Detection and Recognition of Facial Emotion using Bezier Curves

Yong-Hwan Lee
Dankook University
152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 448-701, Korea
Tel: +82-10-3072-1458
hwany1458@empal.com

Abstract

Extracting and understanding of emotion is of high importance for the interaction among human and machine communication systems. The most expressive way to display the human's emotion is through facial expression analysis. This paper presents and implements an automatic extraction and recognition method of facial expression and emotion from still image. There are two steps to recognize the facial emotion; (1) Detecting facial region with color feature-map, and (2) Verifying the emotion of characteristic with Bezier curve. To evaluate the performance of the proposed algorithm, we assess the ratio of success with emotionally expressive facial image database. Experimental results shows average 78.8% of success to analyze and recognize the facial expression and emotion. The obtained result indicates the good performance and enough to applicable to mobile environments.

Keywords: Emotion Recognition, Face Recognition, Facial Expression Analysis, Bezier Curve.

1 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 [2]. 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 [3]. 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 [17].

This paper proposes a scheme 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.

This paper is organized as follows: next section reviews related works about emotion recognition through facial expression analysis. Section 3 presents the proposed system that uses two main step to recognize and classify the facial emotion from still image. Experimental results are shown in section 4, and section 5 summarizes the study and its future works.

IT CoNvergence PRActice (INPRA), volume: 1, number: 2, pp. 11-19

2 Related Works

A number of recently papers exist on automatic affect analyze and recognition of human emotion [6]. Research on recognizing emotion through facial expression was pioneered by Ekman [4], who started their work from the psychology perspective. Ira et al [3] proposed an architecture of Hidden Markov Models for automatically segment and recognize human facial expression from video sequences. Yashnari [10] investigated a method for facial expression recognition for a human speaker by using thermal image processing and a speech recognition system. 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. Spiros et al [8] suggested the extraction of appropriate facial features and consequent recognition of the user's emotional state that 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 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 [14] or profile view face images [13]. 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. Valstar et al [19] employs probabilistic, statistical and ensemble learning techniques, which seem to be particularly suitable for automatic action unit recognition from face image sequences.

3 Proposed Method

The proposed method for recognition of facial expression and emotion is composed of two major steps: first one is a detecting and analysis of facial area from original input image, and next is a verification of the facial emotion of characteristic features in the region of interest. A block diagram of our proposed method is depicted in Figure 1.

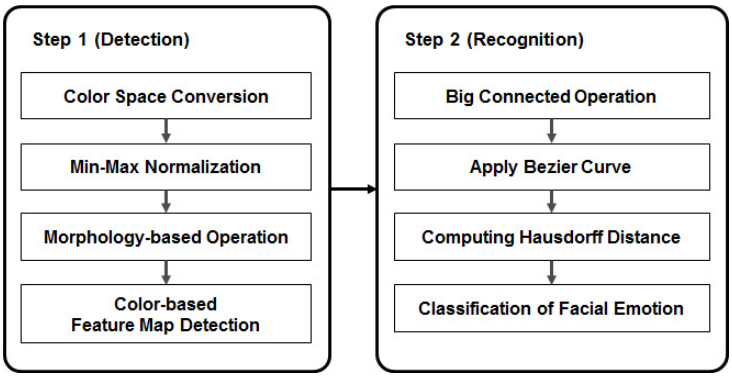

Figure 1: The proposed system flowchart

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 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 [5].

$$\begin{aligned} Y &= 0 + (0.299 \cdot R) + (0.587 \cdot G) + (0.114 \cdot B) \\ C_b &= 128 - (0.168736 \cdot R) - (0.331264 \cdot G) + (0.5 \cdot B) \\ C_r &= 128 + (0.5 \cdot R) - (0.418688 \cdot G) - (0.081312 \cdot B) \end{aligned} \tag{1}$$

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;

$$Y' = (\frac{y - min_1}{max_1 - min_1})\dot{(}max_2 - min_2) + min_2 \ , \ if(y \leq K_l \ or \ K_h \leq y) \tag{2}$$

where min_1 and max_1 are minimum and maximum value of Y component on input image, min_2 and max_2 are the value of the transformed space, $K_l = 30$ 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,

$$P_k(r_k) = \frac{n_k}{n} \ , \ 0 \leq r_k \leq 1 \ , \ k = 0, 1, \cdots, l-1 \tag{3}$$

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.

$$s_k = \sum_{j=0}^{k} \frac{n_k}{n} = E(r_k) = \sum_{j=0}^{k} P_r(r_j) \tag{4}$$

