International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-6, Issue-2, October 2016

Emotion Detection System using Facial Action Coding System

Vedant Chauhan, Yash Agrawal, Vinay Bhutada

***Abstract*— Behaviors, poses, actions, speech and, facial expressions; these are considered as channels that convey human emotions. Emotions are an extremely important part of human life and so immense research has been carried out to explore the relationships between these channels and emotions, which have led to important real world applications. Facial expressions are the most varied way for micro-expressions. They are closely accurate indicators of emotion. This paper proposes a system which recognizes the emotion represented on a face. Thus a Facial Action Coding System (FACS) used in classifying the universal emotions: Happiness, Anger, Sadness, Fear, Surprise and Disgust. Individual differences in every component of a face like eyes, face, cheeks etc. combines to detect a particular emotion. Colored frontal facial images are given as input to the FACS system. After the image is captured via webcam, facial feature are marked for neutral and emotion face image. This is an image processing step, where the input image is processed so that its pixels are readable by machine. Now this processed image is overlaid on a base image which is used by FACS to differentiate instant changes in facial expressions. Finally, a set of values obtained after processing those marked feature points are compared to recognize the emotion contained. Based on the emotion certain audio is played depicting that emotion. This system can be useful for psychologists, animators, game developers, criminal studies and many more.**

***Index Terms*—Action Units, FACS, HCI, IP.**

I. INTRODUCTION

An emotion is a mental and cognitive state which is private and subjective; it involves a lot of actions, behaviors, feelings, and thoughts. There are six basic emotions which this project will be focusing on happiness, sadness, anger, fear, surprise, and disgust.

Many factors contribute in conveying emotions of an individual. Speech, pose, behavior, actions, and facial expressions are some of them. From these above mentioned factors facial expressions have a higher importance since they are easily perceptible.

Computer Vision experts are now being attracted towards Facial Expression Analysis. A number of facial features like eyes, lips etc. are being tracked by multistate face and facial component models.

The idea of this project stems from the fact that a person's emotion is being recognized on his or her state of mind, or rather, "emotion". The significance of facial expressions in determining the mood of a person combined with current technologies allows us to provide an automated solution for the above mentioned task.

Vedant Chauhan, Computer Engineering, Payoda Technologies, Chennai, India, 9047078556

Yash Agrawal, Computer Science, IUPUI, Indianapolis, United States of America, 2016391619

Vinay Bhutada, Computer Science Engineering, University of Texas, Arlington, United States of America, 4699541144

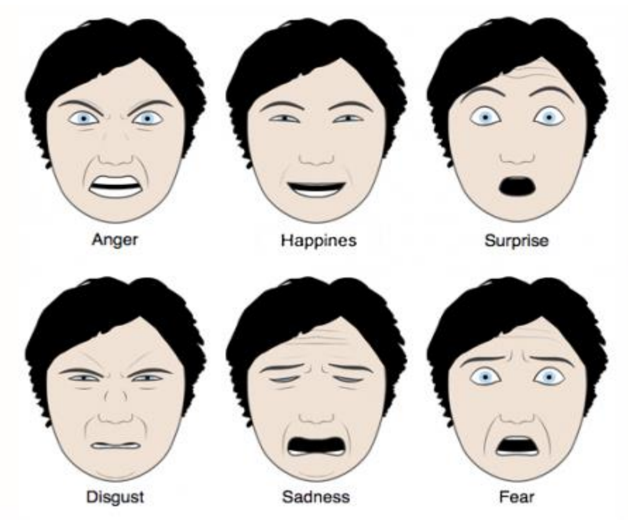


Fig. 1 Six basic emotions

II. PROPOSED SYSTEM

A. Image Processing

Image processing is one of the form of signal processing. The input is an image (photograph or video frame); the output is either an image or a set of parameters related to the image. Standard signal processing techniques are applied to images since an image can be treated as a 2-D signal.

An image is considered to be a function of two real variables, for example, p(x, y) with p as the amplitude (e.g. brightness) of the image at the real coordinate position (x, y).

