Hands-Free Gesture and Voice Control for System Interfacing

*Tanvish Thakker,
BE Student, Department of Information Technology,
K.J. Somaiya Institute of Engineering and Information Technology,
University of Mumbai, Maharashtra, India
tanvish.thakker@icloud.com

Shrenik Ashar,
BE Student, Department of Information Technology,
K.J. Somaiya Institute of Engineering and Information Technology,
University of Mumbai, Maharashtra, India

Vaibhav Chopda,
BE Student, Department of Information Technology,
K.J. Somaiya Institute of Engineering and Information Technology,
University of Mumbai, Maharashtra, India

*Nasim Shah
Assistant Professor, Department of Information Technology,
K.J Somaiya Institute of Engineering and Information Technology,
University of Mumbai, Maharashtra, India
nshah@somaiya.edu

Abstract: The proposed system presents a simple prototype system for real-time tracking of a human head and speech recognition for hands-free mouse. This system uses a simple yet an effective Face tracking algorithm. The Haar-classifier algorithm is used to capture the frames of the face and Lucas-Kanade algorithm for marking the features of a human face. The general requirements of a real-time tracking algorithm – it should be computationally economical, should possess the capability to perform in diverse environments and should be able to run itself with a very minimal knowledge about the preexistence of the faces in the head tracking algorithm. This system also makes use of Microsoft Speech SDK 5.1 for speech recognition. It is composed of two fundamental components – voice recognizer and speech synthesizer. The voice recognizer is used to capture the input of voice signals and speech synthesizer is responsible for lexicon management.

Keywords*— face tracking, speech recognition, brain computer interface (BCI), gesture, Lucas Kanade, Haar Classifier*

*****

I. Introduction

Develop a system that uses a camera to visually track a feature on a person's face, for instance, the tip of the nose and blinking of the eyes, and use the movement of the tracked feature to directly manage the mouse pointer on a computer, together with the speech recognition. All the operations of the mouse are done using the face movement and speech recognition. For face tracking, we are about to make use of Haar Classifier algorithm which can capture the frames of the face and Lucas-Kanade rule which can capture the features from the frame being captured; they are computationally cheap and possess the flexibility to perform in numerous environments[1].

For speech recognition, we are about to use Microsoft Speech SDK 5.1, that is offered as free software from MS. people that are a handicapped person and somewhat physically disabled, for instance from traumatic brain injury or stroke, have restricted motions they can make voluntarily. Some individuals can move their heads. Some can blink or wink voluntarily. Several computer access strategies are developed to assist people that are handicapped person and nonverbal external switches, devices to notice tiny muscle movements or eye blinks, head pointers, infrared or close to infrared camera primarily based systems to notice eye movements, conductor based systems to live the angle of the eye in the head, even systems to find features in electroencephalogram. These have helped many people access the computer and have made tremendous enhancements in their lives. Up to the present time, there are many people who are handicapped and physically disabled. These people don't have any dependable solution to access the computer and benefit from it [2]. The proposed system is inclined towards growing the capabilities of computer systems that allows these people to not only efficiently access the system but also make optimum use of it.

II. Literature Survey

EEG (Electroencephalograms):

The current computer interface set up of a mouse and keyboard needs the user to possess full use of his or her hands. Sadly, many people don't have sufficient use of their hands as a result of injury and are therefore unable to use a computer using traditional hardware. Some different interfaces are developed using electroencephalograms (EEGs) and eye motion, but these need a great deal of expensive hardware, require significant processing time, and solely offer the user-restricted control. EEG and the connected study of ERPs are used extensively in neuroscience, cognitive science, psychological science, nerve cell linguistics and psychophysiological analysis.

BCI (Brain-Computer Interface):

EEG is also used for system interfacing in the system as BCI (Brain Computer interfacing). Brain-computer interfaces (BCIs) acquire brain signal, evaluate them, and interpret them into commands that are relayed to output devices that do preferred actions. BCIs don't use traditional neuromuscular output pathways. The main goal of BCI is to switch or restore useful function to people disabled by neuromuscular disorders like amyotrophic lateral sclerosis, brain disorder, stroke, or spinal cord damage. From first demonstrations of EEG-based spelling and single-neuron-based device control, researchers have gone on to use electroencephalographic and diverse brain signals for more and more complex control of cursors, robotic arms, prostheses, wheelchairs, and different devices[3]. Brain-computer interfaces can also prove helpful for rehabilitation after stroke and for other disorders. In the future, they could augment the performance of surgeons or other medical professionals. Its future achievements will depend on advances in three crucial areas. BCI must be thoroughly checked and tested by people with disabilities to make sure that advancements can take place. However, there are issues associated with BCI such as research in BCI at the instant is at a reasonably basic level considering the complexness of the problem and BCIs are currently fairly inaccurate in terms of classifying neural activity. However, in the proposed system such information collection module is completely eliminated this implies proposed system primarily has two main modules that are gesture recognition module and voice recognition module [4]. The proposed system uses java as a platform for the development of gesture recognition module in which Lucas-Kanade rule is used for face detection and Haar classifier is used for feature extraction. Lucas-Kanade algorithm generates a rectangular box that is employed for face detection. Once the face detection step is completed Haar classifier rule is employed for feature extraction, which implies it, detects eyes, nose, and mouth in a face [5].

