ISSN No. 0976-5697

Volume 2, No. 5, Sept-Oct 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Face Emotion Recognition Using Optimality Parameters Eye and Lip in Different Geographical Area

Ahmad Habibizad Navin*
Department of Computer, Science and Research Branch, Islamic Azad University, East Azerbaijan, Tabriz, Iran
habibi@iauotash.ac.ir

Mehdi Akhari Oskuyee
Department of Computer, Tabriz, Islamic Azad University, Tabriz, Iran
sadegnet@yahoo.com

Mirkamal Mirnia
Department of Computer, Science and Research Branch, Islamic Azad University, East Azerbaijan, Tabriz, Iran
mirnia-kam@tabrizu.ac.ir

Nima Jafari Navimipour
Department of Computer, Science and Research Branch, Islamic Azad University, East Azerbaijan, Tabriz, Iran
n.jafari@srbiau.ac.ir

Abstract: The detection of emotion is becoming an important field for human – computer Interaction. Emotion recognition can be achieved by a number of methods such as facial expressions, vocal, gesture and physiology signal recognition. Facial expressions possess a number of advantages against other emotion recognition methods and suitable for a number of environmental conditions. In this paper, we describe a procedure for face emotion recognition through eye and lip optimal features. This process involves four stages: pre-processing, feature extraction, classification and comparison of emotions. Firstly a series of pre- processing tasks such as adjusting contrast, filtering, skin color segmentation and edge detection are done. One of the important tasks at this stage after pre- processing is to extract features. To extract features with high speed projection profile is used. Second particle swarm optimization (PSO) is used to optimize eye and lip ellipse characteristics. In the third stage, with using the features obtained of the optimal ellipse eye and lip, a person emotion according to experimental results and emotions represented by Ekman (sadness, angry, joy, fear, disgust and surprise without consider natural emotion) is classified. Finally, we will compare the Indian and Japanese people face emotions by PSO algorithm and also present experimental results to test. A major advantage of this method is that for each geographical area can used. To use this method in a particular geographic area, the standard parameters of the same geographic area must be set. In this case the comparison between Indian and Japanese shows that accuracy for Japanese is relatively lower than Indian.

Keywords: Feature extraction, Projection profile, Eye and lip ellipse, Emotions classification, Particle swarm optimization algorithm, Emotion recognition**.**

I. INTRODUCTION

Recognizing human facial expression and emotion by computer is an interesting and challenging problem. Facial expressions provide rich information about human emotion and play an essential role in human communication. A human can express his/her emotion through lip and eye. A category of emotions which universally developed by Ekman are: sadness, angry, joy, fear, disgust and surprise without consider natural emotion. The main purpose of this paper is whether the method face emotion recognition using eye and lip for each geographical area and every nationality is possible. This method consists of mainly four parts. The first part describes various stages in image processing include preprocessing, filtering, edge detection and projection profile is used to extract features. The second part discusses a PSO- based approach to optimize the eye and lip ellipse characteristics. In the third part, we using of the eye and lip optimal parameters to classify the emotions. Finally, we will compare the face emotions between the two countries. The general process for emotion recognition is shown in Fig. 1. The rest of this paper organized as follows. Section 2 is an overview of related works. The method with PSO algorithm is described in section 3. Efficiency analysis and results of the method is discussed in section 4 and section 5 contains conclusions.

