

Emotion Detection Algorithm

Namrata Mahajan, Harshad Mahajan

M.tech (Communication), B.Tech (EXTC)
Nagpur University-India, Amravati University India

Abstract: Human can see emotion as associated with mood, temperament, personality and disposition. To detect emotions is easy for humans but it's quite difficult for computer as world is three dimensional but Computer has only two dimensions. Computer seeks to emulate the human emotions by digital image analysis. Humans can use vision to identify objects quickly and accurately. Human can detect emotion using voice based on different parameters like tone, pitch, pace and volume; but in case of digital images detecting emotion just by analysing images is a novel way.

This algorithm has two major parts. First, Template database generation and another is emotion detection. In this, we first extract face from an image using some basic image processing operations and color models in it. Here we define thresholds to separate our region of interest. Then we perform lip detection on cropped face and this extracted lips are stored in our database with emotion name in database generation phase while in emotion detection phase this extracted lips are compared with series of stored template in database and on the basis of best correlated template emotion is recognized. This method of detecting emotions is simple and fast as compared to previous methods i.e. using brain activity or by speech. Size of database will affect the effectiveness of the proposed algorithm.

Keywords: Emotion Detection, Face Detection, Lip Detection, Skin Segmentation.

I. INTRODUCTION

Human has ability to see emotions by temperament, disposition, personality and mood. Computer seeks to emulate the human emotions by digital image analysis. The problem with the computer vision is due to the fact that world is three dimensional but Computer has only two dimensions.

In our day to day life we interact with each other directly (for e.g. face to face) or indirectly (for e.g. phone calls). In some profession interaction like call centers interaction with people is important. With great advancement in technology in terms of different techniques of people interacting with each other it is quite necessary that one should be aware of current emotions of the person he/she is interacting.

It is widely accepted from psychological theory that human emotions can be classified into six archetypal emotions: surprise, fear, disgust, anger, happiness, and sadness. Facial motion and the tone of the speech play a major role in expressing these emotions.

The muscles of the face can be changed and the tone and the energy in the production of the speech can be intentionally modified to communicate different feelings. Human beings can recognize these signals even if they are subtly displayed, by simultaneously processing information acquired by ears and eyes. Based on psychological studies, which show that visual information modifies the perception of speech [1], it is possible to assume that human emotion perception follows a similar trend. This paper analyses the use of digital image to recognize four different human emotions: sadness, happiness, anger and neutral state, using a database recorded. The more challenging task of capturing salient visual information directly from conventional videos is a topic for future work but is hoped to be informed by studies such as in this report. The primary purpose of this paper is to define a simple and fast automated emotion recognition system.

II. PROCESS AUTOMATION

In order to implement an automated emotion detection system, a series of logical steps have to be developed. The flow of steps in this automated system is defined below. Our project is mainly divided in two parts Database Generation (Fig.1) and Emotion Recognition (Fig.2). In Emotion recognition we are going to match resulting output of input image with our templates which already exist in our database.

