need a reliable skin color model [3] that is adaptable to people of different skin colors and to different lighting conditions. We generate a statistical skin-color model by means of a supervised training, using a set of 140 skin-color regions of size 45 x 45, obtained from a color face database. Such images were obtained from people of different races, ages and gender. Thus, a skin color distribution can be represented by a Gaussian model N (m, C), [3] where,

Mean:

$$\mathbf{m = E\{x\} \text{ where } x = (Cb \quad Cr)^T\}} \tag{1}$$

Covariance:

$$\mathbf{C = E\{(x - m)(x - m)^T\}} \tag{2}$$

Figure 2 shows the Gaussian Distribution N (m, C) fitted by our data. With this Gaussian fitted skin color model, we can now obtain the likelihood of skin for any pixel of an image. First, the input color image is converted from RGB to YCbCr color space.

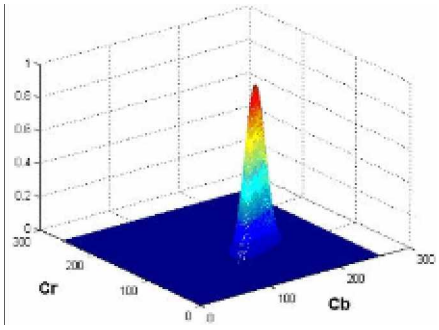


Figure 2: Skin Model

Therefore, if a pixel, in the chromatic color space, has a value of (Cb, Cr), the likelihood [1] of skin for this pixel can then be computed as follows:

$$L(i, j) = \exp(-0.5 (x - m^T C^{-1} (x - m) \quad (3)$$

where the matrices **x, C** and the mean, **m** are obtained from equations (1) and (2). This skin-likelihood image will be a gray-scale image whose gray values represent the likelihood of the pixel belonging to skin. Since the skin regions are brighter than the other parts of the image, the skin regions can be segmented from the rest of the image through a thresholding process. To process different images of different people with different skin, a fixed threshold value is not possible to be found. So we use an optimal global thresholding process [8]. After thresholding, we get a binary image with white pixels representing the skin regions.

2.1.2 Connected Component Analysis

Using the result from the previous section, we proceed to determine which region can possibly determine a frontal human face. To do so, we need to determine the number of skin regions in the image. A skin region is defined as a closed region in the image, which can have 0, 1 or more holes inside it. Its boundary is represented by pixels with value 1 for binary images. All holes in a binary image have pixel value of zero (black). The process of determining how many regions we have in a binary image is by labeling such regions. A label is an integer value. We used an 8-connected neighborhood (i.e., all the neighbors of a pixel), in order to determine the labeling of a pixel and find the connected components. Next, we determine the number of holes in each connected component. To determine the number of holes inside a region, we compute the Euler number of the region, defined as,

$$E = C - H \quad (4)$$

where E is the Euler number, C is the number of connected components and H is the number of holes in a region. The development tool (Matlab) provides a way to compute the Euler number. For our case, we already set the number of connected components (i.e. the skin region) to 1 since we are considering 1 skin region at a time. The number of holes is then,

$$H = 1 - E \quad (5)$$

2.1.3 Face Localization

Next, based on the fact that a skin region corresponding to the face has at least two holes and the assumption that face occupies a significant portion in the input image, we retain only that connected component (skin region) that has at least two holes and the largest filled area. Once the connected skin component representing the face is obtained, we then proceed to localize the face region by extracting the exact rectangular face region boundary by using the height and width of the connected component.

2.2 Facial Feature Extraction

After getting the exact face region, we first convert it into a grayscale image, and its corresponding edge image is obtained by applying the Prewitt edge operator on the cropped gray scale image. Then the facial features that correspond to a facial expression, namely the eye and mouth blocks are extracted from the face image using this edge information of the face. The actual process is described in following sections.

2.2.1 Extraction of Eye block

Based on our knowledge that the eyes are present in the upper portion of the face region, we search for eyes only in the upper half portion of the edge image. The method employs an iterative search algorithm which traverses in the vertical direction and counts the number of white pixels in horizontal overlapping blocks, as shown in Figure 3. Now, the block which contains maximum number of white pixels is the required block which contains the two eyes (For e.g. Block 4 in Figure 3).This block is extracted from the grayscale image. Then, the same algorithm is applied on the left half of this image, since we are aiming to get only one eye block. Here we use ‘canny’ method for obtaining the edges of the left half of the entire eye block. The Canny method finds edges by looking for local maxima of the gradient of the intensity image. The gradient is calculated using the derivative of a Gaussian filter. After smoothing the image and eliminating the noise, the next step is to find the edge strength by taking the gradient of the image.

