

Virtual Mouse for Physically Disabled Using MediaPipe and PyAutoGUI

Indrajith M Dinesh

Dept. of Computer Science and Engineering
Mangalam College of Engineering
Ettumanoor, India
indranseason@gmail.com

Kenus Roy

Dept. of Computer Science and Engineering
Mangalam College of Engineering
Ettumanoor, India
kenusroy33@gmail.com

Muhammad Fairouz

Dept. of Computer Science and Engineering
Mangalam College of Engineering
Ettumanoor, India
fairouz2150@gmail.com

Suraj Tiwari

Dept. of Computer Science and Engineering
Mangalam College of Engineering
Ettumanoor, India
surajtiwari96059@gmail.com

Dr.Padmamal .S,
Professor,

Dept. of Computer Science and Engineering,
Mangalam Collage of Engg, Ettumanoor,
Kottayam Dist, Kerala
splaal71@gmail.com

Abstract— The Eye Blink Detection System focuses on the interaction between people and computers, it provides an interface between them. The main goal of our paper is to design a software that helps the people with physical disabilities to interact with computers in an easier and more convenient way. The main components are image processing to detect the eyes, and eye-blinking, face movements and opening and closing of the mouth. An eye blink is used as the click on the mouse. All the functions of the mouse can be performed by the proposed system. OpenCV is used for image processing after the image is captured from the webcam. Blazeface model of the MediaPipe is used for creating face mesh and feature extraction of the captured image. PyAutoGUI is a library provided by python for programmatically controlling mouse and keyboard.

Keywords—Eye blink, Head Tracking, Face Mesh

I. INTRODUCTION

In today's world each and every day different innovations are happening in the field of computer and technology. But the people with physical disabilities are unable to make use of these computer technologies due to the absence of a convenient way for allowing them to use computers. The existing system for the physically disabled to interact with the computers is a virtual keyboard [1] that allows them to type characters. This system uses the computers camera to capture the images for eyeblink. There is a virtual keyboard that highlights each character in the keyboard sequentially and the user have to blink their eye when the desired character is highlighted to perform the enter function and print that character on the screen.

The existing system is too much time consuming for typing each character as the user have to wait until the desired character is highlighted. Then only the user will blink his eye to enter the character on the screen. And this system can only be used for typing purpose no other functions can be performed using this existing system.

Brain-computer interface (BCI) technology works on signals from the brain. The electrical activity of the brain is determined by an electroencephalogram (EEG). A special cap sensor is placed on the scalp to read the signals, which are then transmitted to the computer. Next, the electrical activity of the brain is translated into a command to perform the action required by the BCI. A study in implemented a virtual keyboard based on BCI components with an eMotiv EPOC Neural Headset. The current implementation of the proposed virtual keyboard still needs some improvements to expand the accuracy and selection rate. Another virtual keyboard was introduced that used a brainwave sensor to connect and write. However, current systems might not have enough control precise for gaze, good control provides the ability to use available commercial eye-tracking based systems. Therefore, some studies and experiments have been built based on the rapid serial visual presentation (RSVP) model, which does not require exact gaze control to differentiate between different characters. The paradigm is called an RSVP KeyboardTM, an EEG-based BCI typing system. The RSVP KeyboardTM gives the user the capability to sequentially scan the options until the

desired symbol is chosen. However, RSVP has limitations on presentation speed and selection speed. Generally, BCI technology has many disadvantages, such as cost and the time to read the signal and give the required action, and system setup is not an easy process.

With the vast adoption and application of Artificial Intelligence (AI), we are witnessing how technology has become an integral part of our everyday life and how it makes people's lives easier. AI advancements increasingly impact society. Artificial Intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment. Human Computer Interaction (HCI) is an emerging technology field focused on designing and enhancing the interaction process between humans and computers. HCI is currently being shaped while also shaping AI applications, leading to the rapid emergence of new, exciting technology and research topics. The field contains different techniques to develop systems that can meet users' needs and requirements. HCI is widely implemented in many areas, such as medical technologies, robotics, urban design, gaming, and assistive technologies. AI and HCI have an apparent positive effect on the lives of people with disabilities in many ways, either potential therapeutic or non-therapeutic users of advanced applications.

In this world filled with computer technology it is so difficult for the physically disabled to use a computer system. The voice command system may be used for this purpose but the accuracy it can provide is doubtful and the language is another big problem even though English is a universally accepted language many people are still not good at English so the voice recognition system should be in their own native language.