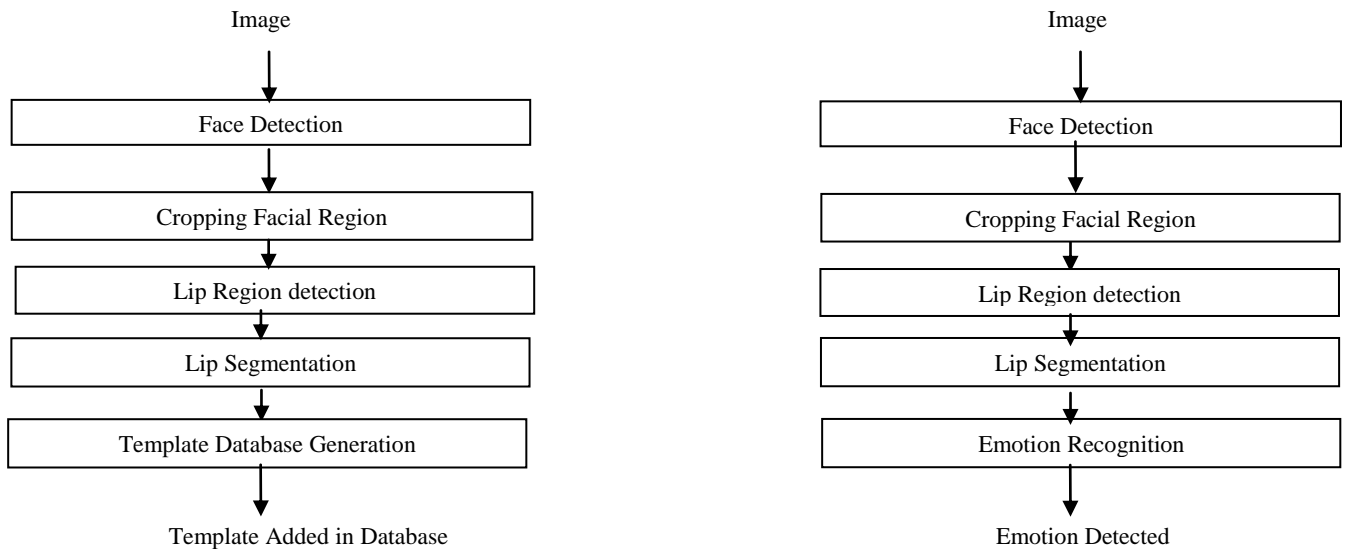


Fig.1 Database Generation Flowchart Fig.2 Emotion Recognition Flowchart

A. Face Detection

We faced many problems related to the face detection. To locate the face image should consists of single face. Hence in this paper our assumption is that an input image is having single face. If this assumption is not satisfied it will take either of any one image from input not any specific one. Although we provide a single face image for face detection but face detection means identifying a region of image which contains face no matter what is the background of image, its lighting conditions or orientation. This task is bit critical as different faces have different colours, shapes and texture.

At this time, many approaches involve automatic face detection followed by lip detection .But most of approaches define the scope of upright view of faces such approaches failed to detect faces rotated in plane or out of image plane. Also sometimes it failed while detecting faces from image when only part of face is there in image.

Hence, in this project we are using thresholding to separate skin region from image in order to detect face.

1. Skin Color Classification

Skin color and textures are important cues that people use consciously or unconsciously to infer variety of culture-related aspects about each other. Skin color and texture can be an indication of race, health, age, wealth, beauty, etc. [3]. However, such interpretations vary across cultures and across the history. In images, skin color is an indication of the existence of humans in such media.

The color of human skin is created by a combination of blood (red) and melanin (yellow, brown). Skin colours lie between these two extreme hues and are somewhat saturate Facial skin takes on is clearly a subspace of the total color space. The human skin is a fraction of the actual color cube, about 0.25 % of the total colours. Except for extremely hairy subjects, which are rare, skin has only low-amplitude texture. Three color spaces i.e. RGB, HIS and YCrCb are defined in literature and they are useful to achieve our goal.

2. Skin Based Segmentation

To detect Human faces, their skin color is unique and efficient feature. Although skin color may vary person to person, major difference lies largely between their intensity rather than their chrominance. Most of the face detection algorithms

are on the basis of skin based segmentation. Many of these algorithms use a range of color values to define skin color.

3. Segmentation Rules

Segmentation procedures partition an image into its constituent parts or objects. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually. In general, the more accurate the segmentation, the more likely recognition is to succeed. Segmentation rules determine the formation of regions. These rules can be defined on the basis of color and edge properties of region. Using skin based segmentation we can reduce the efforts of first determining faces within image in which faces are out of frame or tilted.

4. Segmentation Rules for different Colour Models

After some experiments we found following rules for different colour models which give one step closure to our goal.