III. Materials and Methods

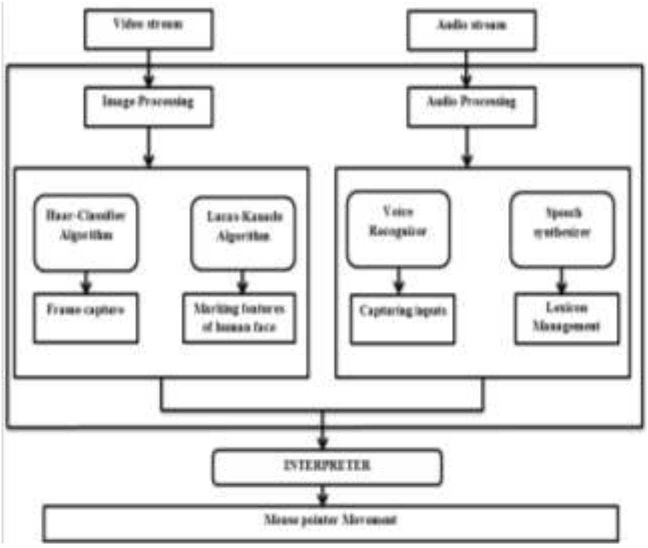

Fig: 4.1: Proposed System Architecture

The various sub-components are:

Video Stream: Video stream is captured through a web camera which feeds images in real time to a computer. This video stream is then saved and processed to detect cursor position.

Audio stream: A continuous voice stream is sensed in real time via a microphone which is then processed by speech synthesizer and speech recognizer module [6].

Image processing: In this step, multiple frames are captured via a web camera which is processed by a morphological operation which detects face position and move cursor accordingly.

Head tracking: Head tracking is done with help of Lucas-Kanade algorithm and Haar classifier. Haar classifier is used to generate a boundary which is primarily used for tracking the position of head and position of eyes. Lucas-Kanade algorithm is used for feature extraction which detects change is the position of eyes which in-turn result in the movement of a cursor on the screen.

Audio processing: Audio processing is an initial step of filtering audio so as to eliminate noise and isolate necessary information conveyed in speech.

Speech analysis: It is a process of extracting information conveyed in an audio stream which in turn is used to launch user specific application in real time.

Interpreter: It gives a combined output of both video processing module and audio processing module by interpreting and processing the results obtained.

Mouse pointer movement & button: Mouse pointer movement is based on coordinates provided by the interpreter. This allows the user to easily manage the system without using primary peripheral devices. To launch any application user can double click on the icon by winking the eye twice. This removes any limitation that specially-abled user may face while using the system [7].

A. *Techniques Used:*

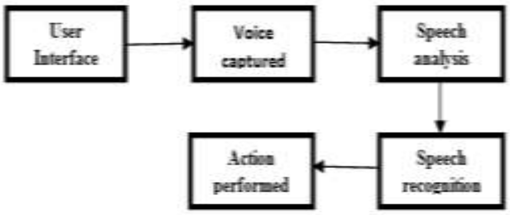

Fig 4(A)-1: Voice Module

Figure 4(A)-1 shows the overview of the speech module. When this module is run it will first check whether the microphone is connected or not, once it's detected voice will be captured, the captured voice will be analyzed and accordingly action will be performed.

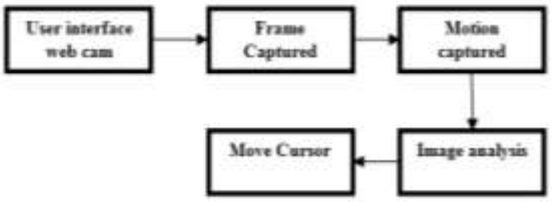

Fig 4(A)-2: Face Recognition Module

Figure 4(A)-2 shows the overview of face recognition module. When this module is running it will first check if the webcam is connected or not, once it detects the status of webcam frames are captured, after the frames are captured motion is captured and accordingly cursor moves.

