Face Enable Mouse using Motion Detection and Speech Recognition

Prof. Kiran Kumari Sinha
Department of Information Technology
K J Somaiya College of Engineering,Vidyavihar (E),Mumbai – 400077.

Bhavesh Lalwani
Department of Information Technology
K J Somaiya College of Engineering,Vidyavihar (E),Mumbai – 400077.

Rishi Shah
Department of Information Technology
K J Somaiya College of Engineering,Vidyavihar (E),Mumbai – 400077.

Mayur Jalora
Department of Information Technology
K J Somaiya College of Engineering,Vidyavihar (E),Mumbai – 400077.

Abstract

In this contemporary world of human – computer interaction, where abatement of efforts is eulogized, 'Face Enabled Mouse using Motion Detection and Speech Recognition' provides a brilliant asset for an advanced interaction with the modern day computers. This system, in its' real world scenario, is capable of tracking the human head with the assistance of a simple yet an operative Face Tracking Algorithm. The objective of this project is to make the real time computing uncomplicated and more facile. Not only, will the computer conduct the actions as per the head movements but also correspond to the speech as input commands. This makes the system more versatile and easy to use through all the sections of the society.

Keywords: Perceptual User Interface, Face Detection, EEG (Electro Encephalograms), Blink Tracking, Open CV, Speech Recognition

I. INTRODUCTION

Human to human interaction has always been considered to be the most efficient means of communication. They rely on the physical characteristics of their bodies viz. face gestures, speech, hand gestures, etc. collectively, termed as "Biometrics". These days, there has been a rapid development in incorporating the biometrics in computers for easier access and optimizing the usability performance of computers. We have seen, how hands-free [4] Bluetooth devices have brought a tremendous revolution in way we handle the compact mobile phones. Similarly, this system is an endeavor to empower the computers to understand us even better.

Let's now graze through the entire spectrum of the biometrics currently established in human – computer interaction [3], [5]. Biometrics used can be easily characterized as 'Behavioral' and 'Physiological'. A behavioral biometric can be further classified as voice, key stroke and signature whereas, a physiological biometric can be further classified into face, fingerprint, hand, iris, DNA.

This system aims at having a complex amalgamation of both the primitive biometrics, viz. Voice from Behavioral biometric and Face from physiological metric. Though the amalgamation of these varied dimensional fields could be difficult, but in contrast to it, we can obtain a very useful and practically a viable solution to modern day computing.

Face Recognition will be the first and the most foremost area of interest for us. Since, in particular it has received tremendous attention from both the industrial as well as the research community. A face recognition system is the one that identifies gestures and understands them from the digitized images or videos that are captured by cameras in frames.

Speech recognition [2], [5] is by far the most widely accepted and used biometric amongst all of the above mentioned. Whether it is the Contextual Voice Commands by Google or the Voice Recognition by Microsoft we find a supreme acceptance of this biometric too. Easy availability of hardware viz. the microphone and rapid practical development has given a way to get an entirely new perspective of handling the computers.

This system is capable of integrating both these massive fields into a singular domain of computing.

II. PROPOSED SYSTEM

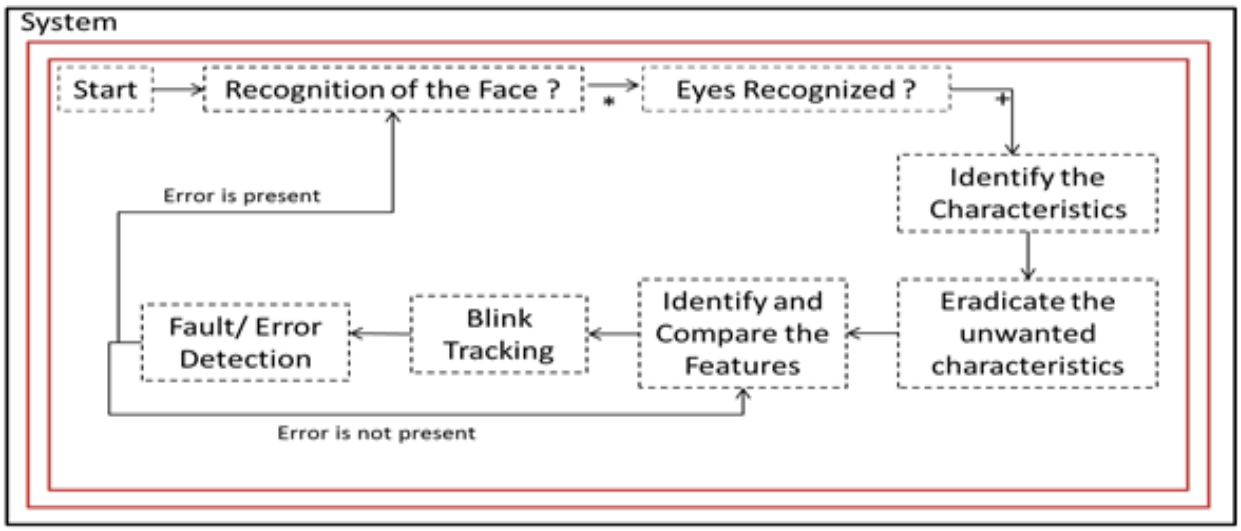
The entire system is going to be examined in two conditions, occurring subsequently and eventually resulting in the output.
The initial step of the system includes the face recognition. During this process, the facial arrangements considered are rotation of the head and blinking of the eyes. The three dimensional position of the head is tracked and converted to 2D coordinates on the computer screen; at the same time, intentional blinks are recognized and interpreted as an action. The tracker works uniquely using the real time video of the person sat in front of the screen.