In this project, image processing helps in extracting facial features from an image with emotions. This emotional image is overlapped on the base image which notifies the image processing tool to find the differences between the two images. The extracted data is passed to FACS system.

B. Facial Action Coding System (FACS)

FACS [3] coding is the state of the art system for manual measurement of facial action. It is, however, is labor concerted and difficult to systematize across coders. Goal of automated FACS [3] coding is to remove the need for manual coding and apprehend automatic recognition and analysis of facial actions. Success of this effort depends on retrieving reliably coded collection of FACS-coded images from well-chosen observational scenarios. Completing the necessary FACS [3] coding for testing and training algorithms has been a rate-limiter. Manual FACS [3] coding remains expensive and slow.

Fast-FACS [1] uses advances in Computer Vision and machine learning to increase efficiency and reliability of FACS coding. It includes 3 parts:

Active Appearance Tracking: There exist a variety of methods for facial feature tracking. Appearance models have become increasingly important in computer vision and graphics over the past few years. Parameterized Appearance Models (PAMs) have been proven useful for alignment, detection, tracking, and face synthesis. In particular, Active Appearance Models (AAMs) have proven an excellent tool for detecting and aligning facial features. AAMs typically fit their shape and appearance components to an image through a gradient descent, although other optimization approaches have been employed with similar results.

Peak, Onset and Offset Coding: The user annotates the peak of a facial action. The system then automatically determines the remaining boundaries of the event, that is, the onset and offset (extent) of the AU (Action Unit). The estimation of the position of the onset and offset of a given event peak is based on a similarity measure defined on features derived from the AAM mesh of the tracked face and on the expected distribution of onset and offset durations (for a given AU) derived from a database of manually coded AUs.

Learning a metric for onset/offset detection: It describes the procedure to learn a metric for onset and offset estimation.

Fig. 2 shows the main idea of project. The specific aims are to:

First, reduce time and effort required for manual FACS [3] coding by using novel computer vision and machine learning techniques. Second, increase reliability of FACS [3] coding by increasing the internal consistency of manual FACS [3] coding. Third, develop an intuitive graphical user interface that is comparable to commercially available packages in ease of use, while enabling fast reliable coding.

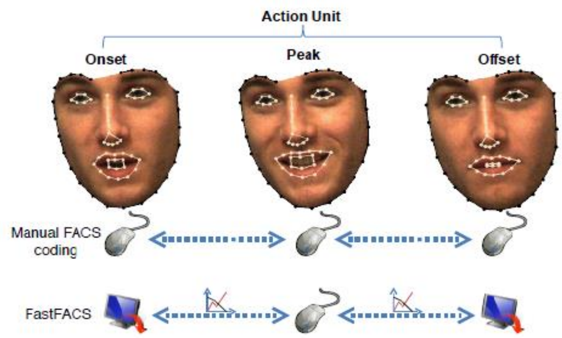


Fig. 2 Emotions at time frames

FACS [3] coding typically involves frame-by-frame inspection of the video, paying close attention to subtle cues such as wrinkles, bulges, and furrows. Left to right, evolution of an AU 12 (involved in smiling), from onset, peak, to offset. Using Fast-FACS [1] only the peak needs to be labelled and the onset/offset are estimated automatically.

C. Human Computer Interaction (HCI)

HCI involves the study, planning, and design of the interaction between people (users) and computers. It is often regarded as the intersection of computer science, behavioral sciences, design and several other fields of study. HCI aims to improve the interactions between users and computers by making computers more usable and receptive to users' needs.

The basic idea of the project was to take human and machine interaction to a whole new level. The primary purpose was to integrate the system developed with a virtual assistant like Siri or Cortana. Once that is implemented, the virtual presence will no longer be a just an application but will become more human than ever. It will be more of a companion or friend to the user as and when needed. This takes interaction between humans and computers to a whole new level.

This project further will perform events like showing funny video, images, playing music etc. based on user emotions.