RGB Model: $0.836G - 14 < B < 0.836G + 44 \Rightarrow \text{Skin}$ OR $0.79G - 67 < B < 0.78G + 42 \Rightarrow \text{Skin}$

HIS/HSV Model: $19 < H < 240 \Rightarrow \text{Not skin}$

YCrCb Model: $140 < Cr < 165 \Rightarrow \text{Skin}$ OR $140 < Cr < 195 \Rightarrow \text{Skin}$



Fig 3.Original Image



Fig.4 Image after Skin Thresholding

5. Erosion and Dilation

After skin based segmentation, for fast processing we convert our output image into binary image. The noises in this black and white image are removed using suitable window sizes. The pixels in this image which were falsely detected as skin color pixels are removed by erosion and dilation by moving window over image matrix. If it is found that the number of white pixels is more than fixed threshold then whole window is made black. The face detection is performed on skin detected areas. Hence to locate face we always consider the occurrence of white region in output image after erosion and dilation. This process has assumption that image on which we are going to perform process should have single face with minimum skin exposure.

Skin colors in this project are using RGB, HIS/HSV and YCrCb models. Values are unique for this project obtained using fresh toned digital photos. For larger skin tone range HIS model gives better results but these models made it difficult to constraint color for flesh-like tones. RGB models are preferred in such situations.



Fig.5 Image after Dilation



Fig.6 Image after Erosion

B. Cropping Facial Region

After detection of Face we are going to extract lip region .We are aware about the thing that lip region contents maximum of red color content. If red color content is present in the image outside the facial region,it will interfere in our lip extraction process it may also cause increase in time and complexity.Thus we decide to crop the region of our interest only.So focusing on this limited area, results will be more accurate. Also Computational time is directly proportional to the size of image. If processing image size is reduced due to cropping region of interest,computational time will also reduce.

Before cropping we used to apply erosion and dilation operations on detected face image. These functions help us to remove noise which is quite helpful for region cropping.Major work in cropping is to find four boundary points.We found first top, bottom, right and left four points. On the basis of these four points height and width of detected region can be calculated. After calculating height and width we create rectangle and then crop required area of facial region.

C. Lip Detection

After face detection, we have our region of interest; on this image we applied our most important algorithm of lip detection. Human emotions reflect changes in lip region and eyes. But maximum changes around lips are visible rather than eyes. Hence a detected lip gives us accurate results.

After face detection and getting our region of interest, we now move towards the most important part of the project i.e. lip detection. As already discussed in the 1st chapter, the variations seen during different emotions are around the lip region as well as the eyes region. But the maximum changes that are visible will be around the lip region, thus giving us more accurate results.

Same as we used RGB thresholds in skin based segmentation to detect face. Here we used thresholds on red color as lip region contains maximum of red colors.

In field of computer vision, automation of the process of locating and tracking person's lip is somewhat difficult task. There several methods proposed for lip detection in last 10-12 years. But we found important and effective way of detecting lip based on the segmentation directly from the color space. In this algorithm we can enlarge difference between skin and lip using color filters or color transformations.

The processing time in this algorithm is main advantage .Lip segmentation is very sensitive to changes in colors. Hence it may create problem if for some faces lip and skin contrast is lower.



Fig. 7 Image after Lip Detection

D. Lip Segmentation

After completion of lip detection, now our next step is lip segmentation. After this segmentation we will use result of lip segmentation to generate template if we are creating database otherwise it will be used to detect emotion if we are in emotion detection phase of project. Thus this cropped lip region makes us closer to our aim.

Again the process of lip segmentation is same as face detection .In which we need to define first four boundary points.i.e. top, bottom, left and right points of detected lip. From this point we will find width and height of lip region which is further used to create rectangle and extract lip region.

It may possible in case of noise occurs on the image, it will create confusion between lips and any other area on face. To avoid this confusion we used a fact that width of height is 3 times than lips height. But for some worst case we modified this condition as lip width should be greater than the twice of its width, if this condition is true then we create rectangle window and crop lip region.