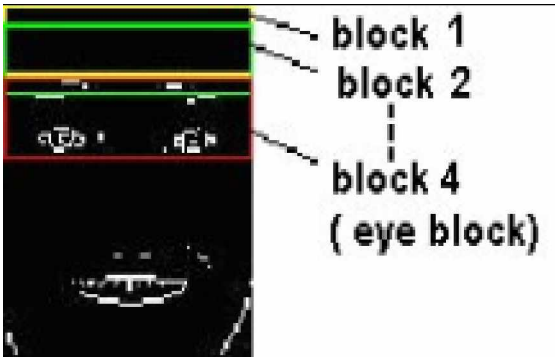


Figure3: Illustration of Feaure Extraction

This is done using a Sobel edge operator. After finding the edge, we count the number of white pixels in vertical overlapping blocks traversing in horizontal direction. The block which contains the maximum white pixels represents the exact eye block.

2.2.2. Extraction of Mouth block

The same algorithm, which is used to extract eye block, is employed here. Based on the criterion that mouth is present only in the lower half of the face, we take into consideration only the lower half of the edge image of the original grayscale image containing the face region. On applying this algorithm, we get the block containing the mouth and from its edge image, we get the exact mouth block.

2.3 Emotion Recognition

In this section, we have described our method to recognize the emotion from the extracted facial features. Though various methods exist for emotion recognition, neural networks hold its position due to its robustness. So, we employ a neural network based approach for recognizing emotions.

2.3.1 Preprocessing for Neural Network

Before the extracted features are fed as inputs to the neural network, they have to be preprocessed. This preprocessing is nothing but resizing the extracted eye and mouth blocks to a fixed size. We resize the eye block to a fixed size of 28 x 20 and the mouth block to 20 x 32, using 'nearest neighbor interpolation' method. These 2-D matrices are converted into 1-D vectors such that each row follows one another sequentially to form a single column. Thus, a 560 x1 column vector is obtained from the eye block and a 640 x 1 column vector is obtained from the mouth block. Then, we append the two column vectors of eye and mouth, which results in 1200 x 1 column vector, which is given as the input to the neural network.

S. No	Layer	Number of Neurons
1.	Hidden layer 1	120
2.	Hidden layer 2	16
3.	Output layer	6

Table 1: Neurons in each layer for our Three layer neural network

2.3.2 Network Architecture and Training algorithm

We have chosen multilayer feed forward network as the network architecture. The number of neurons that has tobe in the output layer is fixed as we know the number of emotions that we are going to consider - Happy, sad, anger, fear, disgust, surprise. Therefore, in our case, the number of neurons in the output layer is chosen as 6. The number of hidden layers in the network and the number of neurons in each layer is chosen by trial and error method based on the performance function until it reaches the specified goal. By trying various combinations, we have chosen the efficient architecture, which is a three layer feed forward network where there are two hidden layers and an output layer. The neurons in each layer are shown in Table 1. We have chosen Back-propagation training algorithm for training the network because of its simplicity and efficiency.

2.3.3Training Samples and Network Simulation

For training the network to recognize various emotions, we used different images from different databases [2], [4],[11],[13] (available in the World Wide Web belonging to various Universities) and created a new database which includes the images from our own database [12] of face images. Some of the faces from which training samples were extracted are shown in Figure 4.

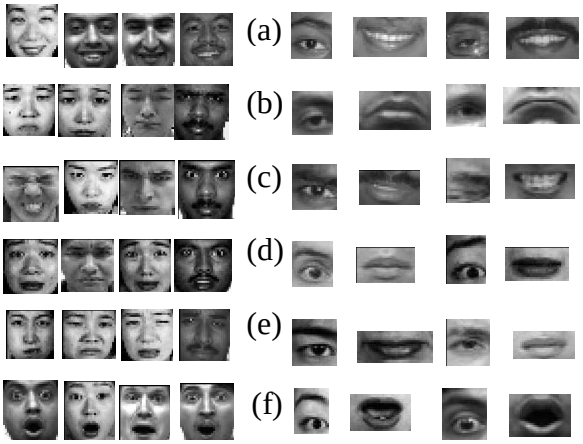


Figure 4 Training Samples for Six Emotions–(a) Happy, (b) Sad, (c) Anger, (d) Fear, (e) Disgust, (f) Surprise.

