

A Neural Network Approach for Real Time Emotion Recognition

Atul Chavhan¹, Sneha Chavan², Samiksha Dahe³, Santosh Chibhade⁴

BE Scholar, Department of Computer Engineering, S.R.E.S.'s College of Engineering, Kopergaon, India^{1,2,3,4}

Abstract: Effective Human Computer Intelligent Interaction (HCII) requires the information about the user's identity, state and intent which can be extracted from images, so that computers can then react accordingly, e.g. systems behaving according to the emotional state of the person. The most expressive way humans display emotions is through facial expressions. In this paper we have presented the technique of recognizing emotion using facial expressions which is of prime focus in human interactions. By creating machines that can detect and understand emotion, we can enhance the human computer interaction. In this paper, we have discussed a framework for the classification of emotional states, based on images of the face and the implementation details of a real-time facial feature extraction and emotion recognition application. This application automatically detects faces from the captured images and codes them with respect to 7 dimensions in real time: Neutral, Anger, Disgust, Fear, Smile, Sadness, and Surprise. Most interestingly the outputs of the classifier change smoothly as a function of time, providing a possibly worth representation of code facial expression dynamics in a fully automatic and unnoticeable manner. The main objective of our work is the real-time implementation of a facial emotion recognition system.

Keywords: Real Time, Emotion Recognition, Human Computer Interaction (HCI), Feed-Forward Neural Network, Largest Connected Region (LCR), Cubic Bezier Curve.

I. INTRODUCTION

According to Dr. Charles Darwin, the facial expression indeed contribute in communicating one's emotions, opinions as well as intentions to each other in an effective way. In addition to this, his study on human behavior explicitly states that such expressions also provide information about the cognitive state of a person.

This includes the states like- boredom, stress, interest, confusion so on so forth. Hence, on the similar lines, considering the importance and the escalating need for advanced Human Computer Interaction, in this paper we have proposed techniques to recognize human emotions based on their facial expressions. Furthermore, it explains the necessity of Real-Time Systems in order to achieve high levels of interactions with the machines.

It is of prime importance that the interaction of humans with the computers should be free from latency, thus taking it to the level of face-to-face communication [1]. Hence, we are going to develop a robust real-time perceptive system, which will take into account the facial expressions, detect human face and code the expression dynamics.

Thus, systems working on these lines have a wide range of applications in basic and applied research areas, including man-machine communication, security, law enforcement, psychiatry, education and telecommunications [1].

In the field of emotion recognition, two general techniques are used to determine the emotional state of a subject.

One approach is to track the motion or shape of facial features, while another would be to consider the motion within the entire area of the head region. In this paper, we

are going to focus on the technique of facial feature tracking and the recognition of emotional states from its motion. Facial feature extraction deals with the problem of attempting to locate the position of important facial features, such as eyes, eyebrow, lip area etc. which are important in determining facial expression [6].

II. SYSTEM OVERVIEW

Fig. 1 shows the fundamental architecture of proposed system. There are 10 basic building blocks of the emotion recognition as enlisted below:

1. Live Streaming: Import input image from Webcam or Local Drive.
2. Skin Color Segmentation: It distinguishes between face and non-face parts.
3. Largest Connected Region: Check the probability to become a face.
4. Face Detection: Locates position of face in the given image frame using skin pixels.
5. Eye Detection: Identifies position of eyes in the image frame.
6. Lip Detection: Determines lip coordinates on face.
7. Bezier Curve Algorithm: Apply cubic Bezier curve equation on eye and lip to get coordinates.
8. Database Training: Stores values derived from Bezier curve equation that are used for comparison in emotion detection phase.
9. Emotion Detection: Emotion is detected by trained feed forward neural network as well as pattern matching using values from database.
10. Output Display: Shows output from emotion detection phase on screen.