E. Template Database Generation

Our project is mainly divided in two parts. 1. Database Generation and 2. Emotion Recognition. In Emotion recognition we are going to match resulting output of input image with our templates which already exist in our database. But for this we need to create database. In order to generate database, templates have to be generated. Roipoly function is used to create template. It is out of box function in MATLAB.Roipoly function specifies the polygonal region of interest (ROI)within an image. Its output is in the form of binary image which is used as mask for filtering.

BW = roipoly creates an interactive polygon tool, associated with the image displayed in the current figure, called the target image. With the polygon tool active, the pointer changes to cross hairs, '+' when you move the pointer over the image in the figure. Using the mouse, you specify the region by selecting vertices of the polygon or any area of interest. You can move or resize the polygon or figure using the mouse. When you are finished positioning and sizing the polygon/figure, create the mask by double-clicking, or by right-clicking inside the region and selecting Create mask from the context menu. Roipoly returns the mask as a binary image, BW, the same size as I. In the mask image, roipoly sets pixels inside the region to 1 and pixels outside the region to 0. This binary image BW so returned is in logical format and as such cannot be used for implementation. It needs to be converted to a format that can be processed on i.e. double or uint8. After this the masks are cropped in the similar manner used for face cropping and lip cropping i.e. finding four points and creating rectangle and finally cropping.[5]

F. Emotion Recognition

In Emotion recognition we are going to match resulting output of input image with our templates which already exist in our database. This comparison is made by using cross-correlation.

This option calculates the cross correlation function for two images. Each of the images is divided into rectangular blocks. Each block in the first image is correlated with its corresponding block in the second image to produce the cross correlation as a function of position.

The cross correlation may be used to determine the degree of similarity between two similar images, or, with the addition of a linear offset to one of the images, the spatial shift or spatial correlation between the images.

G. Emotion Detection

After finding cross-correlation of the detected and cropped lip region with templates, we have to find the index number on which we get maximum value offered by cross-correlation of two images. Emotion is detected by checking the corresponding emotion across the index number with maximum value of cross-correlation

III. Final Remarks

We found that this algorithm is faster and easy to implement than existing algorithm based on speech recognition and by ECG in which physical contact with human brain is needed which is quite expensive than this. We worked on five emotions i.e. happiness, grief, anger, surprise and neutral. These emotions give us successful results on majority of images. Generally no false emotion were found or wrongly interpreted.

References

- [1] "Human Emotion Recognition using Frequency & Statistical Measures of EEG Signal", Monira Islam, Tazrin Ahmed, Sheikh Shanawaz Mostafa*, Md. Salah Uddin Yusuf, and Mohiuddin Ahmad
- [2] "Human Emotion Recognition using Frequency & Statistical Measures of EEG Signal", Monira Islam, Tazrin Ahmed, Sheikh Shanawaz Mostafa*, Md. Salah Uddin Yusuf, and Mohiuddin Ahmad
- [3] Bernhard Fink, K.G., Matts, P.J.: Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior* 27(6) (2006) 433–442
- [4] Jones, M.J., Rehg, and J.M.: Statistical color models with application to skin detection. *International Journal of Computer Vision (IJCV)* 46(1)(2002) 81–96
- [5] <http://www.mathworks.in/help/images/ref/roipoly.html>
- [6] A. Yuille, "Feature Extraction from Faces Using Deformable Templates," *Int. Journal of Computer Vision*, (1992), 8(2):99-111.
- [7] Bracewell R.N.: "The Hartley transform", Oxford University Press, 1986.
- [8] Gocke R., Millar J.B., Zelinsky A., Robert-Ribes J, "Automatic extraction of lip featurepoints". *Proc. Austral. Conf. Robotics and Automation*, Melbourne, Australia, August 2000, pp. 31–36.