The actual training samples are the eye and the mouth blocks extracted from the set of gray scale face images. Some of these samples are also shown in Figure 4. The network is trained with the samples corresponding to various emotions, to their respective targets. The next step is to simulate the trained network using the test samples and the network output determines the exhibited emotional state.

3. Experimental Results

We have evaluated our algorithm on various color images containing the frontal view of the human face. The simulations were performed using the Image Processing Toolbox and the Neural Network Toolbox of Matlab 6.0. The images were obtained from the databases available in the World Wide Web and also from our own database. The results for face detection for a color image obtained from the AR Face Database(http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html) are given in Figure 5 and for feature extraction in Figure 6. Also, the results for two images from our face database are given in Figure 7 and Figure 8.

Next, for emotion recognition, the neural network is simulated. Before simulation, the network has been trained with samples, corresponding to various emotions as targets. The training performances of various emotions are given in Table 2. For the final step, i.e. Network Simulation, the image vector, which is of length 1200 x 1, is given as the input to the trained neural network and the network is simulated to obtain the recognized emotional category. When the network was simulated with the test input vectors corresponding to various emotions, as mentioned above, the obtained recognition rates are mentioned in Table3 .

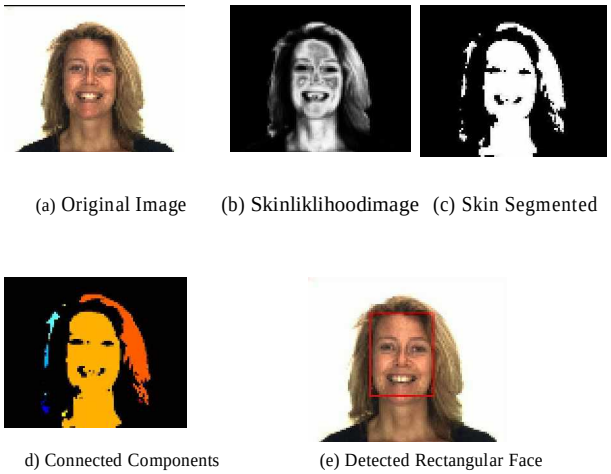


Figure 5: Face detection - Experimental Results for image from AR Face database (Purdue University)

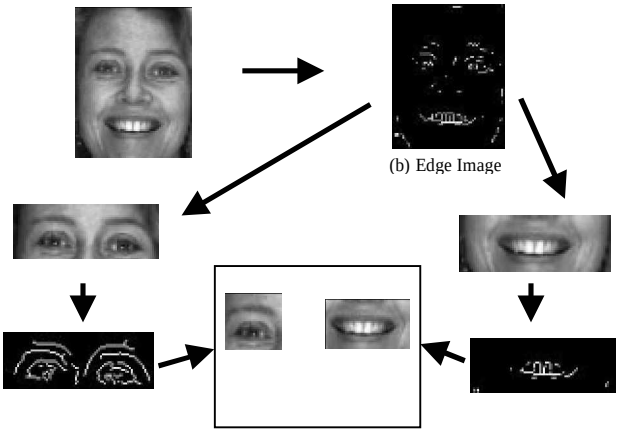


Figure 6.Results for Feature Extraction

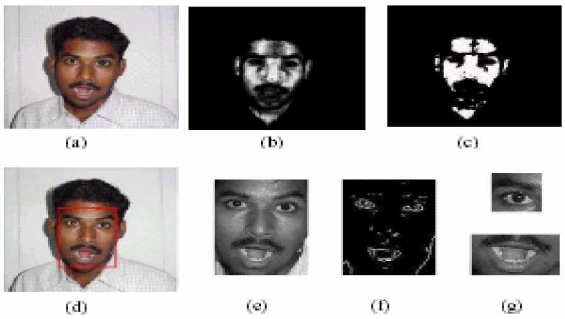
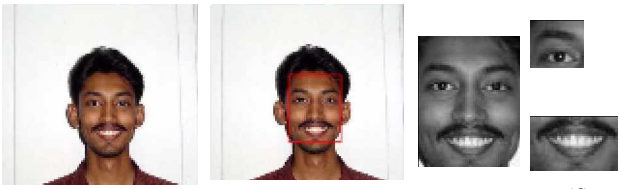


Figure 7: Results for an Image from our face database: (a) Input Color Image, (b) Skin-likelihood Image, (c) Skin-segmented binary Image, (d) Face Localized Image, (e) Cropped Face Region, (f) Edge Image, (g) Extracted exact Eye and mouth block.



