IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 1, January 2018

A novel approach to Real Time Human Emotion Recognition based on displacement of facial feature points

Anurag De[1], Dr. P. Satheesh[2], Dr. B. Srinivas[3], M. Vamsi Krishna[4]

Assistant Professor, CSE, MVGRCE, Vizianagaram, India [1, 4]

Associate Professor, CSE, MVGRCE, Vizianagaram, India [2, 3]

Abstract: Human emotion recognition is one of the vital areas of research in the broad field of image processing and human-computer interaction. In this paper, an implementation methodology is proposed for designing a real-time emotion recognition system through facial expression detection. In implementing this system, a webcam is used to capture the frontal face of the person and the YCbCr color model is used to detect the face part in the captured image and then various facial features are extracted from the face image and based on the distances measured between the various feature points of the face, the facial expression is detected and eventually classified into six basic emotions such as happiness, surprise, anger, fear, sorrow and neutral. The classification results varies in a time function thus reducing the latency between the man and machine making the system behave in a real-time manner.

Keywords: Real-time, YCbCr color model, Facial Expression, Emotion Recognition, Human-Computer Interaction.

I. INTRODUCTION

A Facial Expression of a human being conveys his emotional state to a person with whom he is communicating or to any observer. This emotional state is generated due to the movement of the facial muscles lying beneath the skin of the person. Emotion Recognition through facial expression detection is one of the new concepts which is becoming very crucial in the wide field of research in man-machine interaction or Human Computer communication. Facial expression recognition refers to classification of the facial features into six basic emotions i.e. happiness, sadness, fear, disgust, surprise and anger, as introduced by Ekman [1-4]. This perception is assumed due to the fact that the appearance of human emotions is universal across peoples of different ethnics, races and cultures. This paper proposes an implementation methodology of real-time emotion recognition. Real-time emotion recognition is very much important to reach higher levels of Human Computer Interaction by removing the latency between man and machine and making the system more robust and reducing the processing time eventually delivering a real-time output of the emotion recognized [5]. Here in this paper, the person whose emotion is to be recognized expresses his emotions through facial expressions in front of a webcam and based on the facial feature extraction and distances measured between the various feature points of the face, the facial expression is recognized and classified into emotions such as: happiness, surprise, anger, fear, sorrow and neutral.

II. PROPOSED METHODOLOGY

This section states the various phases of this real-time emotion recognition system. The first phase is to live stream the video in real time for acquiring the input image frames in real time by live video feed through a webcam till the required frame is captured [6]. The next step proceeds for detecting the face region in the captured image frame by identifying the skin color pixels. The next phase deals with segmenting the face region by elimination of unnecessary areas of the skin region which are not necessary for processing. Then we proceed towards segmenting or detecting the eye region from the skin region identified which occurs to be a very important feature for facial expression recognition. Once the eyes region is segmented the next phase is to detect the mouth region including the lips area as because the mouth is a very important feature which involves in producing various facial expressions. After all the feature needed are segmented, we can now proceed with finding out the longest binary pattern which results in converting the non-skin color pixels to black pixels and skin color pixels to white pixels. As because non-skin pixels will be those of the eye and lip area and these regions are most important for producing facial expressions, the Bezier curve equation is thus applied on the mouth and lip region of the face to find out the curves or features needed to calculate the feature point displacements and accordingly perform the emotion recognition by comparing the test image displacements with the dataset images designated for various expressions and the closest match will thus be selected as the desired emotion for a particular test image in real time. Figure 1 shows the flowchart of the proposed methodology for implementing this system.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified

Vol. 7, Issue 1, January 2018

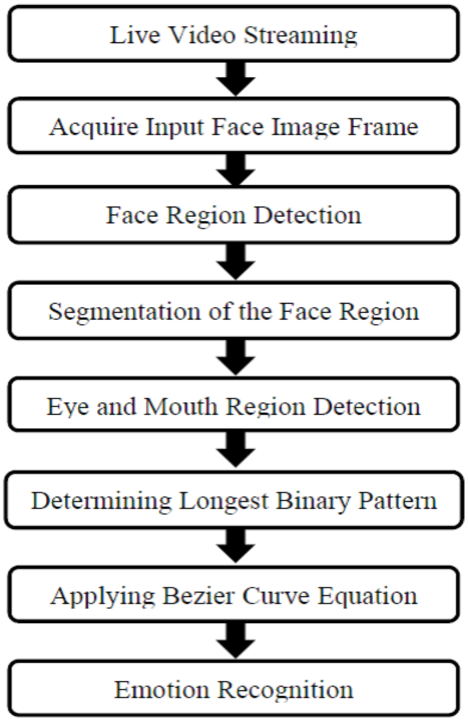


Fig. 1 Proposed Methodology

III. FACE REGION DETECTION

The initial objective of the system is to detect the face area from the input image frame captured by the webcam, for which the YCbCr (luminance (Y channel) and chrominance (Cb and Cr channels)) color model is used to extract the skin color region of the person's image. Thus the YCbCr color model is used to detect the skin area of the person's face. Three parameters, namely y, Cb and Cr are been identified based on the following formula (1), (2) and (3) to convert the image from RGB to YCbCr [7].

$$y = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (1)$$

$$Cb = -0.172 \times R - 0.339 \times G + 0.511 \times B + 128 \quad (2)$$

$$Cr = 0.511 \times R - 0.428 \times G - 0.083 \times B + 128 \quad (3)$$

For each face pixel the computation is done based on the above equations. Therefore, a face pixel is denoted as a skin pixel provided the values for parameters y, Cb and Cr of the pixel satisfies the following three inequalities, which are $y > 80$, $85 < Cb < 135$ and $135 < Cr < 180$.

Parts of the body other than the face and neck region present in the input image frame also comes under the skin regions as the skin region detection is solely based on the color value matching which denotes the pixels as skin and non-skin pixels. Figure 2 shows the extracted skin region.

Fig. 2 Extracted skin area

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 1, January 2018

IV. SEGMENTATION OF THE FACE AREA

The next objective is to segment the face region from it by elimination the unnecessary parts of the skin region. Therefore, to remove the unnecessary skin regions, the column wise sum is being computed for all the columns. Next, the non-zero column sum windows are marked by grouping the adjacent columns having non-zero column sum value. Similarly, to find out the maximum row window same operations are carried out row wise based on the non-zero row sum values. In this way, the skin color patches, spots and unwanted skin parts are removed from the face image. Figure 3 displays the segmented skin region after elimination of the unwanted skin parts [8].

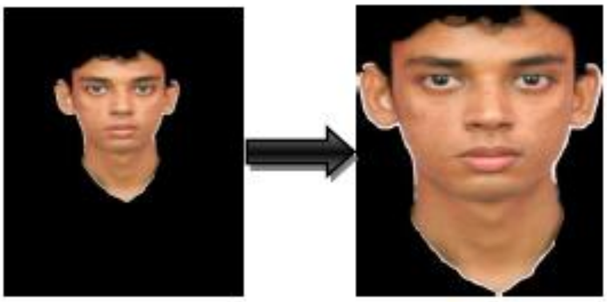


Fig. 3 Segmented face area

V. EYE AREA DETECTION

After the face region is segmented by removal of the unnecessary skin area we need to locate the eyes of the person which are a very important feature for recognition of emotions by the system. Here both the left and right eye are detected. To locate the eye region a drastic change in color is being considered. Such as, when the pixel intensity values from the forehead region is marked, the eyebrow part possess a separate color than the forehead area, thus, minimizing the search region by redefining the minimum and maximum column values. Then the column-wise sum of pixel values is computed for each of the columns. It again further minimizes the search region by redefining the minimum and maximum column values. Next, we have to determine the sharp color changes for which we take the row sums enclosed by the column boundaries. Then we calculate the gradient between two consecutive row sum values such as difference between sums of i^{th} and $(i+1)^{th}$ row. Picking up the ten maximum gradients and their corresponding row values, it is found that the minimum row value among these values returns the row locating the eyebrows. Figure 4 displays the detected eye region [8].

Fig. 4 Detected Eye area

In case, if the left eyebrow and right eyebrow are not aligned in the same row, the row value of the upper eyebrow is picked up to determine the upper boundary of the search area. The lower limit of the search area will be bounded by upper limit in addition to half of the column window width. After the eye region is detected, it is clipped and the left and right eye are segmented for further processing. Figure 5 shows the segmented left and right eye respectively.