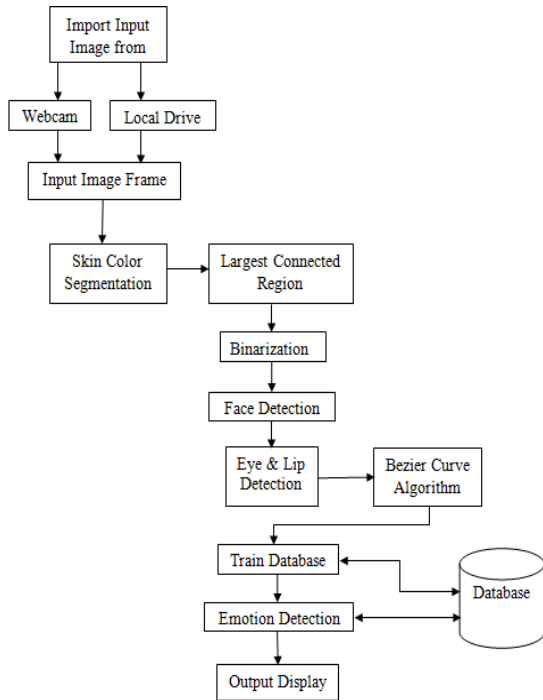


Figure 1: System Architecture

A. Import Input Image:

Image Data:

We are using the live image that is the image captured from the Webcam. When live streaming runs then we have to capture image from the webcam and this image will be processed for the next module.

Image data is conceptually a three-dimensional array of pixels, as shown in following Figure 2. Each of the three arrays in the example is called a band. The number of rows specifies the image height of a band, and the number of columns specifies the image width of a band. Monochrome images, such as a gray scale image, have only one band. Color images have three or more bands, although a band does not necessarily have to represent color. For example, satellite images of the earth may be acquired in several different spectral bands, such as red, green, blue, and infrared. In a color image, each band stores the red, green, and blue (RGB) components of an additive image, or the cyan, magenta, and yellow (CMY) components of a three-color subtractive image, or the cyan, magenta, yellow, and black (CMYK) components of a four-color subtractive image. Each pixel of an image is composed of a set of samples. For an RGB pixel, there are three samples, one each for red, green, and blue.

An image is sampled into a rectangular array of pixels. Each pixel has an (x, y) coordinate that corresponds to its location within the image. The x coordinate is the pixel's horizontal location; the y coordinate is the pixel's vertical location. Within JAI, the pixel at location $(0,0)$ is in the upper left corner of the image, with the x coordinates increasing in value to the right and y coordinates increasing in value downward. Sometimes the x coordinate is referred to as the pixel number and the y coordinate as the line number

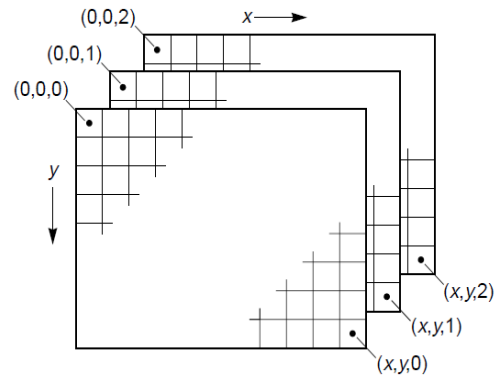


Figure 2: Image Read and Display

B. Skin Color Segmentation:

It discriminates between face and non-face parts. A general human skin is characterized by a combination of red and melanin (yellow, brown) and there is somewhat a range of hue for skin and saturation that represent skin-like pixels. YUV is the color space used in the PAL system of television broadcasting which is the standard in most of Europe and some other places [9].

Steps to perform Skin Color Segmentation:

- Contrast the image.
- Check the probability to become a face.
- Face will open in new form with LCR.
- If LCR height & width is larger or equal than 50.
- And ratio of height/width is between 1to 2, and then it is a face.

As we have detected Skin so now we have to select only that part of image which contains skin because that part will contain Face, so we will discard other portion by Finding Largest Connected Region [2].

C. Detecting Connected Component:

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

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

Where E is the Euler number, C is the number of connected components and H is the number of holes in a region. For our case, we already set the number of connected components (i.e. the skin region) to 1 since we

are considering 1 skin region at a time. The number of holes is then,

$$H = 1 - E \quad (2)$$

D. Face Detection:

In this module the face detection will perform, for which system will identify that the detected largest connected component having face or not?

- Locates position of face in the given image frame using skin pixels.
- Convert RGB image to binary image.
- Calculate average value of RGB for each pixel.
- Scan the binary image for forehead.
- Search for maximum width of continuous white pixels till we reach eye brows.
- Now cut the face in such a way that its height is 1.5 times that of its width.
- Connected component operators are non-linear filters that eliminate parts of the image, while reserving the contours of the remaining parts.
- This simplification property makes them attractive for segmentation and pattern recognition applications.
- Faces are nearly circular in shape and hence face components exhibit a high value for this operator.
- A threshold is fixed for this operator based on the observations on various face components.
- If a particular component shows a compactness value greater than this threshold it is retained for further analysis, else discarded. By this we will get the Skin Color which can have the Face [2].

