

Artificial Sensory Head

Sayani Manna

*M.Tech (CSE), RCC Institute of Information Technology, Beliaghata, Kolkata,
West Bengal, India
sayanimanna1309@gmail.com*

Abstract

This paper proposes a novel approach to prepare an Artificial Sensory Head which can recognize senses (speech, vision) and change its facial expression according to it. These are basic human expressions like happy, sad, angry, fear, surprise and disgust. The ASH receives input from user which may be in the form of speech or image. On receiving input from user of the system the type of emotion is detected. A bit pattern corresponding to each emotion is set. This generated bit is sent to the target application through usb port. The microcontroller is programmed with the logic to activate the driver at the end. In our methodology, in the end a face model is implemented as driver. This face model has some optical device which is placed at the positions where change would occur during change in facial expressions. These locations are identified from pre determined motion vector of still images and optical devices are placed accordingly. On detection of emotion, an impulse is sent to these optical devices and the ones placed at the points which changes during change in facial expression would glow denoting the direction and location of change in emotions. All throughout the process, we have a face model that can be activated by speech and image.

Keywords: Action Unit (AU), Actuator, Artificial Sensory Head (ASH), Feature point, USB, Atmega 8

1. Introduction

Many research works have been done and many are going on to detect various facial expression. Emotions play a fundamental role in human cognition and are thus essential in studies of cognitive science, neuroscience and social psychology. Emotions are expressed through different facial expressions. A facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person. It plays a communicative role in interpersonal relations. 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 creation of realistic animated faces is still one of the most challenging tasks for visual effect professionals. Many elements in a human face contribute to the realistic appearance of a facial expression. The shape of the mouth, the look of the eyebrows, as well as the gaze is just the most important clues that an observer perceives. These visual clues have also been used for recognizing facial expression.

We present a generic method that can transfer expressions from a source 3D face model (named as Artificial Sensory Head in our project) to a target 3D face model. ASH receives input in form of speech and produces output by changing facial expression. The various facial behaviours and motions can be parameterized based on muscle actions. This set of parameters can then be used to represent the various facial expressions. Till date, there have been two important and successful attempts in the creation of these parameter sets:

1. The Facial Action Coding System (FACS) developed by Ekman and Friesen in 1977.
2. The Facial Animation parameters (FAPs) which are a part of the MPEG-4 Synthetic/Natural Hybrid Coding (SNHC) standard, 1998.

Facial Action Coding is a muscle-based approach. It involves identifying the various facial muscles that individually or in groups cause changes in facial behaviors. These changes in the face and the underlying (one or more) muscles that caused these changes are called Action Units (AU). Thus an emotion is detected through change in different AUs.

2. Literature Review

In the past decades, an enormous amount of scientific work has been presented in the field of facial animation. All these methods include detection of various emotions from images and allow them for the transfer of a facial expression from a source face to a target face. This phenomenon is known as expression analogy. A target face (2D image or 3D model with/without texture) with the same expression as the source is necessary to perform expression analogy. For ease of acquisition, a neutral target face is usually employed in the related work. Existing technique for facial expression recognition can be broadly classified as [2] geometric-based approaches and appearance-based methods. Geometric-based approaches are techniques in which facial image is actually converted into an image with points on the various facial organs be it eyes, eyebrows, nose and lips or the shape of facial organs [3]. Classification is done by analyzing the distances between these feature points or the relative sizes of the facial components. Pantic *et. al.*, [3] proposed a method for detecting facial actions by analyzing the contours of facial components, including the eyes and the mouth. A classifier is then used to recognize the individual facial muscle action units (AUs). However, these methods require obtaining correct location of the feature points especially when the image is of low quality or with a complex background. Appearance-based methods process the entire image by applying a set of filters to extract facial features.

