

Controlling Mouse and Virtual Keyboard using Eye-Tracking by Computer Vision

Mr. R. Raj Bharath, M. E.(Ph.D)

Associate Professor

Computer Science and Engineering

ManakulaVinayagar Institute of Technology, Puducherry, India

Dinesh Kumar .L, Jayasuriya .C, Meenatchi Sundaram.T

Computer Science and Engineering

ManakulaVinayagar Institute of Technology, Puducherry, India.

Received: 2022 March 15; **Revised:** 2022 April 20; **Accepted:** 2022 May 1

ABSTRACT - The most common used input device in a computer is the virtual keyboard and mouse, whereas the usage of the same input device is complex for people suffering from (Paralyzed and persons with disabilities) by enabling them to carry out basic functions used in the conventional input system. The proposed system provides an alternate solution for the people who are suffering from paralysis and with physical disabilities by using their facial expression through the web camera as the basic input system instead of a physically handled virtual keyboard and mouse. The system works by facial expressions like movement of eye-ball and mouth movement. By applying Haar classifier the system identifies the region of the face, eyes, and mouth are detected and extracted for processing. It controls the action of the virtual mouse and virtual keyboard by providing a hands-free interaction between humans and computers. It allows disabled people to scroll up, scroll down, scroll left, scroll right, right-click, left-click, and to perform cursor movement in virtual mouse. The virtual keyboard portion is selected either as a left or right position of the eye-ball. The desired key is selected from the virtual keyboard by blinking the eye the user can type without the need for fingers or hands.

Keywords: Eye-tracking, Computer vision, Haar Classifier, Face Detection.

1.INTRODUCTION

Technology is an important evolution that enables human beings to explore the life and live life. Now a day's usage of the personal computer system has been increased, but some peoples like people with disabilities are unable to operate the computer system virtualKeyboard and mouse in same time. The mouse and virtualkeyboard can be operated by moving the eye by means of computer vision. A human's face transmits a lot of information about the facial expression, in order to operate a mouse and a virtualkeyboard function using the face, eyes and mouth are detected. The visual control framework with the computers will make them work without the assistance of anyone else. Human computer interaction is an implementation of a vision-based system for detecting eye movements through this technology, it was a boon for a person with a different impairment (one hand people).

Computers arrived in 1938 to comfortably improve our lives. There have been stringent improvements. They originally put in a lot of effort to find solutions to word processing and arithmetic difficulties, and now they've come a long way towards making our lives an essential component of their personal and

professional requirements, such as Internet browsing, link the world via social media. In due time, operations on Computers can simply be done by people with hands and eyes and can operate mouse and virtualkeyboard and also can see, what's on the screen. Blind individuals increasingly utilise computers with spoken text alternatives to speak-out the screen's layout. However, since it has a mouse and mouse characteristics, persons without hands cannot use it.

By introducing technology and making them compatible to their computers, they don't have to get down from their dreams and expectations because of disability that they can learn and live.

The idea of creating natural interactions between people and machines has gained popularity recently. There are many studies for universal computing's human-computer interface. The vision-based interface technology uses an input video picture to extract motion data without the need of expensive equipment. Eye and face tracking, as well as their identification, are done in order to construct a vision-based bidirectional human-computer interaction system. All common gadgets need manual operation and are inoperable for people with

mobility disabilities. It is necessary to establish other means of human-computer interaction that are appropriate for people with motor disabilities and that would enable them to participate in the information society.

The installation of a vision-based system for removing the action of eye and facial movements as a Human-Computer Interface for individuals with impairments is given. In order to control a – anti human-computer interface, the proposed work incorporates face identification, face tracking, eye blink detection, and real-time blink sequence interpretation. Using the human face & eye motions to interface with the computer instead of the conventional mouse. It is designed to make computer usage efficient and simple for those who are physically disabled and lack hands.

It deals with how computers can gain a deeper understanding of visual data. The study of computer vision focuses on developing electronic systems that can interpret, analyse, and comprehend visual data (such as photos or videos) in a manner similar to that of humans.