III. DESIGN CONSIDERATIONS

The project is using FACS [3]. The image is captured by webcam for neutral face and emotion face. Then using Fast-FACS [1] technique for feature points is created and features are marked for a set of values. The values of both neutral face and emotion are compared which classify the emotion and based on the emotion certain audio is played depicting that emotion.

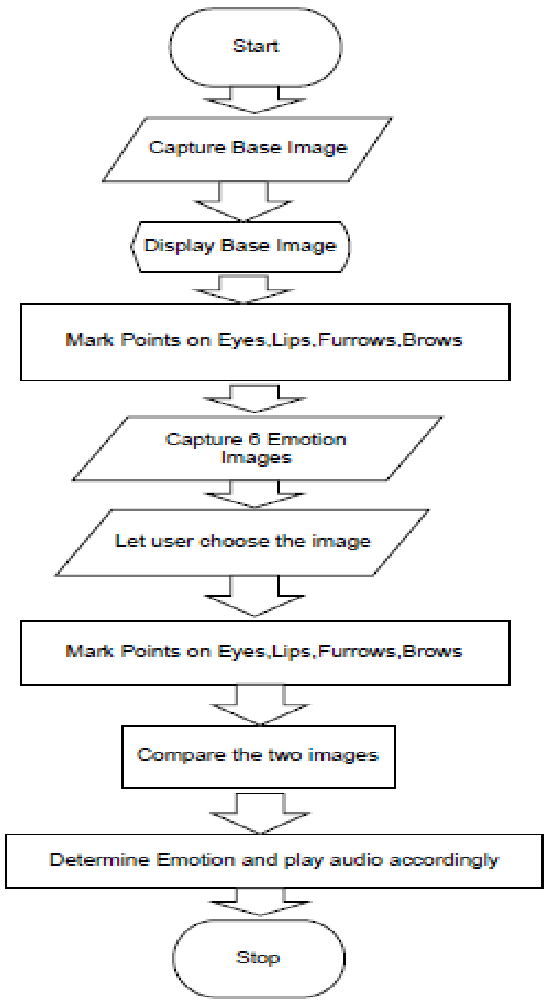


Fig. 3 Design flowchart

A simple approach is followed while designing and consists of the steps in order:

A. Image Acquisition

Image is captured via webcam and with the help of image acquisition tools image is enhanced and stored with proper resolution.

B. FACS Feature Extraction

The technique used for extracting facial features is FACS [3]. It describes facial activity on the basis of 44 unique AUs, as well as several categories of head and eye positions and movements. It is slow and less efficient in traditional manner. So, Fast-FACS [1] is used.

Fast-FACS [1] uses advances in Computer Vision and machine learning to increase efficiency and reliability of FACS [3] coding. Now we are considering two images, base image show in Fig. 4 and emotion image shown in Fig. 5.

Fig. 4 Base Image

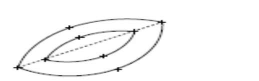
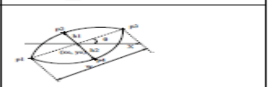
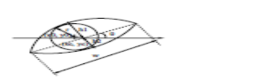
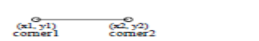
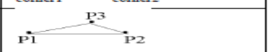
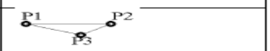
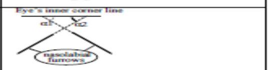
Fig. 5 Emotion Image

On both images features like eyes, lips, furrows and brows are marked and processing is done using Fast-FACS [1] technique.

For detecting emotions we have created an algorithm which determines the emotion of a person and plays an audio or displays text depicting the emotion.

The Table I shows a Multi-state facial component model of a frontal face. Contraction of the facial muscles produces changes in both the direction and magnitude of the motion on the skin surface and in the appearance of permanent and transient facial features. Examples of permanent features are the lips, eyes, and any furrows that have become permanent with age. Transient features include any facial lines and furrows that are not present at rest.