Another method named, FCA (facial coding algorithm) [4] defines an algorithm to recognize facial expression using ANN. Initially the frontal facial region is selected which forms the actually region of interest. Then the facial image is divided into two parts as the upper portion and lower portion of the face. The upper portion of the face includes eyes and eyebrows in [portion(X)] while the lower portion includes the nose and the mouth [portion (Y)]. The image is further segmented. Now portion (X) of the image is bilaterally divided. This forms portion (X) left and portion (X) right where former includes the left eye and left eyebrow while the latter includes the right eye and right eyebrow. Each portion is individually taken to form the bounding box and the Region of Interest (ROI) for the various facial organs. This is done using the CascadeObjectDetector. It also includes finding the ROI using in-plane rotation or rotation along left or right axis (out-of-plane rotation) or both. After finding the ROI we locate the facial organs using the same technique. Then we locate feature points on each of the facial organs. Each eyebrow has 3 points and each eye has 4 points. The points on the left eyebrow: P1, P2 AND P3. The points on the right eyebrow: P4, P5 AND P6. The points on left eye: P7, P8, P9 and P10. The points on right eye: P11, P12, P13 and P14. Similarly, feature points for the portion (Y) of the face are obtained. This portion (Y) includes nose and mouth. Mouth region has four points P15, P16, P17 and P18. In totality there are 18 feature points which are obtained from various facial organs and based on these facial points there are 16 facial vectors obtained and calculated based on the Euclidean distance (root mean square value). This is best method for obtaining the distance between two points. These distance vectors are given as input to the input layer

of the network and for classification technique feed-forward back propagation algorithm is used.

3. Problem Statement and Overview

In a nutshell, our project describes a method which is able to detect the type of emotion from the input and accordingly transform the emotions on ASH. Initially our ASH is a source face with neutral expression. Each facial expression for an emotion has some points which differ from the neutral face. These points are known as feature points which are to be identified for each facial expression. So to make our ASH express an emotion we change the particular set of feature points in specified direction for that emotion. The set of feature points are located at the following positions in our project.

1. On forehead.
2. On inner and outer eyebrow of both eyes.
3. On upper eyelid and lower eyelid.
4. On nose.
5. On cheeks
6. On upper and lower lips.
7. On jaws.
8. On chin.
9. On left and right corner puller muscle.
10. On left and right corner depressor muscle.

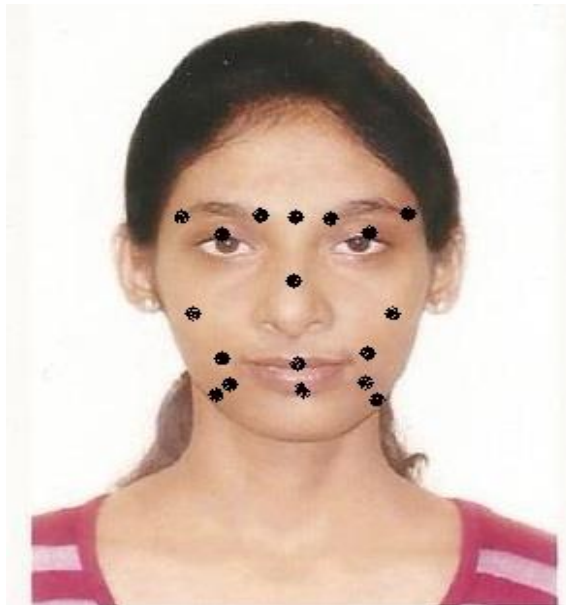

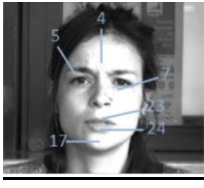
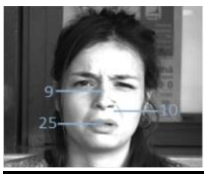
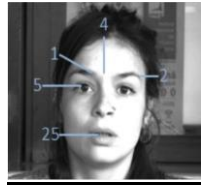
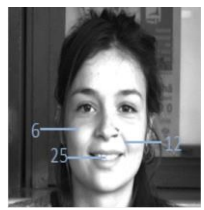

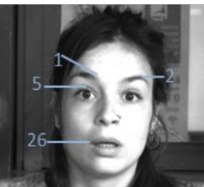


Figure 1. Feature Point are Identified and Marked on the Face

Table 1. Images of Different Emotion with Different Action Unit

Emotions	Image	Change in different Action Unit
Neutral		
Anger		<ul style="list-style-type: none"> • Brow Lower • Upper Lid Raiser • Lid Tighten • Chin Raiser • Lip Tighten
Disgust		<ul style="list-style-type: none"> • Nose Wrinkle • Upper Lip Raiser • Lips Part
Fear		<ul style="list-style-type: none"> • Inner Brow Raiser • Outer Brow Raiser • Brow Lower • Upper Lid Raiser • Lips Part
Happiness		<ul style="list-style-type: none"> • Cheek Raiser • Lip Corner Puller • Lips Part
Sadness		<ul style="list-style-type: none"> • Inner Brow Raiser • Brow Lower • Lip Corner Depressor • Chin Raiser
Surprise		<ul style="list-style-type: none"> • Inner Brow Raiser • Outer Brow Raiser • Upper Lid Raiser • Jaw Drop