Computer vision allows computers to sense their environment and recognize objects like the human eye. To detect,

discriminate, and categories objects, computer vision employs machine learning techniques and algorithms. One of the hardest challenges in computer vision is object recognition. Humans perceive things since they are and remember what they recognize, storing it in their brains until people come across it again. Human vision is the process through which the eyes detect and work together with the intellect to convert them into pictures. Humans easily recognize- items and have little trouble explaining them in a situation. The computer vision generates the information from an image or a video.

2. RELATED WORKS

In this section we will discuss about the existing work done related to this model and will have a brief look at their advantages and drawbacks, functioning, proposals etc.

Mehta et al., proposed the detection of Drowsiness presents an algorithm which takes account of the different drivers when driving drowning detections in real time. A deep waterfall convolutions network has been developed to detect the face area that prevents low accuracy due to artificial extraction. The attributes of the face in an image are found on the basis of the Dlib toolbox. Driver Sleepiness is calculated on

the basis of eye dots using a new metric in called the Eyes Aspect Ratio. The Eyes Aspect ratio is qualified by the same fatigue designation using vector support machines(SVM) suggested by Wang et al. Sharma et al., projected the additional power of EGT-based and EMG-based cursor control systems also resulted in the development of a system that integrates all kinds of input users and makes a more effective use of the cursor in different circumstances. The hybrid EMG/EGT method is ideal for the use of gradable (stop) position controls from the EMG subsystem for the small-cursor moves in the restricted area of a cursor site. The cursor accuracy and usefulness of the EMG evaluation shown by the hybrid approach is inherited.

Pantic et al., proposed a first study of Paikinje's image of the visual field, based on certain points of reference, reveals that a detailed ocular and pupilometer system that doesn't disturb head movements or intervene visually, is possible. Where two or more reference points are in the centre of the visual field, the eye camera picks up the picture of reference points and the pupil overlaps. In this study, four IR LED's are used as references to not disturb the person's focus. The four reference

points serve as highlights in the pupil picture taken.

Balamurugan et al., projected the image capture mechanism is stored when the palm is sensed by the sensor. Input for the further processing shall be the processed images. The method of detection is then focused on the centres and borders. The hand detection moves the removed features, used for the next step for the only caught frame.

Swaroop et al., provides an analysis on the measures of pupil size differences dependent upon visual focus used for the identification of mental and cognitive approaches (e.g. measurements, fixtures, shootings, etc...). We would analyse their involvement and the processing processes used for a detailed emotional and cognitive evaluation. Details were collected and data specifications and other similar information established that were open to the public and used for the relevant projects of study. The accuracy in eye-tracking are diverse methods and other instruments (e.g., bio-signals) are needed to be improved.

According to Lee et al., the two most fundamental mouse operations are clicking and moving the mouse. With the use of an OpenCV, cutting-edge technology switches out this mouse movement for eye

movements. Any facial emotion, including such blinking eyes, expanding lips, or head movement, will activate the mouse button. To replace all functions of standard console mouse and virtualkeyboard operations, the camera mouse is used. The suggested framework can provide allvirtual keyboard and mouse click operation occurrences and console capabilities.

In this strategy, the camera mouse framework alongside the clock goes about as left snap occasion and blinking as right snap occasion. The continuous eye-stare assessment framework is utilized for an eye-controlled mouse and to operate thevirtual keyboard for helping disabled persons.

The action primarily uses facial aspects to interface with the PC, so there will be no need is for hands to use the mouse andvirtual keyboard, according to Ramesh et al., who are trying to compensate the requirement of people who have hands incapacities and are unable to use PC resources without help.

The loss of muscular function in a specific area of your body is known as paralysis. It occurs when there is a problem in the transmission of signals between the brain and muscles. The muscles surrounding the

eyes, as suggested by Vahid et al., are the only ones that can govern movement when this occurs.

