Emotion Detection Using Facial Expression

Deepika Ishwar, Dr. Bhupesh Kumar Singh
CSE Department, M.D.U Rohtak,
Haryana, India

Abstract—

Human emotional facial expressions play an important role in interpersonal relations. This is because humans demonstrate and convey a lot of evident information visually rather than verbally. Although humans recognize facial expressions virtually without effort or delay, reliable expression recognition by machine remains a challenge as of today. To automate recognition of emotional state, machines must be taught to understand facial gestures. In this paper we developed an algorithm which is used to identify the person’s emotional state through facial expression such as angry, disgust, happy. This can be done with different age group of people with different situation. We use Principal Component Analysis (PCA) for facial expression recognition and Bezier Curve Algorithm for feature selection.

Keywords — Principal Component Analysis, Human Computer Interaction, Universal Emotions, Bezier Curve

I. INTRODUCTION

Recognition and analysis of human facial expression and emotion have attracted a lot of interest in the past few decades, and they have been researched extensively in neuroscience, cognitive sciences, computer sciences and engineering[1]. These researches focus not only on improving human-computer interfaces, but also on improving the actions which computer takes on feedback by the user. While feedback from the user has traditionally been occurred by using keyboard or mouse, in the recent, Smartphone and camera enable the system to see and watch the user’s activities, and this leads that the user can easily utilize intelligent interaction. Human interact with each other not only through speech, but also through gestures, to emphasize a certain part of the speech, and to display of emotions. Emotions of the user are displayed by visual, vocal and other physiological means [2]. There are many ways to display human’s emotion, and the most natural way to display emotions is using facial expressions, which are mostly based on video sequences[3].

Our scheme is to automatically segment an input still image, and to recognize facial emotion using detection of color-based facial feature map and classification of emotion with simple curve and distance measure is proposed and implemented. The motivation of this paper is to study the effect of facial landmark, and to implement an efficient recognition algorithm of facial emotion with still image, while most of researches are using video sequences due to utilize the differences between frames.

II. LITERATURE REVIEW

A lot of research has been done on automatic affect analyze and recognition of human emotion[4]. Research on recognizing emotion through facial expression was pioneered by Paul Ekman[5], who started their work from the psychology perspective. He developed the Facial Action Coding System (FACS) where movements on the face are described by a set of action units (AUs). One of the first researchers who used image processing techniques to recognize facial expressions was Mase[6]. With 11 windows manually located in the face, the muscle movements were extracted by the use of optical flow. K – nearest neighbour rule was employed for the classification task of four emotions, with an accuracy of 80%.

An architecture of Hidden Markov Models for automatically segmentation and recognition of human facial expression from video sequences was proposed by Ira et al[7]. A method for facial expression recognition for a human speaker by using thermal image processing and a speech recognition system was investigated by Yashnari[8]. He improved speech recognition system to save thermal images, just before and just when speaking the phonemes of the first and last vowels, through intentional facial expressions of five categories with emotion. The extraction of appropriate facial features and consequent recognition of the user’s emotional state was suggested by Spiros et al [9] which could be robust to facial expression variations among different users. They extracted facial animation parameters defined according to the ISO MPEG-4 standard by appropriate confidence measures of the estimation accuracy. But, most of these research are used as a method to recognize emotion based on videos using extracting frames. Several prototype systems were published that can recognize deliberately produced action units in either frontal view face images[10] or profile view face images[11]. These systems employ different approaches including expert rules and machine learning methods such as neural networks, and use either feature-based image representation or appearance-based image representation. An optical flow region-based approach, by applying a Radial Basis Function Network and a Hidden Markov Model, respectively was developed by Rosenblum and Otsuka. AUs recognition by using permanent and transient facial features (lips, wrinkles) was developed by Tian et al. Geometrical models were used to locate their shapes and appearances. The Bayesian Network classifiers in the static settings and a multi-level HMM classifier to automatically segment an arbitrary long sequence to the corresponding facial expressions was introduced by Cohen[12].