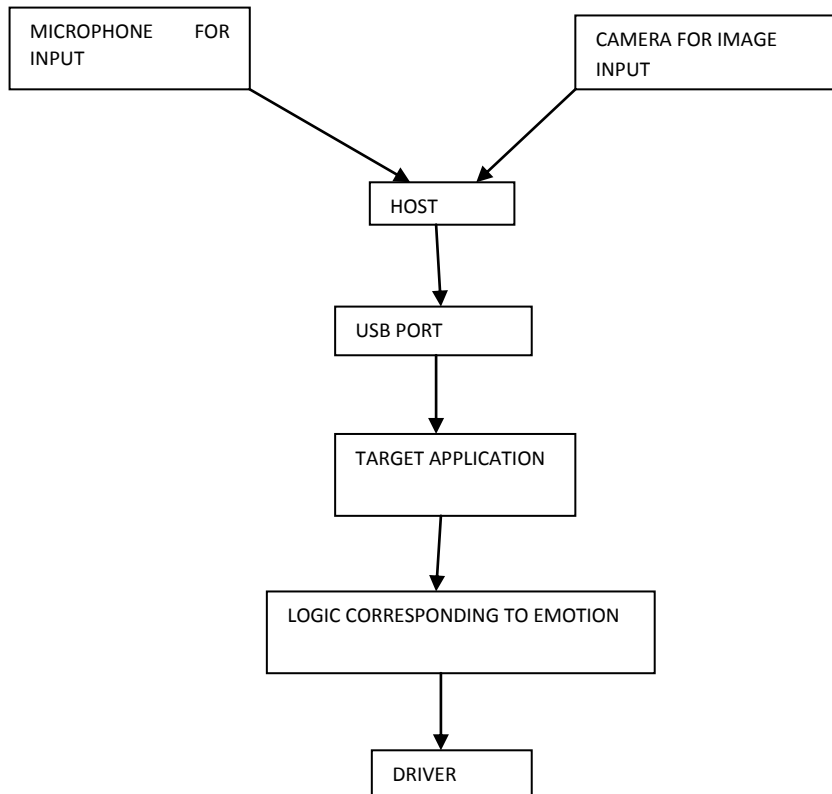


Figure 2. The Block Diagram of the Project

The project includes two parts: Software implementation And Hardware implementation

A. Software Implementation

The entire project is based on the type of input. The inputs are:

- Speech
- Still digital image

When speech is given as input, it is processed and accordingly the type of emotion is detected. Similarly when still image is provided as input it is matched and then accordingly the emotion is detected. After detecting emotion an 8-bit data is generated this carries the information regarding the point of change in our ASH and its direction of change.

B. Hardware Implementation

The hardware involved in our proposed methodology are microphone to provide speech as input to the system, camera for image input, USB port communication, ASH to implement the change in facial expressions and computer system to control the entire process. The process of hardware implementation includes interfacing the device with the port of computer to control the device through voice commands. For our project we have used USB port for communication between computer and hardware. In our ASH (Artificial Sensory Head) some actuators are placed at the feature points detected as points on our face model which involve in changing facial emotion for each emotion.

4. Result

The input to the system may be speech in form of voice command provided by user of the system. The presented methodology was tested with a set of voices (10 set of voice) collected from different people (including adult male, female and children). With this set of voices the recognition rate of speech is as follows.

Table 2. Recognition Rate of Voice Command

Spoken words for different facial expressions	Recognition Rate for speech (%)	Recognition Rate for still image (%)
Anger	80	90
Disgust	60	70
Fear	80	80
Happiness	80	90
Sadness	100	90
Surprise	60	80