According to Hossain et al., Real-time hardware & mouse cursor movement, EOG signal capture and amplification, and analogue to digital conversion. Five disposable Ag-Agcl electrodes are used to measure the potential differences in between cornea and retina in order to identify eye movement. The EOG signal's frequency range is thought to be 0.3 to 15Hz, hence this bandwidths is captured using an active low pass and high pass filter in order to produce an accurate EOG signal. The output from the f there is turned into signal serialisation from the analogue output of the EOG. Arduino offers a point that is utilised for hardware control and serialises the EOG data for calibration. The Classification module, for example, Machine Vector Machine (SVM) and Linear Discriminant Analysis (LDA), classify data in terms of its horizontal and vertical dimensions.

IoanaBacivarov, Physical Activity A 2D affine face model that can rapidly match the texture and contour of a detected facial area is offered by models. The eyes are one example of a subregion of a recognised face that may have similar

statistical models created. Given that the eyes are one of the most communicative aspects of the human face, knowledge regarding them might be crucial in consumer applications. Examples include the study of facial expressions, driver awareness systems in computer animation, the film and advertising business, or helping those with disabilities communicate via an eye-based interface.

A web camera is used by Rishikesh et al. to create a virtual human interaction device that offers a hands-free connection between the humans and computers, particularly for those with physical disabilities.

VahidKazemi and others suggested Face alignment is completed in milliseconds with accuracy that is on par with or better than that of state-of-the-art techniques on common datasets. By identifying the key elements of earlier face alignment algorithms but then combining them in a simplified formulation into a series of high-capacity regression functions trained through gradient boosting, faster approaches were developed than those used before.

Local Binary Pattern (LBP) is a basic texture operator that is considered to be highly efficient. It labels the pixels of an image by thresholding the neighbourhood

of each pixel and then treats the result as a binary integer. We are able to represent the facial photos using a straightforward data vector by utilising the LBP in conjunction with histograms. The first phase in the computational process of the LBPH is the creation of an intermediate picture. This image will describe the initial image in a more accurate manner by focusing on the features of the face. The method does this by using the idea of a sliding window, which is determined by the parameters radius and neighbours.

3. PROPOSED METHODOLOGY

Includes face identification, face tracking, eye detection, and comprehension of an eye sequence in the proposed system. For operating a non-intrusive human-computer interaction, blinks in real-time. Human eye movements take the role of the standard mouse and virtual keyboard technique of computer interaction. Using OpenCV code with pupil detection, this approach will assist the paralysed person in focusing on the eye inside the picture. As a result, the human eye is positioned in the middle (pupil). Physically challenged individuals, particularly those without hands, may calculate quickly and easily by using the centre location of the pupil as a reference and moving the pointer left and right in

accordance with that. The picture is first taken by the camera.

The process starts by using a face detection program to determine a user's face. Once the face has been recognised, the location of the eyes and lips are recorded in order to control virtualkeyboard and mouse functions including choosing keys and moving forward and backward as well as mouse actions like left- and right-clicking and pointer movement. Sensors and specialised

equipment are not required. It is a hands-free solution that is functional for those with various forms of disability. Opening the mouth must be done before using the mouse or virtualkeyboard. Simply turn your head in the direction that a mouse pointer is anticipated to go to left or right. The camera must be squinted in order to start the scrolling mode, and once it is started, it will indicate that the mode is active. You may now look up, climb to the top, and do something similar. Scroll lower, as seen in picture 1.