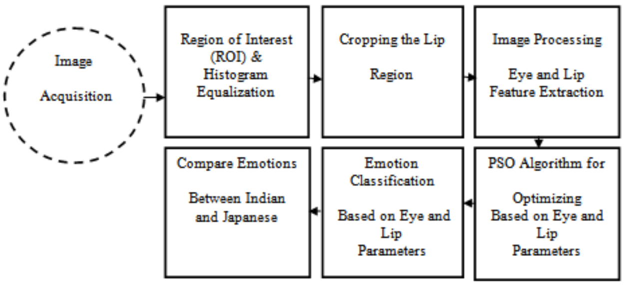

Figure1. Emotion recognition process

II. RELATED WORKS

Facial expressions give important clues about emotions. Therefore, several approaches have been proposed to classify human affective states. The features used are typically based on local spatial position or displacement of specific points and regions of the face, unlike the approaches based on audio, which use global statistics of the acoustic features. For a complete review of recent emotion recognition systems based on facial expression the readers are referred to [1]. Mase proposed an emotion recognition system that uses the major directions of specific facial muscles [2]. With 11 windows manually located in the face, the muscle movements were extracted by the use of optical flow. For classification, K-nearest neighbor rule was used, with an accuracy of 80% with four emotions: happiness, anger, disgust and surprise. Yacoob et al. proposed a similar method [3]. Instead of using facial muscle actions, they built a dictionary to convert motions associated with edge of the mouth, eyes and eyebrows, into a linguistic, per- frame, mid-level representation. They classified the six basic emotions by the used of a rule-based system with 88% of accuracy. Black et al. used parametric models to extract the shape and movements of the mouse, eye and eyebrows [4]. They also built a mid- and high-level representation of facial actions by using a similar approach employed in [3], with 89% of accuracy. Tian et al. attempted to recognize Actions Units (AU), developed by Ekman and Friesen in 1978 [5], using permanent and transient facial features such as lip, nasolabial furrow and wrinkles [6]. Geometrical models were used to locate the shapes and appearances of these features. They achieved a 96% of accuracy. Essa et al. developed a system that quantified facial movements based on parametric models of independent facial muscle groups [7]. They modeled the face by the use of an optical flow method coupled with geometric, physical and motion-based dynamic models. They generated spatial-temporal templates that were used for emotion recognition. Without considering sadness that was not included in their work, a recognition accuracy rate of 98% was achieved. A method that extracts region of eye and lip of facial image by genetic algorithm has been suggested recently [8]. The obtained results show that the success rate and running speed in face emotion recognition using eye and lip by particles swarm optimization algorithm in comparison with the genetic algorithm has better performance

III. THE PROPOSED METHOD

The main purpose of this paper is whether the method face emotion recognition using eye and lip for each geographical area and every nationality is possible. The proposed method in this paper is design a optimize method to emotion recognition for the Indian and Japanese by using particle swarm optimization. In the end compare the results. As the first step in image processing, for skin color segmentation, first we contrast the image. A histogram equalization method has been applied. This method usually increases the global contrast of many images, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities can be better distributed on the histogram. This allows for areas of lower local contrast to gain a higher contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values. The histogram equalized image is filtered using average and median filters in order to make the image smoother. And we have to find the largest connected region. Then we have to check the probability to become a face of the largest connected region. If the largest connected region has the probability to become a face, then it will accept as the largest connected region. If the largest connected regions height & width is larger or equal than 50 and the ratio of height/width is Between 1 to 2, then it may be face. For face detection, first we convert binary image from rgb image. For converting binary image, we calculate the average value of rgb for each pixel and if the average value is below than 110, we replace it by black pixel and otherwise we replace it by white pixel. By this method, we get a binary image from rgb image. Then, we try to find the forehead from the binary image. We start scan from the middle of the image, then want to find a continuous white pixels after a continuous black pixel. Then we want to find the maximum width of the white pixel by searching vertical both left and right site. Then, if the new width is smaller half of the previous maximum width, then we break the scan because if we reach the eyebrow then this situation will arise. Then we cut the face from the starting position of the forehead and its high will be 1.5 multiply of its width. X will be equal to the maximum width of the forehead. Then we will have an image which will contain only eyes, nose and lip. Then we will cut the rgb image according to the binary image. For eyes detection, we convert the rgb face to the binary face. Now, we consider the face width by W. We scan from the W/4 to (W-W/4) to find the middle position of the two eyes. The highest white continuous pixel along the height between the ranges is the middle position of the two eyes. Then we find the starting high or upper position of the two eyebrows by searching vertical. For left eye, we search w/8 to mid and for right eye we search mid to w – w/8. Here w is the width of the image and mid is the middle position of the two eyes. There may be some white pixels between the eyebrow and the eye. To make the eyebrow and eye connected, we place some continuous black pixels vertically from eyebrow to the eye. For left eye, the vertical black pixel-lines are placed in between mid/2 to mid/4 and for right eye the lines are in between mid+(w-mid)/ 4 to mid+3*(w-mid)/ 4 and height of the black pixel-lines are from the eyebrow starting height to (h- eyebrow starting position)/4. Here w is the width of the image and mid is the middle position of the two eyes and h is the height of the image. Then we find the lower position of the two eyes by searching black pixel vertically. For left eye, we search from the mid/4 to mid - mid/4 width. And for right eye, we search mid + (w-mid)/ 4 to mid+3*(w- mid)/ 4 width from image lower end to starting position of the eyebrow. Then we find the right side of the left eye by searching black pixel horizontally from the mid position to the starting position of black pixels in between the upper position and lower position of the left eye. And left side for right eye we search mid to the starting position of black pixels in between the upper position and lower position of right eye. The left side of the left eye is the starting width of the image and the right side of the right eye is the ending width of the image. Then we cut the upper position, lower position, left side and the right side of the two eyes from the rgb image. For lip detection, we determine the lip box. And we consider that lip must be inside the lip box. So, first we determine the distance between the forehead and eyes. Then we add the distance with the lower height of the eye to determine the upper height of the box which will contain the lip. Now, the starting point of the box will be the ¼ position of the left eye box and ending point will be the ¾ position of the right eye box. And the ending height of the box will be the lower end of the face image. So, this box will contain only lip and may some part of the nose. Then we will cut the rgb image according the box. Finally, sobel edge detection method due to the high speed and small volume of calculations is applied to the eyes and lip image. The sobel edge detection region of lip and eye region are shown in Fig. 2 and Fig. 3.