The Probabilistic, Statistical and Ensemble learning techniques was employed by Valstar et al [13], which seem to be
particularly suitable for automatic action unit recognition from face image sequences. A system that recognizes various action units based on dense flow, feature point tracking and edge extraction was given by Lien[14]. The system includes three modules to extract feature information: dense-flow extraction using a wavelet motion model, facial feature tracking, and edge and line extraction. Techniques for automatically recognizing facial actions in sequences of images was explored and compared by Bartlett. These techniques include analysis of facial motion through estimation of optical flow; holistic spatial analysis, such as independent component analysis, local feature analysis, and linear discriminant analysis; and methods based on the outputs of local filters, such as Gabor wavelet representations and local principal components [15]. A new technique was suggested by author Yang, J. and Zhang known as two-dimensional Principal Component Analysis (2DPCA) for image representation. As opposed to Principal component analysis, two-dimensional principal component analysis is based on 2D image matrices rather than 1D vector. In two-dimensional Principal Component Analysis, Principal Component Analysis must be applied [16].

III. PROPOSED METHOD

The proposed method for recognition of facial expression and emotion is composed of two major steps: first one is detection and analysis of facial area from original input image, and next is verification of the facial emotion of characteristic features in the region of interest.

In the first step for face detection, the proposed method locates and detects a face in a color still image based on the skin color and the region of eye and mouth. Thus, the algorithm first extract the skin color pixels by initialized spatial filtering, based on the result from the lighting compensation. Then, the method estimates a face position and the region of facial location for eye and mouth by feature map. After obtaining the region of interest, we extract points of the feature map to apply Bezier curve on eye and mouth. Then, understanding and recognition of facial emotion is performed by training and measuring the difference of Hausdorff distance with Bezier curve between input face image and facial images in the database.

3.1 Face Detection and Location

Face detection and location is an important premise to extract facial features, therefore, corresponding to the different environment, such as light conditions, face pose, wearing material interference, detection of facial recognition system requirements with high accuracy and robustness. In the view of the complexity of face detection, the present methods of face detection technology can be broadly divided into two categories: Based on the statistical method and based on the knowledge method. Face location based on statistical method is just as the name suggest, it need a large amount of prior knowledge as theoretical basis, and through kinds of methods such as self learning algorithm of neural network, the subspace dimensionality reduction method and template matching to judge, although these methods are robust, but the large amount of calculation, long computing time, so they are not up to the requirements of real-time identification. Knowledge-based face detection method is to make full use of various characteristic parameters of human face of this unique biological identification as a standard to build models, commonly used has the organ distribution method, color and texture, motion method, symmetric method, etc. Then computer analysis of the images extracted effective characteristic parameters, contrast similarity with known facial feature parameters, so we can realize the face location. For our facial features can be detected by using such as the distance between two eyes, mouth and eyes or eyebrows and eye or some other fixed characteristics; and the color as the main features of the face, the skin color detection can use a variety of color space such as HIS color space, the YCbCr color space and CIE color space. Research shows that the YCbCr color space is more suitable for face detection compared to other several color space.

3.1.1 Detection of Skin Color using YCbCr

A first step of our scheme is color space transformation and lighting compensation. Although skin color appears to vary, we assume that there exists underlying similarities in the chromatic properties of all faces and that all major differences lie in intensity rather than in the facial skin color itself. In this case, we have adopted to utilize a skin-color based approach using YCbCr color model. With the color model, luminance information is represented by a single component, Y, and color information is stored as two color difference component, Cb and Cr[17].
After converted color model, an illumination calibration is pre-requisite during the pre-processing for the accurate face detection. Since the illumination condition is an important factor to effect on the performance of detection, we attempt pre-processing to equalize the intensity value in an image as follow:

\[ Y' = \frac{y - \text{min}_1}{\text{max}_1 - \text{min}_1} \cdot (\text{max}_2 - \text{min}_2) + \text{min}_2, \text{ if } (y \leq \text{K}_1 \text{ or } K_h \leq y) \]  

(2)

where min1 and max1 are minimum and maximum value of Y component on input image, min2 and max2 are the value of the transformed space, \( K_1 = 30 \) and \( K_h = 220 \). The values of these experiential parameters are estimated from training data sets of skin patches of sample database. Histogram equalization enhances the performance which brightness is second into one direction,

\[ P_k(r_k) = \frac{n_k}{n}, 0 \leq r_k \leq 1, k = 0, 1, \ldots, l - 1 \]  

(3)

where \( l \) is the number of discrete values for the intensity, \( n \) is the number of total pixels in the image, \( r_k \) is the \( k \) th intensity, and \( n_k \) is the number of pixels which intensity is \( r_k \). Frequency cumulativeness is dependent on \( r_k \), thus, following Equation (3) can be used for the intensity equalization.