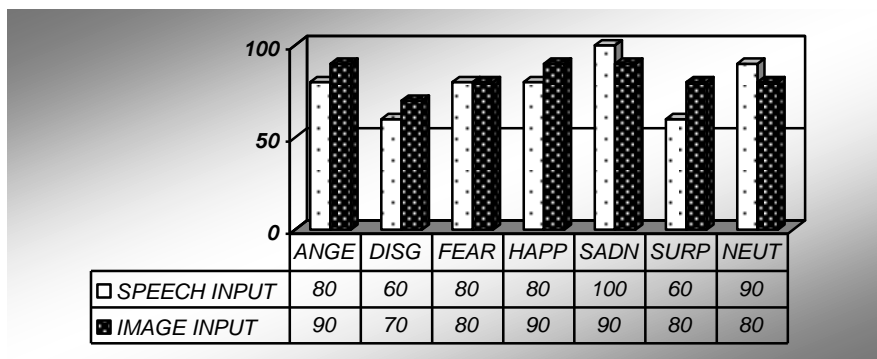


Figure 3. Graph of Recognition Rate

Table 3. Bit Pattern used for Emotions

Types of Emotion	Pattern used
Happy	1
Sad	2
Angry	3
Fear	4
Surprise	5
Disgust	6
Neutral	7

After recognition of the speech a bit pattern is generated this is passed on to the ASH from computer system through serial port. On receiving this bit pattern the change in emotion is expressed in ASH. Initially the ASH is expressed as neutral face. From neutral face it is transformed to target facial expression for recognized emotion. The average execution time for each emotion is noted as followed.

Table 4. Execution Time for Expression of Emotion by ASH

Type of emotion	Execution time (in sec.)
Anger	0.9
Disgust	1.2
Fear	1.0
Happiness	0.8
Sadness	0.8
Surprise	0.9

5. Conclusion

The main focus of the project is to develop animatrix based interacting objects. This is motivated from the trend of designing robots with human appearance, and also for the implications of their use in human environments. The project is designed in such a way that future modification can be done easily. Presently we are successfully able to identify the word being spoken over a dictionary of 6 words, *i.e.*, happy, sad, fear, anger, surprise and disgust when our input was speech. Emotions can also be detected from still image for different facial expressions such as neutral, happy, sad, fear, anger, surprise and disgust. After recognizing the emotion from our input (speech or still image) a bit pattern is generated correctly for respective emotions. The overall recognition rate of the system is 76.67%.

Advantage of Artificial Sensory Head:

- It can be used to communicate with different humans in our absence.
- It can respond on getting input from user by reacting with change in facial expression.
- It does not require any manual action to make it respond.

6. Future Work

Presently we have implemented the ASH by using speech and image as input but it can be extended by using various forms of sense such as touch, taste. The ASH considers six basic emotions such as angry, sad, disgust, happy, fear and surprise. This can be extended with combination of various emotions with its degree of expressing. The ASH shows change in facial expressions by optical devices which in future may be replaced by placing actuators. The delay of changing emotions may be decreased further.

References

- [1] A. Martinez, S. DuA, "Model of the Perception of Facial Expressions of Emotion by Humans: Research Overview and Perspectives", *Journal of Machine Learning Research*, vol. 13, (2012), pp. 1589-1608.
- [2] A. Mehrabian, "Communication without words", *Psychology today*, vol. 2, no. 4, (1968), pp. 53-56,
- [3] J. Mahajan and R. Mahajan, "FCA: A Proposed Method for an Automatic Facial Expression Recognition System using ANN", *International Journal of Computer Applications (0975 – 8887)*, vol. 84, no. 4, (2013) December.
- [4] Y.T. Yu and M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", *Journal of Systems and Software*, in press, (2005).
- [5] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art".
- [6] M. Song, Z. Dong, C. Theobalt, H. Wang and Z. Liu", "A Generic Framework for Efficient 2D and 3D Facial Expression Analysis", *IEEE TRANSACTIONS ON MULTIMEDIA*.
- [7] N. Mahajan and H. Mahajan, "Emotion Detection Algorithm", *International Journal of Electrical and Electronics Research ISSN 2348-6988 (online)*, vol. 2, no. 2, (2014) April – June, pp. 56-60.
- [8] http://www.hitl.washington.edu/research/knowledge_base/virtual_worlds/EVE/I.D.2.d.VoiceRecognition.html.
- [9] Md. A. A. Ahasan, Md. A. Awal and S. S. Mostafa, "Implementation of Speech Recognition Based Robotic System", copyright © 2011 *ijcit*. ISSN 2078-5828 (print) issn 2218-5224 (online), vol. 1, no. 2, manuscript code: 110108.
- [10] P. Li, S. L. Phung, A. Bouzerdom, and F. H. C. Tivive, "2010 IEEE Feature Selection For Facial Expression Recognition".

Author



Sayani Manna, M.tech in Computer Science from RCC IIT under WBUT and B.tech in Computer Science from RERF under WBUT.