Figure 2. The surprise and happy emotion

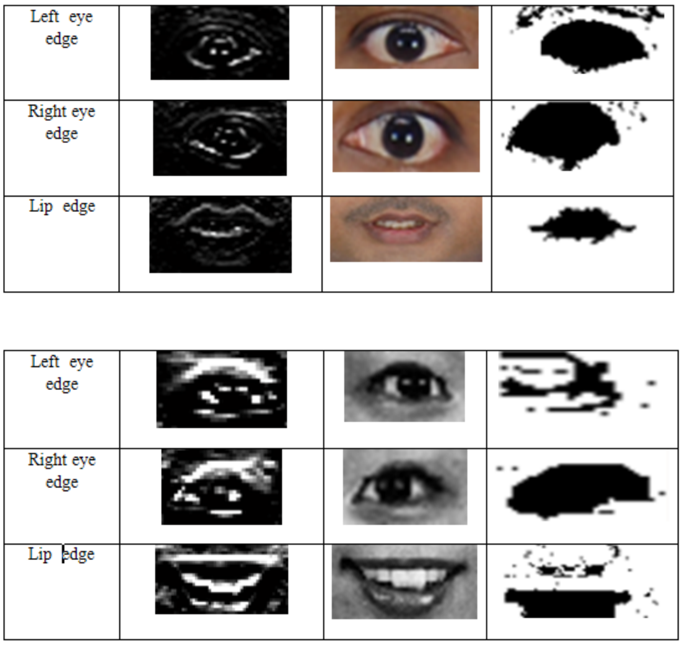

Figure3. Sobel edge detected [9] ,[10].

Feature extraction method is associated with the row-sum and column-sum of white pixels of edge identified image. The pattern of row-sum (M_h) along the column and the pattern of column-sum (M_v) along the row of white pixels are defined as the feature of each region. These patterns are known as projection profiles. Let f (m, n) represents a binary image of m rows and n columns. Then the vertical profile is defined as the sum of white pixels of each column perpendicular to the x-axis which is represented by the vector M_v of size n by (1).

$$M_{vj} = \sum_{i=1}^{m} f(i,j) \quad j = 1,2,3,\ldots\ldots,n$$

The horizontal profile is the sum of white pixels of each row perpendicular to the y-axis which is represented by the vector M_h of size m is calculated by (2).

$$M_{hi} = \sum_{j=1}^{n} f(i,j) \quad i = 1,2,3,\ldots\ldots,m$$

The human eye shape is more like an ellipse (we call this as a regular ellipse), as shown in Fig4. The minor axis is a feature of the eye that varies for each emotion. The major axis of the eye is more or less fixed for a particular person in varied emotions. The ellipse is parameterized by its minor and major axes, respectively, as "2a" (fixed) and "2b" (to be computed) is described by (3).

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

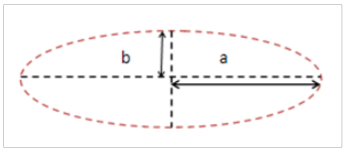

Figure4. The regular ellipse

The shape of human lip towards a combination of two ellipses which is called an irregular ellipse, as shown in Fig5. The word 'irregular' means that the ellipse has two different minor axes wherein the major axes remains the same. The edge detected lip image is considered as an irregular ellipse. Lengths of minor axes of the lip feature for each emotion are computed. The major axis is "2a" (considered to be fix) and two minor axes are "2b1" and "2b2" (to be computed). The suitable values to b1 and b2 are substituted for top and bottom portions respectively. Emotional state on an Image strongly depends on facial expression b1, b2 as expression of lip and b as expression of eye. In the next section PSO algorithm adopted to optimize these expressions.