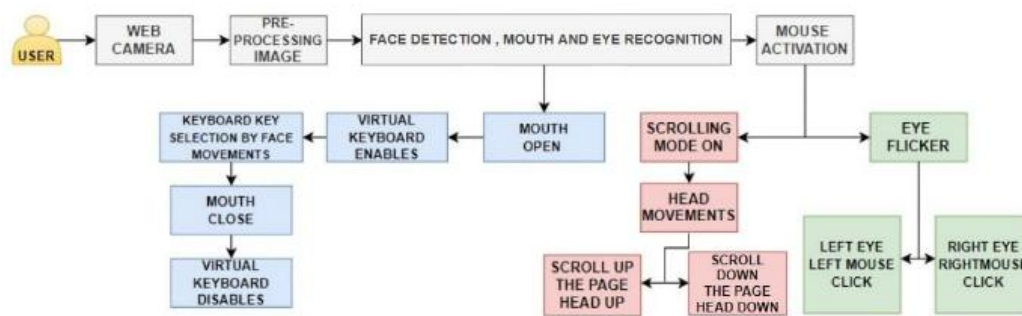


Figure 1: Proposed System Architecture

The image or the video is pre-processed using the camera, then the source is converted into the greyscale image. The greyscale image is converted to a bitmap image from the bitmap image the eye blink detection and the face movement calculation is done, then the mouse event and mouse cursor movement are done.

The following procedure was used to simulate mouse and virtualkeyboard action using mouth and pupil movement:

A. Detection of Face

The user must look at the camera in order to provide a precise and clear picture. The system's preinstalled camera is used to capture the user's picture, which is then processed using Python and OpenCV. An iBUG 300-W which was before set of data

that has 68 variables that are mapped to a person's face is available from Dlib. The facial detector locates and monitors important facial landmarks.

Mouth Detection(MAR):The MAR stands for Mouth-Aspect Ratio is an mouth recognition algorithm, Totally there are 20 x y coordinates.

$$MAR = \frac{(C2 - C8) + (C3 - C7) + (C4 - C6)}{2(C1 - C5)}$$

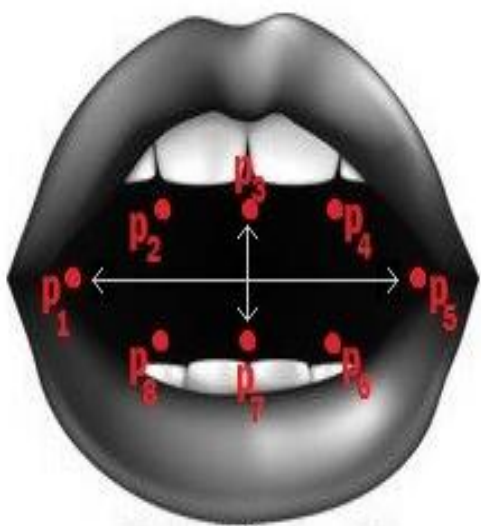


Figure 2: Mouth recognition using MAR algorithm.

Same as eye-aspect-ratio, we have mouth-aspectratio(MAR), which is used to identify whether the mouth is open or not as shown in fig 2.

Haar-cascade algorithm pseudo code:

- Pick f (maximum acceptable false positive rate per layer) and d (minimum acceptable detection rate per layer)
- Lets F_{target} is target overall false positive rate
- Lets P is a set of positive examples
- Lets N is a set of negative examples
- Lets $F_0 = 1$, $D_0 = 1$, and $i = 0$ (F_0 : overall false positive rate at layer 0, D_0 : acceptable detection rate at layer 0, and i : is the current layer)
- While $F_i > F_{target}$ (F_i : overall false positive rate at layer i):
 - $i++$ (layer increasing by 1)
 - $n_i = 0$; $F_i = F_{i-1}$ (n_i : negative example i):
 - While $F_i > f * F_{i-1}$:
 - n_i++ (check a next negative example)
 - Use P and N to train with AdaBoost to make a xml (classifier)
 - Check the result of new classifier for F_i and D_0
 - Decrease threshold for new classifier to adjust detection rate $r >= d * F_{i-1}$
 - $N = \text{empty}$
 - If $F_i > F_{target}$, use the current classifier and false detection to set N

By applying Haar classifier the system identifies the region of the face, eyes, and mouth are detected and extracted for processing. It controls the action of the virtual mouse and virtual keyboard by providing a hands-free interaction between humans and computers. It allows disabled people to scroll up, scroll down, scroll left, scroll right, right-click, left-click, and to perform cursor movement in virtual mouse.

B. Eye Detection

Eye-Aspect-Ratio (EAR) is used in eye detection. As shown in fig. 3, it has been used to determine if the subject's eye is flickering in the video frame.