The system that we propose is such that it controls the movements and clicks of the computers mouse using the users face movements and eye blink respectively. Most of the operations in the system can be done using the mouse such as opening files, directories and even typing using the virtual keyboard that is present in almost every system. The typing operation can be done using the virtual keyboard already existing in the system and click functions are performed using the cursor which is controlled by the users face and eyes.

II. LITERATURE REVIEW

A real-time algorithm to detect eye blinks in a video sequence from a standard camera is proposed. Recent landmark detectors, trained on in the wild data sets exhibit excellent robustness against a head orientation with respect to a camera, varying illumination and facial expressions. We show that the landmarks are detected precisely enough to reliably estimate the level of the eye opening. The proposed algorithm therefore estimates the landmark positions, extracts a single scalar

quantity eye aspect ratio (EAR) characterizing the eye opening in each frame. Finally, an SVM classifier detects eye blinks as a pattern of EAR values in a short temporal window. The simple algorithm outperforms the state-of-the-art results on two standard data sets.

A vision-based human-computer interface is presented in the paper. The interface detects voluntary eyeblinks and interprets them as control commands. [4] The employed image processing methods include Haar-like features for automatic face detection, and template matching based eye tracking and eye-blink detection. [3] Interface performance was tested by 49 users (of which 12 were with physical disabilities). Test results indicate interface usefulness in offering an alternative mean of communication with computers.

The implementation work underlying this system for pupil identification uses raspberry pi board to control the cursor of the personal computer and moreover Eye Aspect Ratio technique is ascertained along with OpenCV to detect the pupil. This system tracks the eye movements of the user with an IP cam (Internet Protocol camera) and simulates the eye movements into mouse cursor movements on screen and also detects user's eye staring on icon and will translate it into click operation on screen. The main aim of this system is to help the user to control the cursor without the use of hands and is of great use especially for the people with disability.

The contribution of the work is an alternative input device for those who have a motor disability and are challenged by traditional input devices. The advantages of a virtual keyboard based on BCI are summarized and we describe its design and implementation. We also present the results of a preliminary study that has suggested several improvements for enhancing the effectiveness of the virtual keyboard.

Humans need communication. The desire to communicate remains one of the primary issues for people with locked-in syndrome (LIS). [10] While many assistive and augmentative communication systems that use various physiological signals are available commercially, the need is not satisfactorily met. Brain interfaces, in particular, those that utilize event related potentials (ERP) in electroencephalography [7] (EEG) to detect the intent of a person noninvasively, are emerging as a promising communication interface to meet this need where existing options are insufficient. Existing brain interfaces for typing use many repetitions of the visual stimuli in order to increase accuracy at the cost of speed. However, speed is also crucial and is an integral portion of peer-to-peer communication; a message that is not delivered timely often loses its importance.

III. PROPOSED SYSTEM DESIGN

The proposed system is a virtual mouse that controls the movement of the cursor by tracking the facial movements and the eye blink of the user of the system. Head tracking and eye blink detection are the main functions or components of the system. The image of the user's face is captured from the web cam and OpenCV is used for the processing of the captured image. The captured image is converted from BRG to RGB

using OpenCV. Machine learning is used to detect the user’s facial landmarks, the Blaze Face Model is used for the feature extraction of the image. MediaPipe generates the face mesh model with 468 3D landmarks. The generated face mesh and the facial landmarks are used to locate the position of the head. Based on the previous position of the head the relative movement of the head is calculated, find the 0th landmark in the first frame then find the 0th landmark in the second frame, relative distance between the two is calculated to find the movement of the head. The mouse movement is performed using PyAutoGui, based on the relative movement the cursor is moved. To perform right click or left click the blink of the corresponding eye is checked.

The full form of OpenCV is Open Source Computer Vision Library. It is an open-source library for machine learning and computer vision. The main purpose of OpenCV is to provide computer vision with a shared infrastructure for applications. OpenCV makes it easier to perform complex tasks such as recognizing and identifying faces, identifying objects, classifying human actions in videos, track moving objects. In this paper we are making use of OpenCV to recognize human face, track the movement of the head and to convert BRG image to RGB image.

MediaPipe face detection is based on Blaze face, a lightweight and well performing face detector tailored for mobile GPU inference. MediaPipe face mesh is solution for mobile devices that estimates 468 3D facial landmarks in real-time. GPU acceleration throughout the pipeline together with the lightweight model architecture that provides a solution for real time performance that are critical for live experience.