\[ s_k = \sum_{j=0}^{k} \frac{n_k}{n} = E(r_k) = \sum_{j=0}^{k} P_r(r_j) \]  

(4)

3.1.2 The Establishment of Skin Color Model

Gauss model is a model that decomposed the specific things into a probability density distribution curve of Gauss (The normal distribution curve) to accurately quantify. Because of the similar face skin color distribution and Gauss distribution, so for each point on the face image can use the model of skin color to calculate each point probability. Then establish skin improved grayscale based on skin color probability values of these points, and then the candidate region of human face is the higher probability regions of skin color in the grayscale.

\[ m = E\{x\}, x = (Cr, Cb)^T. \]  

(5)

\[ C = E\{(x - m)(x - m)^T\}. \]  

(6)

\( m \) is the mean chroma component, \( C \) is covariance matrix. Through the Gauss skin distribution calculation of skin probability image of an arbitrary point in value. For example, for a pixel, it is converted from RGB space to YCbCr color space of skin color probability value of the formula is as follows.

\[ p(Cr, Cb) = \exp[-0.5(x - m)^T C^{-1}(x - m)] \]  

(7)

In the formula :

\[ x = (CbCr)^T \]

3.1.3 Skin Color Segmentation

Through the Gauss model we know the image of all points on the skin probability value, in order to reduce the effect of high frequency noise, first the pictures will be transferred to the YCbCr color space from RGB space through the low pass filter, determine the gray level of each pixel point value, and then processing the skin area and other background regions separated threshold segmentation and binarization, using the following formula to calculate the specific:

\[ f_T(x, y) = \begin{cases} 1, & \text{if } f_T(x, y) \geq T \\ 0, & \text{otherwise} \end{cases} \]  

(8)

\( T \) is the selected threshold, that is greater than \( T \) the pixel is 1 (white), the other is 0 (black). Usually \( T \) algorithm has the Bimodal method, \( P \) parameter method, Otsu method, iterative method, here we use the Otsu method, the basic idea of Otsu method is to use an assumed gray value \( t \) divided image gray into two groups, when two groups of between class variance is maximum, the gray value of image \( t \) is the value of the optimal threshold [18]. Suppose the image’s gray value is \( M \), the range is \( 0-M-1 \), select the gray value of \( T \) within this range, the image is divided into two groups of \( G_0 \) and \( G_1 \), \( G_0 \) contains the pixel gray values in \( 0-t \), \( G_1 \)’s gray value is in \( t+1-M-1 \), with \( N \) as pixel number, \( n_{i} \) said the number of pixel gray value of \( i \) the number of assumptions, \( G_0 \) and \( G_1 \) two groups of pixels which occupies in the whole image respectively, two groups of average gray value, so probability:

\[ \omega_0 = \sum_{t=t}^{M-1} p_i, \omega_1 = \sum_{t=t+1}^{M-1} p_i = 1 - \omega_0 \]  

(9)

The average gray value are:

\[ \mu_0 = \sum_{i=0}^{t} i p_i \text{ and } \mu_1 = \sum_{i=t+1}^{M-1} i p_i \]
$T$ is the optimal threshold.

$$\mu = \mathcal{O}_0 \times \mu_0 + \mathcal{O}_1 \times \mu_1,$$

$$g(t) = \omega_0(\mu_0 - \mu)^2 + \omega_1(\mu_1 - \mu)^2 = \omega_0\omega_1(\mu_0 - \mu_1)^2,$$

$$T = \arg\max(g(t)).$$

Morphological processing: morphology of two value morphology transform image is a view of the process of the collection, its essence is the interaction by the image of objects, or the shape of the set and structural elements of intersection and set operation. The main methods are corrosion $X$ and expansion $Y$:

$$X = E \ominus B = \{x : B(x) \subset E\},$$

$$Y = E \oplus B = \{y : B(y) \cap E = \emptyset\}.$$

Corrosion is a process to eliminate boundary point so that the boundary contraction to interior, it can be used to eliminate small and insignificant objects. Expansion is on the contrary with corrosion, it is a process that all the background point will contact with objects and merged into the object, so that the boundary to the external expansion of the cavity can be used to fill the object.

Through the steps above we can get the effective area of the faces in the image. Figure 2 shows the process of face detection and location.