EAR is an eye recognition algorithm, totally there are 6 x y coordinates when the eye lid is closed the input tends to 0.

$$EAR = (C2 - C6) + (C3 - C5) / 2(C1 - C4)$$

C1-C6 located on the eye

C2-C6 and C3-C5 distance between the eyes in vertical

C1-C4 distance between the eye in horizontal

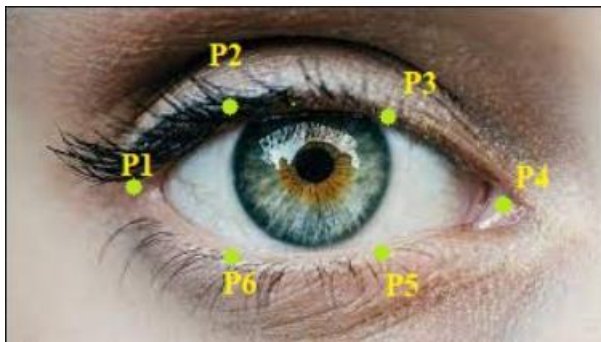


Figure 3: Eye recognition using EAR algorithm.

The minimum EAR value that must be met before the eyes are considered open; otherwise, they are considered closed. It's possible that you'll wish to adjust this setting in accordance with your needs. Determine the value after you have tried to discover EAR in a variety of various contexts. It is important to keep in mind that this EAR does not refer to just one eye but rather represents the total EAR for both eyes together.

The value of EAR is subject to rapid fluctuation. Even if you close your eyes and blink twice, the EAR will decrease significantly. However, blinking does not

always indicate that one is drowsy. Grogginess would be a condition in which a person has closed one or both of their eyes (meaning that their EAR is extremely low) for a period of time that is at least ten video frames long. Therefore, the value of this variable indicates the maximum number of frames in a sequence during which EAR may continue to be lower than MINIMUM EAR without triggering an alarm for sleepiness.

C. Classification of Eye Movement: The support vector machine classifier is used to categorise the different eye motions. Open-eye, closed-eye, left- and right-eyeball motions are among the eye movements. SVM can also be used to categorise the collected data. SVM is a term for a collection of connected supervised learning techniques used in classification and regression. SVM use training files for several classes.

D. Motion Detection: Motion is seen when the identical pixels from the previous frame are deleted from a specific location. The activity at a pixel is recognised when the exact amount of the subtracting is higher than a predetermined threshold.

E. Detection of Blink: The blink detection process can only run when the eye is not moving. The face is used to shift the cursor on the screen instead of the mouse, and then the blink is executed to the click-on eyes.

F. Mouse Operation: The client played out the mouse activities utilizing just facial signals. For starting a left snap, the client can flicker the left eye, also for the right snap, squint the right eye and get the tip of the nose across the screen to move the mouse cursor every which way. This operation is focused heavily on foreseeing the facial tourism attractions of a specific face. With the help of the milestones, we can accomplish a lot. recognising eye flickers in video to anticipating the subject's emotions. The uses, effects, and possible outcomes of face tourist attractions are many and fascinating.

Dlib's prebuilt model, wholly an execution of, not in the least a quick face discovery yet additionally permits us to anticipate 68 2D facial tourist spots precisely. Exceptionally convenient. Using those predicted tourist attractions on the face, we may put together pertinent highlights that will also enable us to distinguish certain actions, such as using the mouth-angle proportions to identify a yawn, the eye-

viewpoint proportional to identify a squint, and so on, or possibly a mope. These actions are changed in this funder takeover and used as mouse cursor triggers. The mouse cursor was managed by the PyAutoGUI module, as shown in image 4.