IV. Results and Discussion

Face Processing:

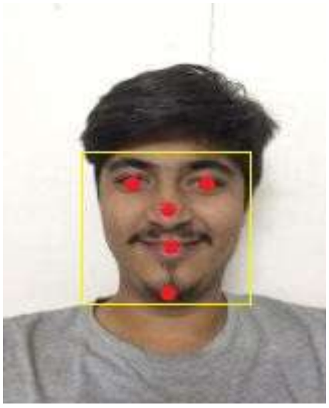

Fig 5.1

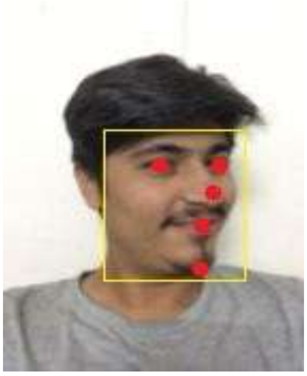

Fig 5.2

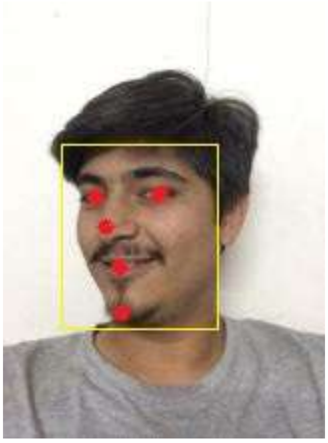

Fig 5.3

The yellow box around the face in the above photos represent the boundary of face detection and feature extraction for the head tracking system. The red dots represent the facial features that can be used to mark the face for the sole purpose of tracker movement. These points are x and y coordinates for the system and these features lie within the graph. If the features are found within the yellow box and if it matches the regular coordinates which are in approximation then the tracker display will say that the "Face is found" and the cursor will start moving according to the eye movement and if the features are not found then it will simply display "Face not found"[8].

Proposed System	Existing System
In proposed system recognition of gesture and voice is faster as well as reliable	In existing system like BCI recognition of brain activity is way slower due to noise in the environment
Hardware requirement of proposed system is less as it requires a webcam and computer system	Earlier system like BCI are complex as it depends on neural activity of brain and hardware required for detecting such neural activity is very expensive
Proposed system is user-friendly and way more accessible	Existing system like BCI are very time-consuming as it require various headgears

Table 1: Comparison of Proposed System with the Existing System

V. Conclusion

According to the research, it is very clear that the proposed system is much better than the existing systems in terms of cost, availability, and ease of use. The project presents a simple and effective low-cost optical system for implementing mouse operations using processed head motion. The system consists of several basic, off the shelf components including a webcam, a headset, and a computer, for handicapped people, and people with the certain illness. Images from the webcam are analyzed using a combination of software in order to determine the position of the user's head. A head position data is then transformed using a non-linear transformation into a corresponding screen position that is used to control the mouse pointer. Clicking operations can be accomplished using either the voice commands or the still Face position.

This project explores the field of computer vision with the broad aim of developing a system capable of interpreting the movements of human facial features.

VI. Acknowledgement

We would like to thank all those people who helped us in successful completion of this research paper on "Hands-Free Gesture and Voice Control for System Interfacing".

We would like to thank Dr. Suresh Ukarande, Principal of K.J. Somaiya Institute of Engineering and Information Technology, Sion for his immense support and motivation.

We would like to thank Mr. Uday Rote, Head of the Department, Information Technology of K.J. Somaiya Institute of Engineering and Information Technology, Sion for his guidance. We have learnt so many things from him and he motivated us and strengthened our confidence in doing the thesis. We express our deepest gratitude for his valuable suggestions and constant motivation that greatly helped the project work to get successful completed. Throughout the project work, his useful suggestions, constant encouragement has given us a right direction and shape to our learning. We would like to thank Mrs. Nasim Shah, Assistant Professor of K.J. Somaiya Institute of Engineering and Information Technology, Sion for her guidance in latest technologies of image processing throughout the thesis.

References

[1] Alvarez, L. (2012). Progress in pattern recognition, image analysis, computer vision, and applications. 1st ed. Berlin: Springer.

[2] Estellers, V., Gurban, M. and Thiran, J. (2012). On Dynamic Stream Weighting for Audio-Visual Speech Recognition. IEEE Transactions on Audio, Speech, and Language Processing, 20(4), pp.1145-1157.

[3] Hussen Abdelaziz, A., Zeiler, S. and Kolossa, D. (2015). Learning Dynamic Stream Weights For Coupled-HMM-based Audio-visual Speech Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, pp.1-1.

[4] International Conference on Image Processing. (2010). IEEE Transactions on Audio, Speech, and Language Processing, 18(1), pp.212-212.

[5] Kim, W., Suh, S., Hwang, W. and Han, J. (2014). SVD Face: Illumination-Invariant Face Representation. IEEE Signal Processing Letters, 21(11), pp.1336-1340.

[6] Khachidze, V. (2012). Contemporary research on E-business technology and strategy. 1st ed. Berlin: Springer.

[7] Sarita, S. and Kumar Kaki, K. (2013). Mouse Cursor & Movements using Voice Controlled Mouse Pointer. International Journal of Computer Applications, 71(7), pp.27-34.

[8] Yupapin, P., Hamedi, Salleh, Tan, Ismail, Ali, Dee-Uam, and Pavaganun, (2011). Human facial neural activities and gesture recognition for machine-interfacing applications. International Journal of Nanomedicine, pp. 3461.