3.2 Eye and Mouth Detection using Feature Map

After reducing the illumination impact to the brightness component, we attempt to extract the region of eye and mouth on an image using *Eyemap* and *Mouthmap* in Equation (5) and Equation (6) [11]. Region of eyes would be easily found due to its intrinsic feature, namely symmetry. This paper restricts that both both eyes should be present inside the image to detect the skin region. As observed in [7], eyes are characterized by low red component and high blue one in the CbCr planes. Thus, *Eyemap* transformation is constructed by:

$$EyeMap = \frac{1}{3}(\alpha \cdot (C_b)^2 + \beta \cdot (\hat{C}_r)^2 + (\frac{C_b}{C_r})) \tag{5}$$

where $(C_b)^2$, $\hat{C}_r)^2$ and $\frac{C_b}{C_r}$ all are normalized to the range [0,255] and $\hat{C}_r$ is the negative of C_r, and α is greater than 1, β is less than 1 of positive constant which emphasizes to increase or decrease the red and blue component, because of Asian skin color, which generally have $R \geq G \geq B$ pattern.

Mouth is characterized by a high red component and low blue one, thus the mouth region has a relatively high response in the $(C_b)^2$ feature, low response in the $\frac{C_r}{C_b}$. *Mouthmap* is constructed by follow Equation (6).

$$MouthMap \;=\; C_r^2 \times (C_r - (\frac{\alpha \cdot C_r}{255 \cdot C_b})) \tag{6}$$

where $\alpha = 0.80$, which is the experiential value.

Figure 2: Detection of face boundary and features using Eyemap and Mouthmap; (a) Original input image, (b) Skin-tone extracted image, (c) Eyemap, (d) Mouthmap, and (e) Face boundary map

3.3 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 [16]. 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:

$$P(t) \;=\; \sum_{k=0}^{L} P_k BEZ_{k,L}(t) \tag{7}$$

where the Bezier blending function $BEZ_{k,L}(t)$ is known as the Bernstein polynomial, which is defined as [15]:

$$BEZ_{k,L}(t) \;=\; (\frac{L}{k}) t^k (1-t)^{L-k} \tag{8}$$

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

$$BEZ_{k,L}(t) \;=\; (1-t) \cdot BEZ_{k,L-1}(t) + t \cdot BEZ_{k,L-1} + t \cdot BEZ_{k-1,L-1}(t) \tag{9}$$

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:

$$\begin{aligned} x(t) \;&=\; \textstyle\sum_{k=0}^{L} x_k BEZ_{k,L}(t) \\ y(t) \;&=\; \textstyle\sum_{k=0}^{L} y_k BEZ_{k,L}(t) \end{aligned} \tag{10}$$

An example of a Bezier curve is given in Figure 3 with 4 control points [1].

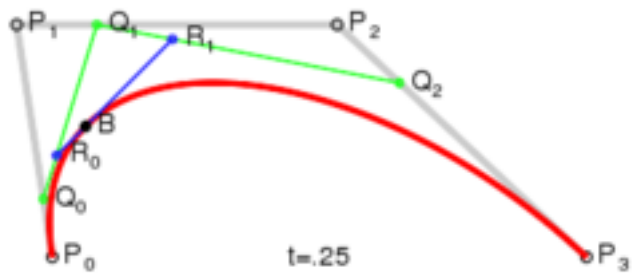

Figure 3: Examples of construction of the Bezier curve

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.

3.4 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 $q(t), t \in [c,d]$ is given in Equation (10) [9].

$$d_H(p,q) \;=\; max\{\max_{s\in[a,b]}\min_{t\in[c,d]}|p(s)-q(t)|,\ \max_{t\in[c,d]}\min_{s\in[a,b]}|p(s)-q(t)|\} \tag{11}$$

4 Experiments and Results

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. The ground truth set for estimating the performance of the algorithm is provided with the categories in the experiments, which are correct if the decision is belonged to the correct category. Figure 5 shows how to move the control points across different subjects (e.g., neutral and smile) and to interpret a tracking facial features with Bezier curve and extracted feature points interpolation.