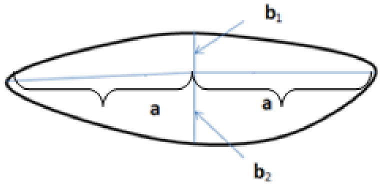

Figure5. The irregular ellipse

Particle swarm optimization as stated before, PSO simulates the behaviors of bird flocking. Suppose the following scenario: a group of birds are randomly searching food in an area. There is only one piece of food in the area being searched. All the birds do not know where the food is. But they know how far the food in each iteration. So what's the best strategy to find the food? The effective one is to follow the bird which is nearest to the food. PSO learned from the scenario and used it to solve the optimization problems. In PSO, each single solution is a "bird" in the search space. We call it "particle". All of particles have fitness values which are evaluated by the fitness function to be optimized, and have velocities which direct the flying of the particles. The particles fly through the problem space by following the current optimum particles. PSO is initialized with a group of random particles (solutions) and then searches for optimal by updating generations. For every iteration each particle is updated by following two "best" values. The first one is the best solution (fitness) it has achieved so far. (The fitness value is also stored.) This value is called pbest. Another "best" value that is tracked by the particle swarm optimization is the best value, obtained so far by any particle in the population. This best value is a global best and called gbest. When a particle takes part of the population as its topological neighbors, the best value is a local best and is called lbest. After finding the two best values, the particle updates its velocity and positions with following equation (4) and (5).

v [] = v [] + c_1 * rand () * (pbest [] - present []) + c_2 * rand () * (gbest [] - present []) *(4)*

present [] = present [] + v [] *(5)*

v [] is the particle velocity, present [] is the current particle (solution). pbest [] and gbest [] are defined as stated before. Rand () is a random number between (0, 1). c1, c2 are learning factors. Usually $c_1 = c_2 = 2$.

Algorithm parameters are shown in Table I.

Table I. Parameter settings for PSO processing

Description	*Parameter*
particle	(x, x_1, x_2)
number of particles	200
Dimension of particles	3
Range of particles	It is also determined by the problem to be optimized. X_1>=0 and X_2<=0
V_{max}=20	It determines the maximum change one particle can take during one iteration.
Learning factors	Usually c1 equals to c2 and ranges from [0, 2].
stop condition	The maximum number of iterations the PSO execute and the minimum error requirement.(one miss-classified)
version	local version is a little bit slower but not easy to be trapped into local optimum.(refine the search)
inertia weight	W_{max}=0.9 , W_{min}=0.4
max iteration number	500
W(iteration)	W_{max}-((W_{max}- W_{min})/ max iteration)* iteration

IV. EXPERIMENTAL RESULTS

In this study on Indian and Japanese subjects, seven emotions and 350 images were examined. The eye and lip features have been given as input to the PSO algorithm to find the optimized values (ellipse optimum). Optimization process was repeated 20 times for each emotion. Thereupon optimal parameters (x, x_1, x_2) come from of optimal ellipsoid axes. In Table II manual measured parameters and PSO optimized parameters (The mean of the parameters) from 350 images Indian are shown. Sample images of Indian are shown in [9]. In Table III manual measured parameters and PSO optimized parameters (The mean of the parameters) from 350 images Japanese are shown. Sample images of Japanese are shown in [10]. By comparing Table II and Table III:

- Various parameters of the eye and lip are used to face emotion recognition.
- In this case the comparison between Indian and Japanese accuracy for Japanese is relatively lower than Indian.
- This method can be implemented for each geographic area with standard parameters.

Table II. Manual and PSO optimal measured parameters for Indian

Emotion	*Manually Computed Mean Value (in pixels)*			*Optimized Mean Value by PSO (in pixels)*			*50 Images For each emotion*	*Duration of Emotion Recognition (sec)*
	b_1	b_2	b	x_1	x_2	x	**Success Rate**	**Mean Time**
Natural	40	44	25	39.8165	43.2366	24.9852	93%	48
Fear	27	44	21	26.2525	43.6355	19.6565	89%	39
Happy	27	50	20	26.9612	48.2256	19.6353	92%	51
Sad	28	37	22	27.1464	36.5598	21.9751	88%	49
Angry	27	36	19	26.1256	35.2684	18.6521	94%	53
Dislike	37	32	18	35.2565	31.2255	17.9850	87%	39
Surprise	46	60	20	45.9680	58.2685	19.1451	94%	52

Table III. Manual and PSO optimal measured parameters for Japanese

Emotion	*Manually Computed Mean Value (in pixels)*			*Optimized Mean Value by PSO (in pixels)*			*50 Images For each emotion*	*Duration of Emotion Recognition (sec)*
	b_1	b_2	b	x_1	x_2	x	**Success Rate**	**Mean Time**
Natural	26	22	18	25.1454	21.8449	17.9852	90%	42
Fear	19	27	17	18.8675	26.0782	15.6565	86%	31
Happy	21	32	23	20.5266	30.8992	22.6353	90%	42
Sad	22	25	27	21.5544	24.7565	26.9751	96%	38
Angry	21	23	25	20.1091	21.9728	24.6521	96%	45
Dislike	21	24	21	20.0370	23.5860	20.9850	86%	34
Surprise	29	29	23	28.7725	28.2365	21.1451	86%	43