PyAutoGUI is a python module that can be used across different platforms for GUI automation. Mouse and keyboard can be programmatically controlled using PyAutoGUI. Different operating systems have different ways to programmatically control mouse and keyboard. The PyAutoGUI hides all the complexities behind a simple API.

The web cam of the system captures the images and the live video of the user of the system. For image processing OpenCV is used. The images from the web ca are captured in BRG format. OpenCV converts the BRG format of the image to the standard RGB format. The image processing is performed by OpenCV. Through the USB bus the processed image is passed on to the MediaPipe face mesh solution for face detection [2] and for plotting facial landmarks. MediaPipe face mesh [9] is solution for mobile devices that estimates 468 3D facial landmarks in real-time. The image is resized into 256*256 using mediapipe solution. The Blaze Face model of the mediapipe is used for feature extraction of the image. The area of the face is cropped on from the given input image, then feature extraction is done suing the attention mesh model. The face mesh is created using the facial landmarks that includes face mesh, face contours and face mesh iris. Face mesh is the overall face landmarks that are spread all over the face. Face contour represents the edges or the boundaries of the face. Face mesh iris represents the face mesh that is concentrated on the eyes of the used that is usefull to detect whether the used have closed their eye or not.

The facial landmarks that are created by Blaze face model is pass on to perform the mouse actions with the help of

PyAutoGUI. PyAutoGUI is a python module that can be used across different platforms for GUI automation. Mouse and keyboard can be programmatically controlled using PyAutoGUI. Different operating systems have different ways to programmatically control mouse and keyboard. The PyAutoGUI hides all the complexities behind a simple API. It is a library provide by mouse and keyboard.

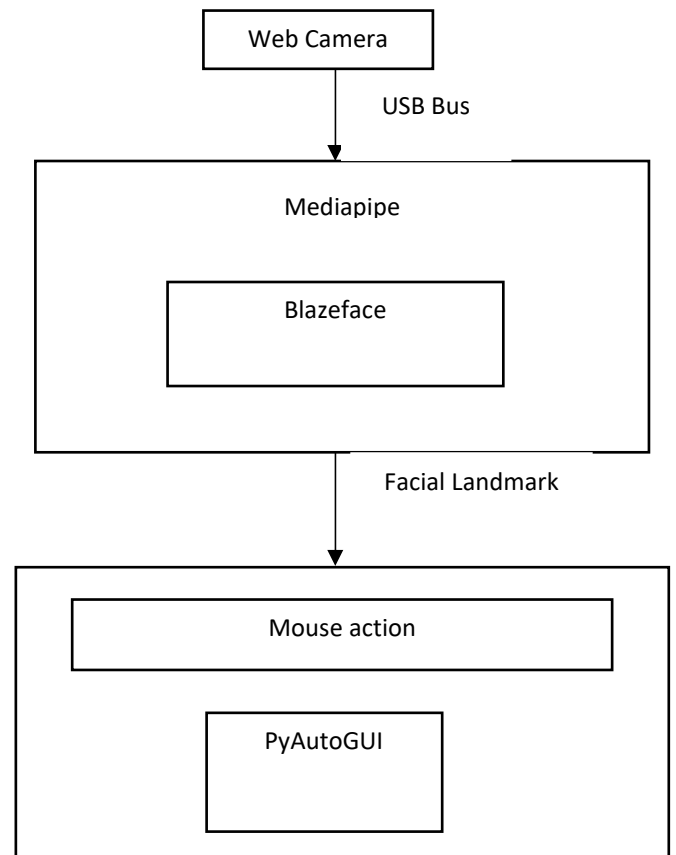


Fig 1. Proposed Architecture Diagram

The Fig 2. represents the system flow diagram of our proposed system. The flow of the working system is represented using the system flow diagram. At first images are captured from the web cam, image processing is done using OpenCV and BlazeFace [8] model of the mediapipe solution does the face recognition, feature extraction, creates a face mesh with facial landmarks. If the face is detected from the received image, then the relative position of the head is calculated otherwise, if there is no face or head in the received image then the next frame of the video is passed through the steps mentioned above. [8] Now if a face is detected then after calculating the relative position of the head the mouse pointer is moved to the corresponding position.