To categorize facial emotion, we need first to determine the expressions from movements of facial control points. Ekman *et al.* [4] have produced a system for describeing visually distinguishable facial movements (called *Facial Action Coding System, FACS*), which is based on enumeration of all action

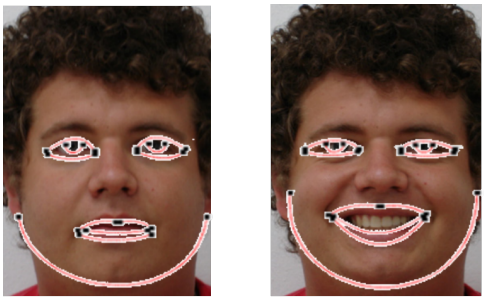

Figure 4: Bezier curve and extracted feature points interpolation across different subjects. (a) line fitting over example image in neutral condition, and (b) in smile condition

Table 1: Rule of facial emotion recognition

Emotions	Movements of AUs
Smile	Eye opening is narrowed, Mouth is opening, and Lip corners are pulled obliquely
Sad	Eye is tightened closed, and and Lower lip corner depressor
Surprise	Eye and Mouth are opened, Upper eyelid raiser, and Mouth stretch

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.

In this work, the facial expressions have been recognized only by static image. Testing of the algorithm is performed on a database of people, which is obtained from FEI face database [18] and JAFFE database [12]. There are 14 images for each of 200 individuals, total of 2,800 images in FEI face database. All images are colorful and taken against a white homogenous background in an upright frontal position. Additionally, they provide a subset of facial images, totally 400 images, each subject has two frontal images (one with neutral and the other with smiling facial expression). JAFFE database contains 213 images of 7 facial expressions posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. We used a subset of facial images from two database which consisted of 250 images and four categories in neutral, smile, sad and surprise. Figure 5 shows sample facial images which expresses an emotion, for example, by neutral and smiling expression. 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, as shown in Figure 6.

The experiments shows the recognition results under different facial expressions such as smile, sad, surprise and neutral. The proposed method recognizes 197 of the 250 faces, which means that a successful recognition ratio of 78.8% is achieved, as shown in Table 2. Experimental result reveals that success ratio is better for smile, because the control points and curves for smile are given more appreciable changes from neutral expression. Among the 53 false dismissals cases, some of faces are gray-scale images from JAFFE database, which leads hard to detect the skin color pattern of facial region in the image.

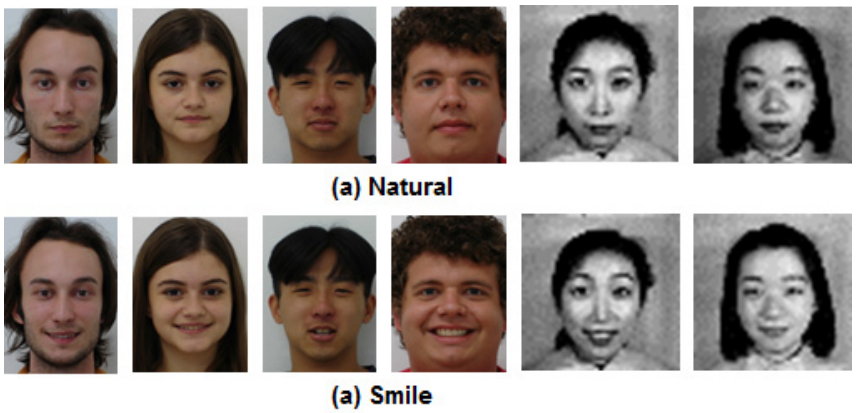

Figure 5: Examples of images for three subjects from sample database: Neutral facial emotion at first row, and smiling for happy at the bottom

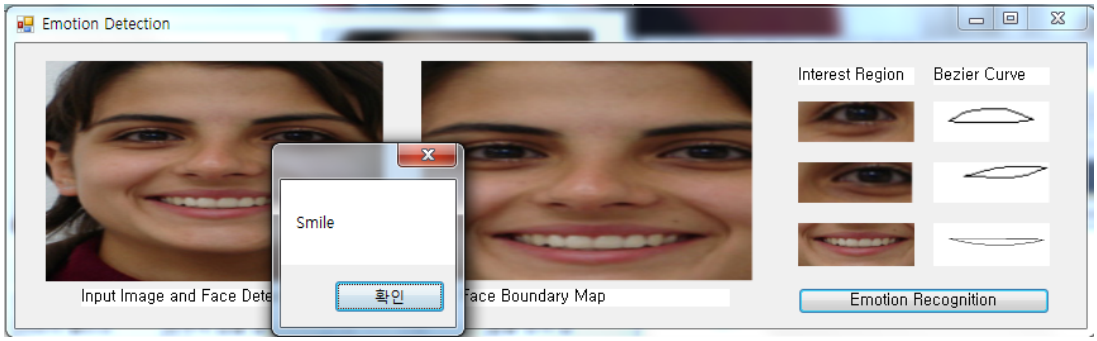

Figure 6: Screenshot of the implemented system: face recognition and analysis for facial emotion

5 Conclusion

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 shows average successful ratio of 78.8% to recognize the facial expression, and this indicates the good performance and enough to applicable to mobile devices.