(a) (b) (c) (d)

Figure 8: Results for another Image from our face database: (a) Input Color Image, (b) Face Localized Image, (c) Cropped Face Region, (d) Extracted exact Eye and mouth block.

S. No	Emotion	Error	Training Performance
1.	Happy	9.78879e-11	Goal Met
2.	Sad	7.01013e-11	Goal Met
3.	Anger	9.89466e-11	Goal Met
4.	Fear	9.5835e-11	Goal Met
5.	Disgust	9.57067e-11	Goal Met
6.	Surprise	9.3645e-11	Goal Met

Table 2: Training performance for the six emotions

S. No	Emotion	Recognition Rate (in %)
1.	Happy	83.33
2.	Sad	66.7
3.	Anger	83.33
4.	Fear	66.7
5.	Disgust	66.7
6.	Surprise	83.33

Table 3: Recognition rates for six basic emotions

We hope that the recognition rates can be improved if the network is trained for the neutral state and also by giving negative training to the network. The performance can be tried to improve by also changing the nature of the input to the neural network, for example, the facial feature contour coordinates of the subject exhibiting the emotion and that of the subject's neutral state.

Conclusion and Future Work

We have proposed an efficient method for recognizing emotion from facial expressions, which can be applied to a color image containing the frontal view of the human face. We have presented a novel scheme for face detection in a color image based upon a feature invariant approach namely, skin-color based segmentation. Face localization is then performed in an effective manner by adopting a knowledge based criterion resulting in the extraction of the exact rectangular boundary of the face region. And then, facial feature extraction has been performed on the edge information of the cropped face region in gray scale by using an iterative algorithm which searches for eyes and mouth through horizontal and vertical overlapping blocks of the edge image. And finally emotion recognition is performed using a robust feed-forward neural network. The results are encouraging for the wide range of images that we have used which makes the proposed work to be possibly used for a broad range of applications in human-computer interaction. Our future work is to improve the efficiency of our face detection method by generalizing it perform well in complex background images with varying illumination and pose and at the same time reducing the false detections. We also plan to improve the recognition rates of the neural networks by trying various networks and also selecting additional facial features.

References

[1] N.Ahuja, D.J.Kriegman, and M.H.Yang, "Detecting faces in images – A Survey", IEEE Trans. on Pattern Analysis and Machine Intelligence, Jan 2002.

[2] AR-Face Database (Purdue University) - http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html.

[3] H.Chang and U.Robes, "Face Detection", Project Report, Stanford University, May 2000.

[4] CMU Facial Expression Database, Apr 2001.

[5] G.Cottrell, C.Padgett, University of California and Ralph Adolphs, Neurology Dept, University of Iowa, "Categorical Perception Facial Emotion Classification", 1996.

[6] G.Cottrell, C.Padgett, "Identifying Emotion in Static face image", University of California, San Diego, Nov1995.

[7] E.D.Cowie, R.Cowie, W.Fellenz, S.Kollias, J.G.Taylor, N.Tsapatsoulis, G.Votsis, "Emotion Recognition in Human Computer Interaction", IEEE Signal Processing Magazine, Jan 2001.

[8] R.C.Gonzales R.E.Woods,"Digital Image ocessing", Second Edition, Pearson Education Inc.,2002.

[9] F.Hara and H.Kobayashi, "Recognition of six basic Facial expressions and their strength by Neural Network", Proc., Intl' Workshop on Robot and Human Computer interaction, 1992.

[10] R.L.Hsu, A.K.Jain,M.A.Mottaleb, "Face Detection in Color Images", IEEE Trans. on Pattern Analysis and Machine Intelligence, May 2002.

[11] JAFFE DATABASE - http://www.mis.atr.co.jp/~mlyons/jaffe.html

[12] TCE-ECE Face Database, (Thiagarajar College of Engineering, Madurai, India), (2004).

[13] YALE Face database – http://cvc.yale.edu