Table I Multi-state facial component model of a frontal face

Component	State	Description/Feature
Lip	Opened	
	Closed	
	Tightly closed	
Eye	Open	
	Closed	
Brow	Present	
Cheek	Present	
Furrow	Present	
	Absent	

As FACS [3] provides different facial features for upper and lower face. Table I shows basic Upper and Lower face action units or there combinations. Now, each emotion is specified by different FACS [3] Action Unit shown in Table III.

Table II Basic Upper Face Action Units or Combinations

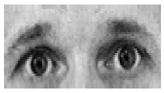
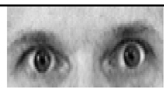
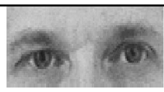
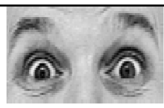
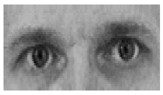
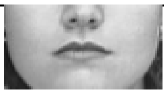
AU 1	AU 2	AU 4
		
Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together
AU 5	AU 6	AU 7
		
Upper eyelids are raised.	Cheeks are raised.	Lower eyelids are raised.
AU 1+4	AU 4+5	AU 1+2
		
Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.	Inner and outer portions of the brows are raised.
AU 1+2+4	AU1+2+5+6+7	AU0(neutral)
		
Brows are pulled together and upward.	Brow, eyelids, and cheek are raised.	Eyes, brow, and cheek are relaxed.

Table III Emotion represented by FACS Action Unit

Emotion	Action Units
Happiness	6+12
Sadness	1+4+15
Surprise	1+2+5B+26
Fear	1+2+4+5+7+20+26
Anger	4+5+7+23
Disgust	9+15+16

Table IV Basic Lower Face Action Units or Combinations

AU 9	AU 10	AU20
		
The infraorbital triangle and center of the upper lip are pulled upwards. Nose wrinkling is present.	The infraorbital triangle is pushed upwards. Upper lip is raised. Nose wrinkle is absent.	The lips and the lower portion of the nasolabial furrow are pulled pulled back laterally. The mouth is elongated.
AU 15	AU 17	AU12
		
The corner of the lips are pulled down.	The chin boss is pushed upwards.	Lip corners are pulled obliquely.
AU 25	AU 26	AU27
		
Lips are relaxed and parted.	Lips are relaxed and parted; mandible is lowered.	Mouth stretched, open and the mandible pulled downwards.
AU 23+24	neutral	
		
Lips tightened, narrowed, and pressed together.	Lips relaxed and closed.	

IV. IMPLEMENTATION

A. Steps in Image Acquisition

Input: Image from webcam.

1) Click on the option to capture the image for both base and emotion image.

2) A preview window is displayed, and the user will be able to fit his face inside the window so that an optimal frontal image of his/her face can be captured.

3) We capture a snapshot using img = getsnapshot(vid). The image "img" will then be stored in a dedicated folder.

Output: Image is captured in desired resolution.

B. Steps in Fast-FACS

Input: Image obtained from webcam.

1) Subtle changes in the facial components are measured; we develop a multistate model based system for tracking facial features.

2) Motivated by FACS [3] action units, these changes are represented as a collection of midlevel feature parameters, facial features such as lips, eyes, furrows and brows are marked on both base and emotion image respectively,

~ 4 points on lips, eyes and furrows

~ 2 points on brows.

3) Calculations are done on both base and emotion images by using Euclidean distance function "pdist(p,'euclidean')".

Output: Values calculated for detection of emotion.

C. Algorithm for Emotion Recognition

Input: Values obtained from Fast-FACS [1] step.

1) Store facial feature points of base image.

2) Store facial feature points for emotion image.

3) for (emotion)

a. compare vertical distance between lips.

b. compare vertical distance between eyes.

c. compare horizontal distance between lips.

d. compare horizontal distance between eye brows.

4) Perform Step 3 for emotion = happy, sad, angry, disgust, shock and fear.

5) Get emotion of image.

Output: Emotion is detected and audio file is played based on emotion.

V. RESULTS

A. Result of Image Acquisition

This subsection displays the output after capturing an image via webcam.