(a) (b)

Fig. 5 Segmented eye (a) Left eye (b) Right eye

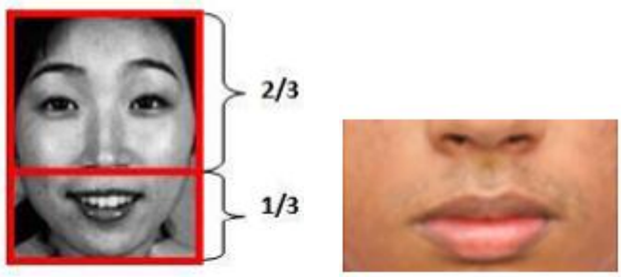
IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 1, January 2018

VI. MOUTH AREA DETECTION

From the study of the geometry of human faces, it is found that the eyes and the eye-brows are present in the upper two-third portion of the human face and the mouth is located in the lower one-third portion of the face. Using this knowledge of facial geometry, the mouth region is detected in the lower one-third position of the face which may contain some portions of the lower nose as well. Figure 6 displays the facial geometry of the human face upon an expression image taken from the jaffe database and the detected mouth region respectively [8].



(a) (b)

Fig. 6 (a) Facial geometry (b) Detected mouth area

VII. DETERMINING LONGEST BINARY PATTERN

This phase deals with determining the longest binary pattern for the eyes and lips in order to segment the lip part and eye part for further processing. The lip is the most varied region of the face whose texture varies in evenness, tone and color. Thus it is very important to make the lip region extremely smooth in order to detect correct lip movements and various trenches it forms on the face[9]. Now the skin color pixels of the lip region are converted to white pixels and remaining pixels to black pixels. As the lip region varies in evenness, tone and color, those pixels which resemble to an extent to skin color pixels are also converted to white color pixels. The upper lip color, chin color and few dents in cheek resemble the skin color pixels which are converted to white color pixels. To find these resembling skin color pixels, average RGB value of those pixels is calculated, and if the difference between two of those pixels is less than or equal to 10, then those pixels are denoted as skin alike pixels. The RGB values of the pixels will differ depending on the image quality and thus while pointing out those skin alike pixels, 10 is not always the best comparison factor. To find out the suitable comparison factor, a histogram is used to find out the minimum and the maximum RGB value. Depending on these values, it is found out that 8 can be a suitable comparison factor which will be suitable to find out the skin like pixels for both low and high quality images. Thus to detect only the lip in the lip region, we determine the longest binary pattern of the lip by finding the largest connected region bounded by black color pixels which is monotone in nature denoting the lip.

Similarly, we also need to detect only the eye in the eye region as the eye also plays an important role in expressing emotions. Similarly, first the skin color and skin color resembling pixels are converted to white pixels and remaining pixels to black color pixels. Then the eyebrows in the eye region need to me removed to detect only the eye and for doing it, continuous black pixels followed by continuous white pixels followed by continuous black pixels are searched and by eliminating the first series of continuous black color pixels the eyebrows can be removed. Now, again the largest connected region of black color pixels is found to denote the eye. In this way, the longest binary pattern of two important facial features i.e. the eye and the lip are determined. Figure 7 shows the longest binary pattern of the lip region and Figure 8 shows the longest binary pattern of the eye region after eliminating the eyebrows from it.

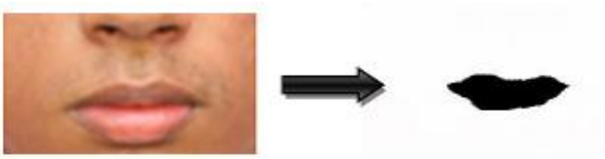


Fig. 7 Longest binary pattern of the lip area

Fig. 8 Longest binary pattern of the eye area eliminating the eyebrow

 DOI 10.17148/IJARCCE.2018.7126

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 1, January 2018

VIII. APPLYING BEZIER CURVE

In this phase, we apply the Bezier curve on the black color lip region and eye region obtained from the previous step. This is done by scanning the series of black color pixels horizontally until a white color pixel is found in order to find the starting and ending point of the lip. Thereafter tangents are drawn for both the lower and upper lip starting from both the end points of the lip. Then 2 points on the lower lip tangent and 2 points on the upper lip tangent are found which does not lie on the lip and thus gradually align both the tangents and the four points lying on it so as to generate a curve shaped outline of the lip. Similarly, the starting and ending pixel of both the eyes (Left Eye and Right Eye) are marked and four points lying on the tangents (2 points on lower eye tangent and 2 for upper eye tangent) for both the eyes are obtained in order to outline the eye with a Cubic Bezier curve [9]. Cubic Bezier Curves is used here and suppose the points are a1, a2, a3, a4, then explicit form of curve is given by equation (4),

$$B(t) = (1-t)^3 a1 + 3(1-t)^2 t a2 + 3(1-t)t^2 a3 + t^3 a4 \qquad (4)$$

If we follow these four points in order we can get the outline curve through which the values in the movement of eyes and lips can be mapped. Figure 9 and Figure 10 shows the Bezier curve applied on the lip and eye respectively.