Table 2: Results of facial expression recognition

Expression	Corrects / Misses	Success Ratio
Smile	67 / 12	84.8%
Sad	22 / 8	73.3%
Surprise	45 / 16	73.8%
Neutral	63 / 17	78.8%
Total	197 / 53	78.8%

References

[1] Bézier curve. Wikipedia, 2013. `http://en.wikipedia.org/wiki/Bezier_curve`.

[2] R. Adolphs. Recognizing Emotion From Facial Expressions: Psychological and Neurological Mechanisms. Behavioral and Cognitive Neuroscience Reviews, 2002.

[3] I. Cohen, A. Garg, and T. S. Huang. Emotion Recognition from Facial Expressions using Multilevel HMM. In *Proc. of the 2000 NIPS Workshop on Affective Computing, Denver, Colorado, USA*, December 2000.

[4] P. Ekman and W. Friesen. Facial Action Coding System (FACS): Investigator's Guide. Conculting Psychologists Press, 1978.

[5] R. C. Gonzalea, R. E. Woods, and S. L. Eddins. *Digital Image Processing using Matlab*. Prentice Hall, 2004.

[6] G. H. and P. M. Automatic Temporal Segment Detection and Affect Recognition from Face and Body Display. *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, 39(1):64–84, 2009.

[7] R.-L. Hsu, M. Abdel-Mattaleb, and A. K. Jain. Face Detection in Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):696–701, 2002.

[8] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis, and S. D. Kollias. Emotion Recognition through Facial Expression Analysis based on a Neurofuzzy Network. *Neural Networks*, 18(4):423–435, 2005.

[9] S.-H. Kim and Y. J. Ahn. An Approximation of Circular Arcs by Quartic Bezier Curves. *Computer-Aided Design*, 39(6):490–493, 2007.

[10] Y. Koda, Y. Yoshitomi, M. Nakano, and M. Tabuse. Facial Expression Recognition for Speaker using Thermal Image Processing and Speech Recognition System. In *Proc. of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'09), Toyama, Japan*, pages 955–960. IEEE, September-October 2009.

[11] Y.-H. Lee, T.-K. Han, Y.-S. Kim, and S.-B. Rhee. An Efficient Algorithm for Face Recognition in Mobile Environments. *Asian Journal of Information Technology*, 4(8):796–802, 2005.

[12] M. J. Lyons, M. Kamachi, and J. Gyoba. Japanese Femail Facial Expressions (JAFFE). Database of digital images, 1997.

[13] P. M. and I. Patras. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man and Cybernetics*, 36(2):433–449, 2006.

[14] P. M. and L. Rothkrantz. Facial Action Recognition for Facial Expression Analysis from Static Face Images. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, 34(3):1449–1461, 2004.

[15] M. R. Rahman, M. A. Ali, and G. Sorwar. Finding Significant Points for Parametric Curve Generation Techniques. *Journal of Advanced Computations*, 2(2), 2008.

[16] B. Sam and B. R. Samuel. *3D Computer Graphics - A Mathematical Introduction with OpenGL*. Cambridge University Press, 2003.

[17] N. Sebe, I. Cohen, T. Gevers, and T. S. Huang. Multimodal Approaches for Emotion Recognition: A Survey. In S. Santini, R. Schettini, and T. Gevers, editors, *Internet Imaging VI, SPIE*, volume 5670, pages 56–67. Society of Photo Optical, January 2005.

[18] C. Thomaz and G. Giraldi. A New Ranking Method for Principal Components Analysis and its Application to Face Image Analysis. *Image and Vision Computing*, 28(6):902–913, 2010.

[19] M. Valster, P. M., Z. Ambadar, and J. Cohn. Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions. pages 162–170. ACM, November 2006.

Author Biography

Yong-Hwan Lee received the M.S. degree in Computer Science and Ph.D. in Electronics and Computer Engineering from Dankook University, Korea, in 1995 and 2007, respectively. Currently, he is a research professor at the Department of Applied Computer Engineering, Dankook University, Korea. His research areas include Image/Video Representation and Retrieval, Face Recognition, Augmented Reality, Mobile Programming and Multimedia Communication.