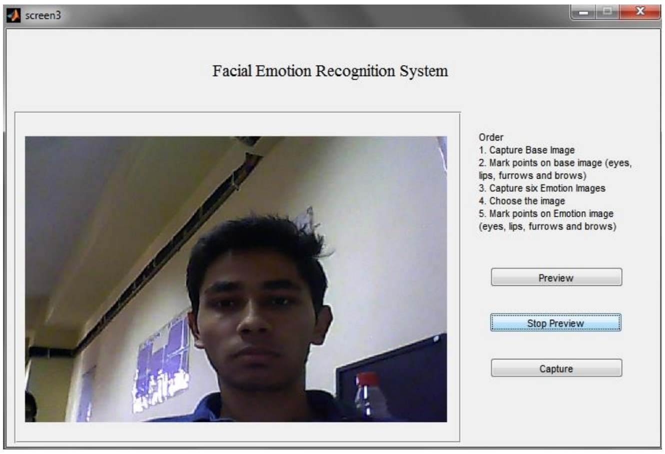


Fig. 6 Result of Image Acquisition

B. Result of Fast-FACS

This section displays various output screens for base image and emotion image.

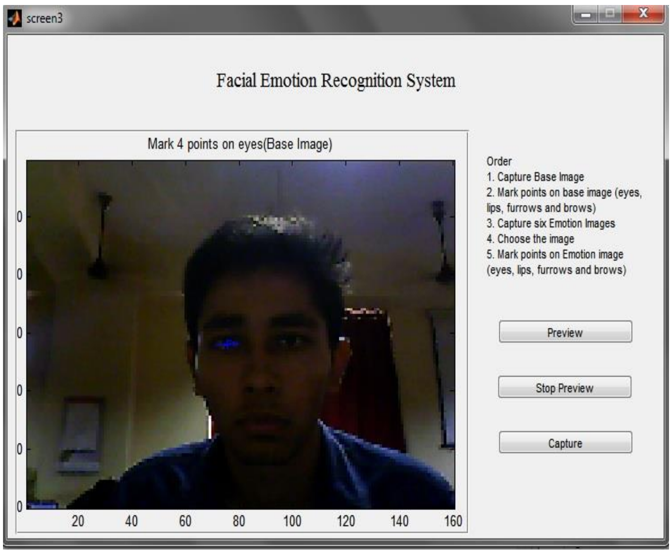


Fig. 7 Marking of points on Emotion image (lips)

Similarly, points are marked for eyes, furrows and brows on base image.

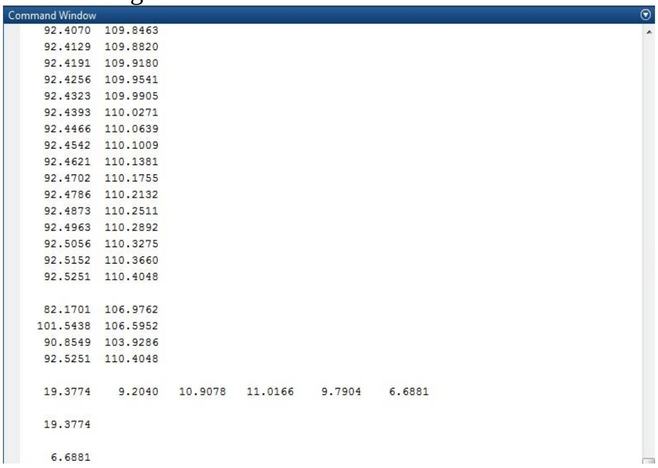


Fig. 8 Values obtained in Fast-FACS step

C. *Result of Emotion Recognition*

This subsection displays final emotion and audio file is played.