V. CONCLUSION AND FUTURE WORKS

This paper discusses the application of sobel filter based feature extraction in combination with PSO algorithm for the recognition of seven different facial emotions (happy, sad, angry, fear, neutral, surprise and dislike). Current methods for emotion recognition are: facial expressions, vocal, gesture and physiology signal recognition. In related works several methods were investigated for the facial expressions. The main purpose of this paper is whether the method face emotion recognition using eye and lip for each geographical area and every nationality is possible. Firstly a series of pre- processing tasks such as adjusting contrast, filtering, skin color segmentation and edge detection are done. One of the important tasks at this stage after pre- processing is to extract features. To extract features with high speed projection profile is used. Second eye and lip features are given as input to the PSO to compute the optimized values of b, b_1 and b_2. In the third stage with using the features obtained of the optimized ellipse eye and lip, a person emotion according to results Table II and Table III is classified. By comparing Table II and Table III reached the following conclusions:

- Various parameters of the eye and lip are used to face emotion recognition.

- This method can be implemented for each geographic area with standard parameters.
- In this case the comparison between Indian and Japanese accuracy for Japanese is relatively lower than Indian.

Future work:

- Standard table can be created for each geographic area. Each table contains standard parameters are that in all geographical areas these values are different.
- With giving an image and identify its nationality, Tables can be used to face emotion recognition.

VI. REFERENCES

[1] Pantic, M., Rothkrantz, L.J.M. Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE , Volume: 91 Issue: 9 , Sept. 2003. Page(s): 1370 – Volume: 91 Issue: 9 , Sept. 2003. Page(s): 1370 –1390.

[2] Mase K. Recognition of facial expression from optical flow. IEICE Transc., E. 74(10):3474–3483, 0ctober 1991.

[3] Yacoob, Y., Davis, L. Computing spatio-temporal representations of human faces. Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on , 21-23 June 1994 Page(s): 70 –75.

[4] Black, M. J. and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. In Proceedings of the International Conference on Computer Vision, pages 374–381. IEEE Computer Society, Cambridge, MA, 1995.

[5] Ekman, P., Friesen, W. V. Facial Action Coding System: A Technique for Measurement of Facial Movement. Consulting Psychologists Press Palo Alto, California, 1978.

[6] Tian, Ying-li, Kanade, T. and Cohn, J. Recognizing Lower Face Action Units for Facial Expression Analysis. Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), March, 2000, pp. 484 – 490.

[7] Essa, Pentland, A. P. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transc. On Pattern Analysis and Machine Intelligence, 19(7):757–763, JULY 1997.

[8] Hideaki Tani, Kenji Terada, Shunichiro Oe and Junichi Yamaguchi, “Detecting of one’s eye from facial image by using genetic algorithm”, The 27thAnnual Conference of the IEEE Industrial Electronics Society, 2001, pp.1937-1940.

[9] http://up6.iranblog.com/images/g8lvr74ipnz575yh8kx.pdf

[10] http://www.kasrl.org/jaffe.html

[11] Rafael C Gonzalez and Richard E Woods, “Digital Image Processing”, Pearson Education, Inc, India, 2002.

[12] Keith Anderson and Peter W. McOwan, “A Real-Time Automated System for the Recognition of Human Facial Expressions”, IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, February 2006, 36(1), pp.96-105.

[13] Maja Panti and Ioannis 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—Part B: Cybernetics, April 2006, 36(2), , pp.433-449.

[14] Hironori Takimoto, Yasue Mitsukura, Minoru Fukumi and Norio Akamatsu, “Face Detection And Emotional Extraction System Using Double Structure Neural Network”, Proc. of the International Joint Conference on Neural Networks, 20-24 July 2003, pp.53-57.

[15] Picard, R., & Klein, J. (2002). Computers that recognize and respond to user emotion: Theoretical and practical implications. Interacting With Computers, 14, 141-169.

[16] Zeng, Z. et al. 2004. Bimodal HCI-related Affect Recognition. ICMI’04, October 13–15, 2004, State College, Pennsylvania, USA.

[17] Rafael A. Calvo, Sidney D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," IEEE Transactions on Affective Computing, pp. 18-37, January-June, 2010.

[18] Zhihong Zeng; Pantic, M.; Roisman, G.I.; Huang, T.S.; , "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.31, no.1, pp.39-58, Jan. 2009.