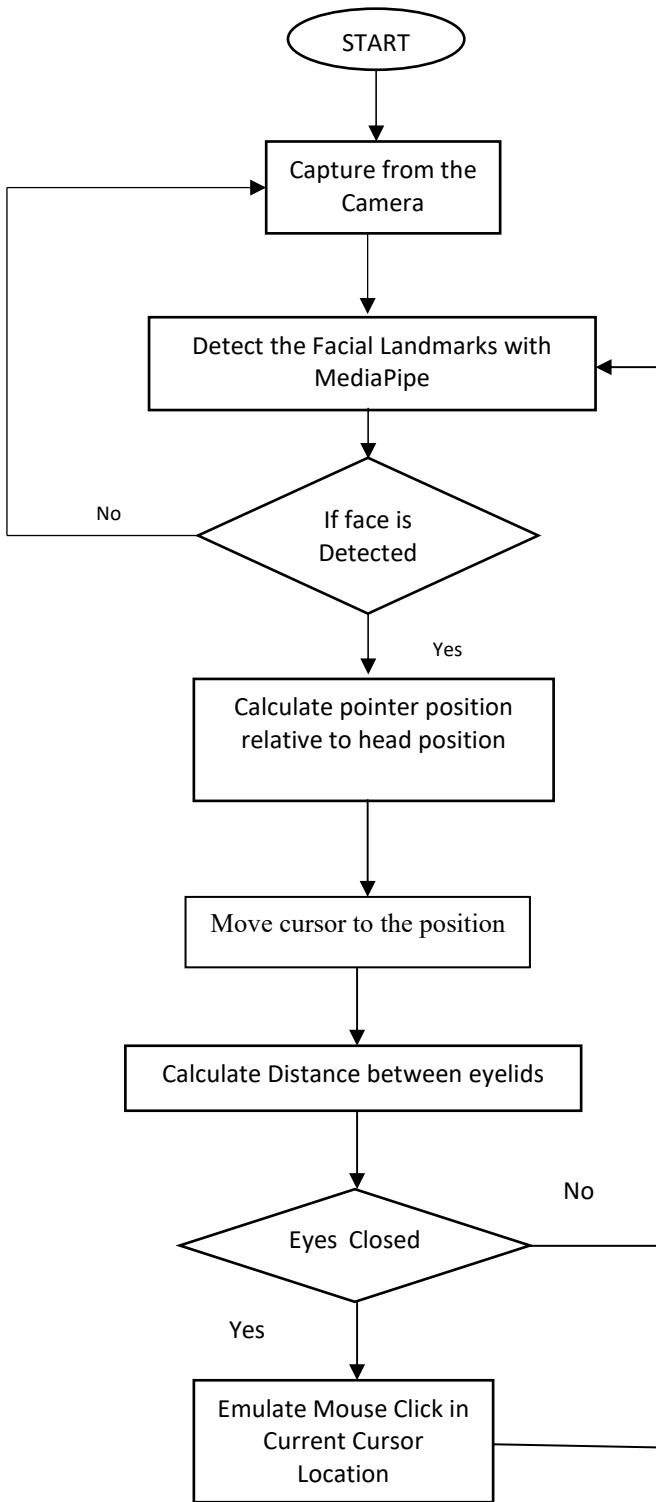


Fig 2. Proposed System Flow Diagram

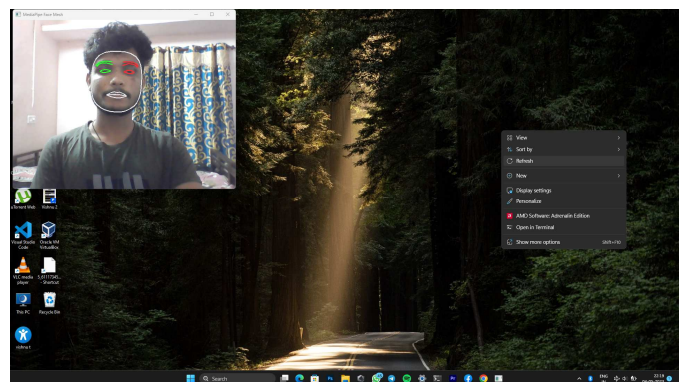
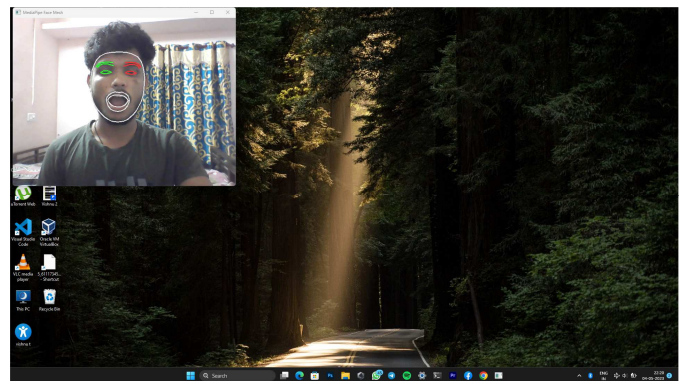
Now the next step is to calculate the distance between the eye lids. It is to know whether the eyes are blinked or not so as to perform the click operation on the mouse pointer. If the eyes are blinked then the mouse click on the corresponding cursor position is emulated. If the right eye is blinked then right click is emulated on the current position of the cursor. If the left eye is blinked then left click is emulated on the current position of the cursor. If the eyes are not closed then the