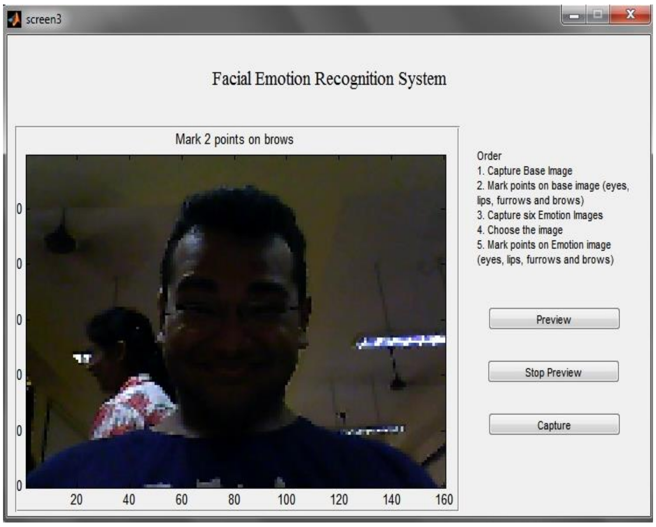


Fig. 9 Emotion image

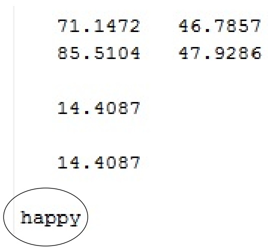


Fig. 10 Command window depicting the emotion

VI. CONCLUSION

A Facial Emotion Recognition System was designed using Fast-FACS [1] technique. Different feature points of the face were used to recognize the various emotions. FACS [3] was used in classifying the universal emotions: Happiness, Sadness, Anger, Disgust, Surprise and Fear. Colored frontal facial images are clicked and given as input to the system.

The face was detected efficiently every time the image is captured was kept free from other faces or intense lighting, and the person was completely facing the camera. There were 6 input images were captured at a time. Out of these 6 images, one image displaying emotion is chosen and it was compared with the neutral image. After the image is captured via webcam, facial feature are marked for neutral and emotion face image. Accessories such as spectacles or sunglasses were proved to be interference to the accurate operation of the system as the user might not mark the points accurately.
Finally, a set of values obtained after processing these marked feature points were evaluated to recognize the emotion displayed. All emotions were classified correctly on basis of feature points values obtained. Based on the emotion recognized, certain audio is played that compliments the user's emotion. For that different audio files were stored and as the emotion was recognized, the audio file was played depending on the emotion.

REFERENCES

[1] Fernando de la Torre, Tomas Simon Kruez, Zara Ambadar and Jeffrey F. Cohn, "FAST-FACS: A Computer-Assisted System to Increase Speed and Reliability of Manual FACS Coding", *Affective Computing and Intelligent Interaction,* 2011.

[2] Y. Tian, T. Kanade, and J. Cohn, "Recognizing Action Units for Facial Expression Analysis", CMU-RI-TR-99-40.

[3] P. Ekman and W. V. Friesen. *The Facial Action Coding System: A Technique For The Measurement of Facial Movement.* Consulting Psychologists Press Inc., San Francisco, CA, 1978.

[4] Y. Tian, T. Kanade, and J. Cohn, "Robust lip tracking by combining shape, color and motion", In *Proc. Of ACCV'2000*, 2000.

[5] M. Pantic 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 - Part B, vol. 36, no. 2, pp. 433–449, April 2006.

[6] M. Valstar and M. Pantic, "Combined support vector machines and hidden markov models for modeling facial action temporal dynamics," in Proceedings of IEEE Workshop on Human Computer Interaction. In conjunction with IEEE ICCV-07., 2007, pp. 118–127.

[7] Y. T. L. Zhang and Q. Ji, "Active image labeling and its application to facial action labeling", in European Conference on Computer Vision, 2008.

[8] F. de la Torre, J. Campoy, Z. Ambadar, and J. Cohn, "Temporal segmentation of facial behavior", in International Conference on Computer Vision, 2007.

[9] Castrillón Marco, Déniz Oscar, Guerra Cayetano, and Hernández Mario, "ENCARA2: Real time detection of multiple faces at different resolutions in video streams". In *Journal of Visual Communication and Image Representation*, 2007 (18) 2: pp.130-140.

[10] P. Ekman, W. Friesen, and J. Hager, "Facial action coding system: Research nexus." Network Research Information, 2002.