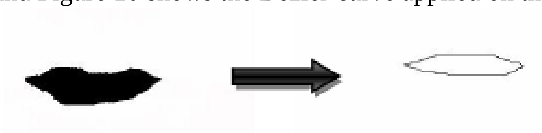


Fig. 9 Bezier curve of the lip

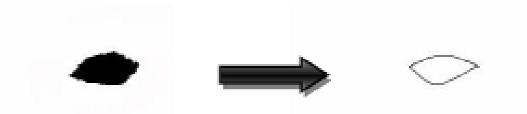


Fig. 10 Bezier curve of the eye

IX. EMOTION RECOGNITION AND RESULTS

The final phase of the system is to detect the emotion of the person through his facial expressions. For recognizing emotions we have to compare all the facial feature detected against values stored in an emotion dataset. The dataset contains six basic emotions such as Happiness, Sorrow, Anger, Fear, Surprise and Neutral and contains corresponding six control points, four on the tangents and two end points for Bezier Curve of lip and both eyes for each emotion and also contains the height and width of the lip and both the eyes and the distance of the left and right eyebrow raise for each emotion. Now to detect the emotion, we compare the features of the input image captured through the webcam with the values stored in the emotion dataset based on the closest matching pattern by averaging the distances of the feature points in the dataset. The closest matching pattern of the emotion is displayed as real-time output.

TABLE I
DISTANCE MEASUREMENT OF FEATURE POINTS

SL NO.	SUBJECT	LEW	LEH	REW	REH	REbR	LEbR	MOW	MOH
1	**Image 1**	12	10	13	10	25	27	45	16
2	**Image 2**	12	10	12	10	25	25	47	16
3	**Image 3**	13	11	14	11	25	27	48	17
4	**Image 4**	17	15	16	14	29	29	50	24
5	**Image 5**	18	15	18	15	27	28	55	20
6	**Image 6**	17	14	17	14	28	29	54	20
7	**Image 7**	13	10	14	11	25	26	48	18
8	**Image 8**	14	10	13	11	26	25	50	21
9	**Image 9**	16	14	17	14	25	27	52	20
10	**Image 10**	13	11	14	12	25	26	49	19

The features of the input face image are Left Eye Width (LEW), Left Eye Height (LEH), Right Eye Width (REW), Right Eye Height (REH), Mouth Opening Width (MOW), Mouth Opening Height (MOH), Left Eyebrow Raise (LEbR) and Right Eyebrow Raise (REbR). Therefore an emotion X, will be represented by eight features as: X={x1 x2 x3 x4

DOI 10.17148/IJARCCE.2018.7126

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 1, January 2018

x5 x6 x7 x8}. Thus every emotion is a vector of (1X8) dimensions. Table 1 display the distance measurements of facial feature points of various emotions and Figure 11 shows the various emotions recognized.