3.2 POSITIONING EYES AND LIPS

3.2.1 Edge Detection

The edge is the most basic feature of image. The so-called edge is the pixels collection of its surrounding pixels having step edge or changes in the roof edge. Edge exists widely in the object and the background, objects and between element and element. Therefore, it is an important feature of image detection and extraction techniques. The step edge on both sides of the gray value changes obviously, and at the junction of roof edge in gray level increase and decrease.

Another is the rise and drop which is formed by the combination of pulse shape edge profile corresponds to a thin strip of gray value mutation. Commonly we use one order and two order derivative to describe and edge detection. If a pixel in the image of a certain object boundary, then its neighborhood will become a gray level change zone. The two most useful features of this change are the variance rate of gray level and direction, and they express by the magnitude and direction of the gradient vector. Two dimensional function $f(x, y)$ is defined as the gradient vector direction derivative. Edge detection operator checks the neighborhood of each pixel and to quantify the variance rate of gray level, including determines direction, most of the use of methods based on directional derivative mask convolution.

3.2.2 Canny Edge Detection

Since the eyes are located in the middle of the top face, lips are in the middle of the bottom face, so the limited area search can improve the detection speed. First estimates the eyes and lips region, and then use the Canny edge detection to determine the exact position of the eyes and lips[19].

(1) Using Gauss filter to smooth the image, The Gauss smoothing function is computed as

$$H(x, y) = e^{-\frac{a^2 + b^2}{2\sigma^2}},$$

$$G(x, y) = f(x, y) * H(x, y).$$
(2) To calculate the gradient magnitude and direction by means of finite first-order partial derivative; A first-order differential convolution template:

\[
H_1 = \begin{bmatrix}
1 & -1 \\
1 & 1 \\
\end{bmatrix},
\]

\[
H_2 = \begin{bmatrix}
1 & 1 \\
-1 & -1 \\
\end{bmatrix}.
\]

(16) \hspace{1cm} (17)

\[
\varphi_1(m, n) = f(m, n) \cdot H_1(x, y),
\]

\[
\varphi_2(m, n) = f(m, n) \cdot H_2(x, y).
\]

(18) \hspace{1cm} (19)

\[
\varphi(m, n) = \sqrt{\varphi_1^2(m, n) + \varphi_2^2(m, n)}.
\]

(20)

\[
\theta_\varphi = \tan^{-1}\frac{\varphi_2(m, n)}{\varphi_1(m, n)}.
\]

(21)

(3) Using gradient direction to suppression non maxima on the gradient magnitude; Figure 3 show the edge direction and angular direction of each point. the purpose is combined with the amplitude and direction of the gradient vector to removal the false edge, specifically to along the gradient direction of eight edge direction of each point one by one to detect modulus maxima, and then compared with the pixels of the partial derivative, then maximum value as the optimal edge points, set the point of gray value as 0.

![Figure 3. Schematic diagram of the edge direction and 8 neighborhood angular direction](image)

(4) Using double threshold algorithm connecting edge; through suppression non maxima we can removed the false edge beside real, for the false edge beside false edge need to further removed to draw the edge contour clearly. The most important things are the two adaptive threshold \(\tau_1\) and \(\tau_2\), and general requirements \(2 \tau_1 \approx \tau_2\), and the strong threshold \(\tau_2\) determined true edge image \(N2[i,j]\), then the weak threshold on the true edge connection processing, making sure the edge is smoothly and completely. According to different profile of the two eyes, testing individually and extract eyes and lips edges[20].

3.2.3 Drawing Bezier Curve on Eye and Mouth

The Bezier curve generates contour points considering global shape information with the curve passing through the first and last control points [21]. If there are \(L + 1\) control points, the position is defined as \(P_k : (x_k; y_k); 0 \leq k \leq L\) considering 2D shapes. These coordinate points are then blended to form \(P(t)\), which describes the path of Bezier polynomial function between \(P_0\) and \(P_L\):

\[
P(t) = \sum_{k=0}^{L} P_k \cdot BEZ_{k,L}(t)
\]

(22)

where the Bezier blending function \(BEZ_{k,L}(t)\) is known as the Bernstein polynomial, which is defined as [22]:

\[
BEZ_{k,L}(t) = \binom{L}{k} t^k (1 - t)^{L - k}
\]

(23)

The recursive formula which are used to decide coordinate locations is obtained by:

\[
BEZ_{k,L}(t) = (1 - t) \cdot BEZ_{k,L-1}(t) + t \cdot BEZ_{k,L-1} + t \cdot BEZ_{k-1,L-1}(t)
\]

(24)

where \(BEZ_{k,k}(t) = t^k\) and \(BEZ_{0,k}(t) = (1 - t)^k\).