Once the connected skin component representing the face is obtained, we then proceed to localize the face region by extracting the exact rectangular face region boundary by using the height and width of the connected component [2].

After getting the exact face region, we first convert it into a gray scale image, and its corresponding edge image is obtained by applying the Prewitt edge operator on the cropped gray scale image.

Then the facial features that correspond to a facial expression, namely the eye and mouth blocks are extracted from the face image using this edge information of the face [2].

E. Eye Detection:

The Eye detection performed after the detection of the Face, the left and right eye will be used for the processing the detection of the Emotion.

- Identifies position of eyes in the image frame.
- Convert RGB face to binary face.
- To find middle position of eye scan image from $w/4$ to $(w-w)/4$. Where, w -face width.
- For left eye black pixel-lines are placed in between $mid/2$ to $mid/4$.
- For right eye lines are between $mid+(w-mid)/4$ to $mid+3*(w-mid)/4$.
- Cut the upper, lower position & left, right side of two eyes from RGB image.

This is done using a Sobel edge operator. After finding the edge, we count the number of white pixels in vertical overlapping blocks traversing in horizontal direction. The block which contains the maximum white pixels represents the exact eye block [2].

F. Lip Detection:

Likewise Eye detection, Lip detection is performed after the detection of the Face, the lip will be used for the processing the detection of the Emotion.

- Determines lip coordinates on face.
- First determine the lip box.
- Lip must be inside the lip box.
- For this determine distance between forehead & eyes.
- Starting point of box will be 1/4 position of the left eye box.
- Ending point of box will be 3/4 position of the right eye box.
- Lip box will contain some part of nose.
- Cut this RGB image according to the image.

The same algorithm, which is used to extract eye block, is employed here. Based on the criterion that mouth is present only in the lower half of the face, we take into consideration only the lower half of the edge image of the original grayscale image containing the face region. On applying this algorithm, we get the block containing the mouth and from its edge image, we get the exact mouth block [2].

G. Cubic Bezier Curve:

As we have to detect emotion the shape of Eye and Lip will play the important role and we will get that shape by using the Bezier Curve.

H. Database Comparison:

This system is working on Neural Network, where we are performing the Database comparison where the Trained image set are compared with the obtain image set from the system and we will get compare that result and will get the result according the Neural Network.

In our database, there are two tables. One table "Person"(Emotions) is for storing the name of people and their index of 7 kinds of emotion which are stored in other table "Position".

In the "Position" table, for each index, there are 6 control points for lip Bezier curve, 6 control points for left eye Bezier curve, 6 control points for right eye Bezier curve, lip height and width, left eye height and width and right eye height and width. So, by this method, the program learns the emotion of the people.



I. Neural Network Architecture, Simulation and Training:

We have chosen multilayer feed forward network as the network architecture. The number of neurons that has to be in the output layer is fixed as we know the number of emotions that we are going to consider - Neutral, Anger, Disgust, Fear, Smile, Sadness, and Surprise. Therefore, in our case, the number of neurons in the output layer is chosen as 7. The number of hidden layers in the network and the number of neurons in each layer is chosen by trial and error method based on the performance function until it reaches the specified goal. By trying various combinations, we have chosen the efficient architecture, which is a three layer feed forward network where there is one input layer, one hidden layer and an output layer.

For training the network, to recognize various emotions, first of all we downloaded the different face images for different emotions from internet and created a new database which includes the images from our own database of face images.

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

J. Emotion Detection:

Emotion is detected by pattern matching using values from database. For this first,

- Convert each width of Bezier curve to 100.
- Convert height according to its width.
- Program will match emotions height in database & displays output.

If emotion is not available in database then program will calculate average height for each emotion in the database for all people and then get a decision according to the average height.

Compare that found features of the face with the training set of our trained features and on basis of that it will show the result.

Though various methods exist for emotion recognition, neural networks hold its position due to its robustness. So, we are going to apply neural network based approach for recognizing emotions [2].

III. EXPERIMENTAL RESULTS

After applying the operations on real time images and stable images (taken from local drive), this system successfully works on all the modules and detects the all type of emotions that are mentioned in the paper.

Figure 3: Form 1 for Emotion Detection System

In the above form face image is captured through webcam and then processed for next module.