Once, the tracker starts working, feature identification is activated. This feature identification works in the following way:

- Retrieve face location: the location of the face is retrieved thanks to the face detection algorithm from OpenCV.
- Identify eyes location: The location of the eyes is calculated relatively to the location of the face.
- Remove unwanted areas edges of the face
- Identify good features to track
- Movement Tracking

Once, the movement tracking is captured, certain sequence of operations are performed as follows:

- The number and the initial position of the facial features are obtained from the previous step.
- The initial configuration of the features is stored in the memory.
- The real-time video clip provided by the camera is analyzed frame by frame by the tracker in order to detect the movements of each single feature.
- For each feature the difference between its current and initial location is calculated, then the average of all the differences is calculated. In this way the tracker detects a small movement when the head performs a roll.



Index	Description
*	One Face is Recognized
+	Eyes are identified

Fig. 1: Block Diagram for Face Recognition Module

The system is intialiazed after starting the application, as soon as the application is started the face detection is executed, once the face is recognized the appplication starts executing the module in which the eyes are identified. Through the face line and eyes outline, the features of the entire face are initiliazed for example chin outline, ears position, lips periphery, etc. The entire face is then slowly recognized and displayed. The unwanted or bad features are eradicated through a removal module. The features which are required are retained and tracked. If, the face detection is facing some kind of error the entire face detection module is restarted and tracked again.

Speech recognition [5] technology has progressed to a level, where users can to some extend control actions on the computer using voice commands. Speech recognition[2], [5] has matured and with the ability to train the parameters of the model using training data to gain optimal performance, it represents a powerful solution. Our prototype system utilizes the Microsoft Speech SDK 5.1, which is available as a free download from Microsoft. By issuing the relevant voice command, the user may at any time perform common tasks, such as click and double-click operations. The need for double-clicking, for example, represents a great challenge for people with reduced dexterity and motor control in their hands. Our system not only enables the user to execute single or double-clicks, but also many more operation. For example, the user will be able to open different applications in his PC such as paint,word,music player and etc.

Voice commands used to activate buttons are:

- "HashClick" – single left click
- "HashDouble – double left click
- "HashRight" – single right click
- "HashWord" – to open a new word document
- "HashMusic" –to run the music application

The commands may be changed to suit the individual user. There is also scope to extend the set of commands for added functionality. We have also used the keywords uncommon so that the system is not confused and performs some unwanted action. For eg. We have used HashDouble instead of a simple command like double that is used in day to day lives of our daily communication.

A. Overview of the Components:

We are using both the hardware and software components for this entire application system.
The hardware components required are:

1) Microphone: A highly sensitive microphone for recording the voice commands for the respective actions to be performed.
2) Web cam: An in – built or external web cam for the purpose of capturing and displaying the face for recognition.
3) Displaying Interface: A monitor or a TV screen is essentially required for the purpose of displaying the results onto the displaying interface.

The software components are as follows:

1) Net Beans: A standard IDE for coding the entire application for both the speech recognition [5] as well as the face detection.
2) Microsoft Speech SDK 5.1: This SDK is important as far as receiving the voice input is concerned. The voice input is recorded and processed.
3) Microsoft Visual Studio: For assisting the Microsoft Speech SDK in performing the various actions to be performed on the Voice Commands received from the user.
4) OpenCV Libraries: It's used for feature tracking.

III. APPLICATION

This entire system takes a diligent effort in understanding the shortcomings of the computer technology which lacks supreme adaptive methods for the disabled. People who are quadriplegic and somewhat physically disabled, for example from cerebral palsy or traumatic brain injury or stroke, have limited motions they can make voluntarily. Some people can move their heads. Some can blink or wink voluntarily. Some can move the eyes or tongue. Unlike, the computers, family, friends, and other care providers usually detect these motions visually. So in order to make the computers a part of regular day to day lives of people in these disabilities our system will provide a new zest and zeal for such people to adopt this brilliant technology, the computers, into their control.

IV. CONCLUSION

In this fast paced world, wherein we are multi – tasking each and everything at much greater efficiency and force, optimum utilization of each and every technology and aspect is of vital importance, nonetheless, it's also imperative that we keep working with a multi – utility dimensional domains of technology. This singular domain of integration of the massive fields of speech and face recognitionis going to aid us in creating a digital world of free movement and less efforts. The higher the sensitive web camera the more quick and fast mouse pointer will react.

REFERENCES

[1] James Gips, Margrit Betke, and Peter Fleming "THE CAMERA MOUSE: PRELIMINARY INVESTIGATION OF AUTOMATED VISUAL TRACKING FOR COMPUTER ACCESS" presented at the RESNA 2000 Conference in Orlando. Published in Proceedings of RESNA 2000, RESNA Press, pp. 98-100
[2] Frank Loewenich and Frederic Maire "Motion-Tracking and Speech Recognition for Hands-Free Mouse-Pointer Manipulation" , pp. 550, November 2008, I-Tech, Vienna, Austria
[3] Sarnaik Supriya Vasantrao, Chalke Minal Prakash, Bombe Akshata Prakash, Anupama Anant "Improved HCI using face detection and speech recognition" from "International Journal of Advance Research in Computer Science and Management Studies" Volume 2, Issue 3, March 2014
[4] Akhil Gupta, Akash Rathi, Dr. Y. Radhika ""HANDS-FREE PC CONTROL" CONTROLLING OF MOUSE CURSOR USING EYE MOVEMENT" International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012
[5] Sarita "Performing Computer Operations with Voice Controlled Mouse" International Journal of Scientific & Engineering Research, Volume 4, Issue 4, April-2013