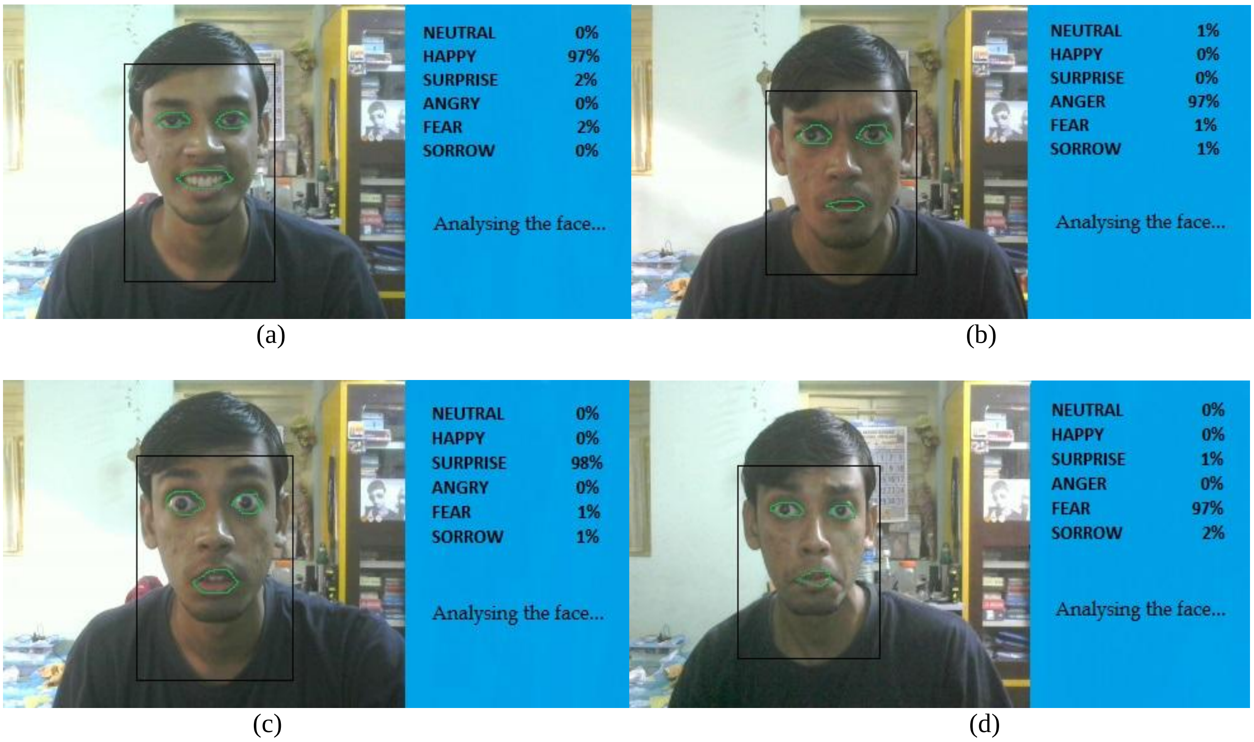


Fig. 11 Recognized emotion (a) Happiness (b) Anger (c) Surprise (d) Fear

X. CONCLUSION

In this paper, we have proposed an implementation methodology for recognizing human emotions based on facial expressions in real time. The algorithm implemented here first extracts the facial features required for recognizing expressions from the face region and then based on the distances measured between the various feature points of the face, the facial expression is detected and classified into six basic emotions such as happiness, surprise, anger, fear, sorrow and neutral independent of facial shapes and color. Cubic Bezier Curve is applied here for the system to work well in detecting emotions in an improvised manner. The model can detect emotions for a single human face and the efficiency reduces when multiple faces are introduced. Thus it can be further improved to support multiple face emotion recognition and crowd face analysis.

REFERENCES

[1] P. Ekman and W. V. Friesen (1971) Constants across cultures in the face and emotion. In: Journal of Personality and Social Psychology, vol 17(2), pp 124–129.
[2] B. Fasel and J. Luettin (2003) Automatic facial expression analysis: a survey. In: Pattern Recognition, vol 36(1), pp 259-275.
[3] P. Ekman (1994) Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. In: Psychology Bulletin, vol 115(2), pp 268-287.
[4] P. Ekman and W. V. Friesen (1975) Unmasking the face: a guide to recognizing emotions from facial clues. In: Englewood Cliffs, NJ: Prentice-Hall.
[5] Gwen Littlewort et al. (2011) The computer expression recognition toolbox (CERT). In: 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), ISBN: 978-1-4244-9140-7. doi:10.1109/FG.2011.5771414.
[6] P. Michel and R. E. Kaliouby (2003) Real time facial expression recognition in video using support vector machines. In: ICMI'03, Vancouver, British Columbia, Canada.
[7] Jorge Alberto Marcial Basilo et al. Explicit Image Detection using YCbCr Space Color Model as Skin Detection. In: Applications of Mathematics and Computer Engineering, ISBN: 978-960-474-270-7.
[8] Ashim Saha, Hillol Das, Nirmalya Kar and Dr. M. C. Pal (2013) An approach of extracting facial components for facial expression detection using fiducial point detection. In: International Journal of Computer Applications (IJCA), vol 80(10). doi:10.5120/13901-1923.
[9] P. M. Chavan, M. C. Jadhav, J. B. Mashruwala, A. K. Nehete and P. A. Panjari (2013) Real time emotion recognition through facial expressions for desktop devices. In: International Journal of Emerging Science and Engineering (IJESE), vol 1(7).