next frame is taken as input and it undergoes same steps as mentioned above.

IV. RESULT AND DISCUSSION

Currently the existing software for the physically disabled people to interact with a computer system is a virtual keyboard that highlights each character in the keyboard sequentially. The user has to blink his eye when the desired character is highlighted so as to enter the character on the screen. This method is so much time consuming and it allows the user only to type on screen.

The proposed system allows the user to do more tasks in the computer as it virtualizes the mouse. It allows to do operations such as opening files and directories, and all other operations that can be done using a mouse of the computer. And the virtual keyboard present in most of the computers can also be used by means of this virtual mouse.



When the mouth of the user is open the cursor will not move. The live video is seen to the user with drawings on their eye mouth and the contours of the face. The above given snapshot shows the proposed systems working.

V. CONCLUSION

The systems developed for the physically disabled have the potential to bring a big change in their life. Main focus of this system is to make the physically disabled people independent to make use of the available data and features of the computer, so that they could perform their work by their own without asking for anyone's help and get familiarized to the technologies available around them. Method proposed in this paper allows them to use their head to control the mouse so as to use the computer system. The accuracy of the eye blink can be increased in future so that the system performs more smoothly.

REFERENCES

- [1] A. Z. Attiah and E. F. Khairullah, "Eye-Blink Detection System for Virtual Keyboard," *2021 National Computing Colleges Conference (NCCC)*, Taif, Saudi Arabia, 2021, pp. 1-6, doi: 10.1109/NCCC49330.2021.9428797.
- [2] D. S. Brar, A. Kumar, Pallavi, U. Mittal and P. Rana, "Face Detection for Real World Application," *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, London, United Kingdom, 2021, pp. 239-242, doi: 10.1109/ICIEM51511.2021.9445287.
- [3] M. Pandey, K. Chaudhari, R. Kumar, A. Shinde, D. Totla and N. D. Mali, "Assistance for Paralyzed Patient Using Eye Motion Detection," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697455..
- [4] C. -L. Xu and C. -Y. Lin, "Eye-motion detection system for mnd patients," *2017 IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, Mauritius, 2017, pp. 99-103, doi: 10.1109/ISCMI.2017.8279606.
- [5] A. Rahman, M. Sirshar and A. Khan, "Real time drowsiness detection using eye blink monitoring," *2015 National Software Engineering Conference (NSEC)*, Rawalpindi, Pakistan, 2015, pp. 1-7, doi: 10.1109/NSEC.2015.7396336.
- [6] A. Udayashankar, A. R. Kowshik, S. Chandramouli and H. S. Prashanth, "Assistance for the Paralyzed Using Eye Blink Detection," *2012 Fourth International Conference on Digital Home, Guangzhou, China*, 2012, pp. 104-108, doi: 10.1109/ICDH.2012.9.
- [7] U. Orhan, K. E. Hild, D. Erdogmus, B. Roark, B. Oken and M. Fried-Oken, "RSVP keyboard: An EEG based typing interface," *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 2012, pp. 645-648, doi: 10.1109/ICASSP.2012.6287966.
- [8] Blaze Face: Sub-millisecond Neural Face Detection on Mobile GPUs Valentin Bazarevsky Yury Kartynnik Andrey Vakunov Karthik Raveendran Matthias Grundmann Google Research 1600 Amphitheatre Pkwy, Mountain View, CA 94043, USA
- [9] J. Cech and T. Soukupova, "Real-time eye blink detection using facial landmarks", *21st Comput. Vis. Winter Work.*, 2016
- [10] Eyeball movement based cursor control using raspberry Pi, Pantech Prolabs technology beyond the dreams, learning center, 2020. [Online]. Available:<https://www.pantechsolutions.net/eyeball-movement-basedcursor-control-using-raspberry-pi>.