The coordinates of individual Bezier curve are represented by the following pair of parametric equations:

\[
x(t) = \sum_{k=0}^{L} x_k \cdot BEZ_{k,L}(t)
\]

\[
y(t) = \sum_{k=0}^{L} y_k \cdot BEZ_{k,L}(t)
\]

(25)

An example of a Bezier curve is given in Figure 4 with 4 control points [23].
For applying the Bezier curve, we need to extract some control points of each interest regions, where locate in the area of left eye, right eye and mouth. Thus, we apply big connect for finding the highest connected area within each interest regions from the eyemap and mouthmap. Then, we find four boundary points of the regions, which are the starting and ending pixels in horizontals, and the top and bottom pixels of the central points in verticals. After getting four boundary points of each regions, the Bezier curves for left eye, right eye and mouth are obtained by drawing tangents to a curve over the four boundary control points.

### 3.2.4 Training and Recognizing Facial Emotion with Hausdorff Distance

In training database, there are two tables which are storing for personal information and indexes of four emotions with his/her own curves of facial expression analysis. For detection of facial emotion, we need to compute the distance a one-to-one correspondence of each interest regions between an input image and the images in the database. The Bezier curves are drawn over principal lines of facial features. To estimate a similarity matching, we first normalize the displacements that converts each width of the Bezier curve to 100 and height according to its width. We then apply the Hausdorff distance to compare the shape metric between them. The distance $d_H(p; q)$ between two curves $p(s), s \in [a, b]$ and $q(t), t \in [c, d]$ given in Equation (25)[24].

$$d_H(p; q) = \max \{ \max_{s \in [a, b]} \min_{t \in [c, d]} |p(s) - q(t)|, \max_{t \in [c, d]} \min_{s \in [a, b]} |p(s) - q(t)| \}$$

### IV. EXPERIMENTS AND RESULT

The expressions such as smile, sad, surprise and neutral are considered for the experiment of face recognition. The faces with expressions are compared against the model face database consisting of neutral faces. All the face images are normalized using some parameters. The Bezier points are interpolated over the principal lines of facial features. These points for each curve form the adjacent curve segments. The Hausdroff distance is calculated based on the curve segments. Then, understanding and decision of facial emotion is chosen by measuring similarity in faces.

To categorize facial emotion, we need first to determine the expressions from movements of facial control points. Ekman et al. have produced a system for describing visually distinguishable facial movements (called Facial Action Coding System, FACS), which is based on enumeration of all action units(AUs). There are 46 AUs in FACS that account for changes in facial expression, and combination rules of the AUs are considered to determine form defining emotion-specified expressions. The rules for facial emotion are created by basic AUs from FACS, and the decision of recognition is determined from their rules, as given in Table 1.

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Movements of AUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smile</td>
<td>Eye opening is narrowed, Mouth is opening, and Lip corners are pulled obliquely</td>
</tr>
<tr>
<td>Sad</td>
<td>Eye is tightened closed, and Lower lip corner depressor</td>
</tr>
<tr>
<td>Surprise</td>
<td>Eye and Mouth are opened, Upper eyelid raiser, and Mouth stretch</td>
</tr>
</tbody>
</table>

In this work, the facial expressions have been recognized only by static image. Testing of the algorithm is performed on a database of people. We used a subset of facial images from two database which consisted of 250 images and four categories in neutral, smile, sad and surprise. The algorithm presented in previous section is implemented with visual C#, and experiments are performed on a Intel Core 3.1 GHz PC with 4 GB RAM. The experiments shows the recognition results under different facial expressions such as smile, sad, surprise and neutral.
Figure 5: Examples of images for three subjects from sample database: Neutral facial emotion at first Row, and smiling for happy at the bottom.

Figure 6: Screenshot of the implemented system: face recognition and analysis for facial emotion

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented and implemented a simple approach for recognition of the facial expression analysis. The algorithm is performed two major steps: one is a detection of facial region with skin color segmentation and calculation of feature-map for extracting two interest regions focused on eye and mouth. And the other is a verification of the facial emotion of characteristic features with the Bezier curve and the Hausdorff distance. Experimental results show successful recognition of facial expressions, indicate good performance and applicability to mobile devices.

REFERENCES