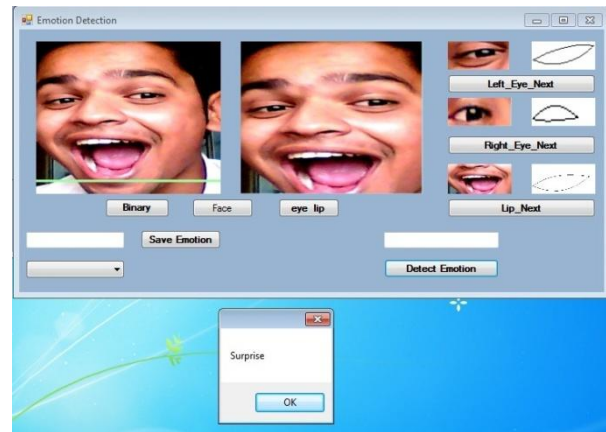


Figure 4: Form 2 for Emotion Detection System

In the above form actual face is detected and then with the help of (eye_lip button) cubic bezier curve we have to find out the exact eye and lip coordinates. After that emotion will be detected and displayed on the screen.

IV. CONCLUSION

Our system verify a wide variety of face images, with many emotions and many different angles other than frontal face image and to provide the safe keeping of person entry or out from the room or lab, whether the person is lawful or not. The system is highly accurate. The system is skilled for using multi-cameras as the capturing device simultaneously and providing the detail of person of recognize whether it is valid or invalid.

In this paper, we present a system that automatically detects human emotions on the basis of facial expressions. Moreover, we propose to improve the emotion detection technique using Cubic Bezier Curve Implementation and Feed-Forward Neural Network which is more adaptive and resurface as the ones with utmost importance in various other fields like that of robotics, computer graphics, automation and animation. This system works well for faces with different shapes, complexions as well as skin tones and senses basic seven emotional expressions.

ACKNOWLEDGMENT

We gratefully acknowledge H.O.D **Prof. D. B. Kshirsagar** of computer engineering department of our college for their kind support for this project. We also thank our project guide **Prof. J. N. Kale** and co-guide **Prof. V. N. Nirgude** for highlighting our path and their gracious guidance. In last we like to thank all the friends who had given some valuable contribution for this system.

REFERENCES

- [1] International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319-6378, Volume-1, Issue-7, May 2013 104 "Real Time Emotion Recognition through Facial Expressions for Desktop Devices".
- [2] Sreevatsan. A.N, Sathish Kumar. K.G, Rakeshsharma. S and Mohd. MansoorRoomi, "Emotion Recognition from Facial Expressions: A Target Oriented Approach Using Neural Network", Thiagarajar College of Engineering, Madurai - 625015, INDIA.
- [3] MarianStewart Bartlett, Gwen Littlewort, Ian Fasel ,Javier R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction, Machine perception laboratory Institute for Neural Computation", University of California, San Diego, CA.

- [4] Gurpreet Singh, RajwinderKaur, DeepinderKaur, “Face Recognition Method using Skin Color segmentation and Connected Region Approach”.
- [5] SaumilSrivastava, “REAL TIME FACIAL EXPRESSION RECOGNITION USING A NOVEL METHOD”, Department of Computer Science, Jaypee Institute of Information Technology, Noida.
- [6] Liyanage C De Silva, Suen Chun Hui, “Real Time Facial Feature Extraction and Emotion Recognition” 2003.
- [7] AbdesselamBouzerdoun, Son Lam Phung, FokHing Chi Tivive, PeiyaoLiFeature, “Selection for Facial Expression Recognition”.
- [8] Irfan A. Essa, “Coding, Analysis, Interpretation, and Recognition of Facial Expressions”.
- [9] M. R. Tabassum, A. U. Gias, M. M. Kamal, H. M. Muctadir, M. Ibrahim, A. K. Shakir, A.Imran, S. Islam, M. G. Rabbani1, S. M. Khaled, M. S. Islam, Z. Begum, “Comparative Study of Statistical Skin Detection Algorithms for Sub-Continental Human Images”, Institute of Information Technology, University of Dhaka, Dhaka-1000, Bangladesh.
- [10] DewiAgushinta R., AdangSuhendra, SarifuddinMadenda, Suryadi H.S, “Face Component Extraction Using Segmentation Method on Face Recognition System”,Information System Department, Informatic Department, Gunadarma University, Indonesia.