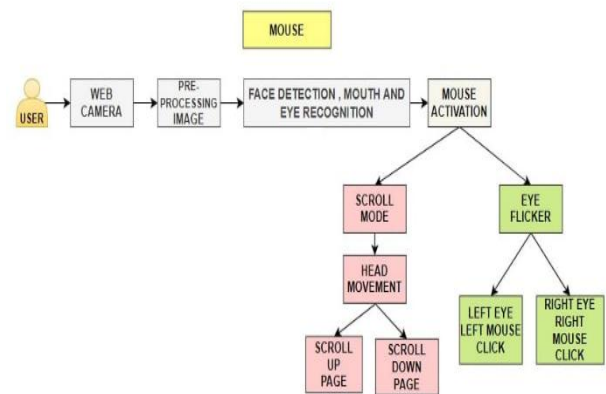


Figure 4: Mouse Architectures

G. Virtual Keyboard Operation: In this task the "Look controlled console with python and OpenCV". This was acknowledged with the main reason to show you how to identify the eyes, their look, and the flickering. A couple of parts in regards to eye discovery have been hardcoded to work explicitly with my eye and with a particular lighting condition and with a specific web camera. Inabilities or physical hindrances brought about by hereditary or horrible injury hold individuals from declaring their fundamental requirements, sentiments, or thoughts. This will present an enabling

framework for physically healthy people with disabilities that is eye gaze controlled. Virtual consoles are suggested in this project as a way for those who are unable to use their four limbs, such as those who are quadriplegic, to express their opinions.

The virtual console that is being suggested contains 40 keys, including an eraser key, English text types, mathematics numbers, and a few Latin symbols. Each key lights up successively in either the forward or backward direction, as illustrated in Fig. 5. The dynamic key is constructed using the client's eye flashing and the client's eye look, which examines the key's initiation. The proposed console includes several impressive features, such the sound that plays when a key is pressed and the erase key that removes characters that were typed mistakenly. The character per minute (CPM), word per minute (WPM), and error rate have been calculated from tests conducted by a small number of customers that provide results that are more advanced than those of the present models. The proposed immersive console is a low-cost assistive system because all that is needed to use it is a webcam.

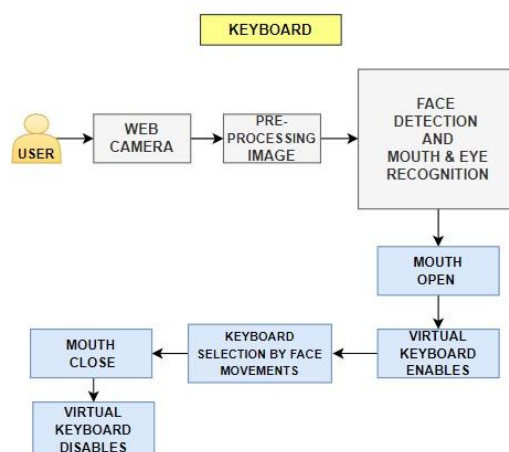


Figure 5: Working of virtualkeyboard

The buttons on the eye controlling virtual keyboards allow us to move the virtual keyboard and remove words and characters. Additionally, a spacebar key and a key to activate shape writing are included. Text suggestions show at the head of the virtual keyboard as you focus on each character. To choose a word, linger over it. Go to Keyboard > Interactions. Choose the virtual keyboard key to use by clicking Warp on key. Check the box next to Perform a click while releasing the key.

3.1 Proposed Algorithm

Step-1: As soon as we start the application the desktop raises a frame window and the camera starts capturing the image and the frame window displays the video capturing that is being captured by the camera.

Step-2: Then Shape Predictor will try to landmark the co-ordinate points on the face and the continuous capturing of the video will depict the changes being made spontaneously.

Step-3: In order to read the input, the application will check for change in mouth aspect ratio, if it is more than the threshold then the frame will start reading the input.

Step-4: In input reading mode, we will have a rectangle around the nose and the nose point movements make the cursor move up, down, left and right.

Step-5: In the reading mode it will also check the change in Eye aspect ratio, if it is below the threshold for either of the eyes then that particular click is made (right-click, left-click).

Step-6: If the Eye aspect ratio of both the eyes is below the threshold (squint) then, it is considered as scrolling mode for scrolling up and down to navigate through documents.

Step-7: If we want to stop the input reading mode, the mouth aspect ratio has to meet the threshold again (this will stop the application frame window).

4. EXPERIMENTAL RESULTS

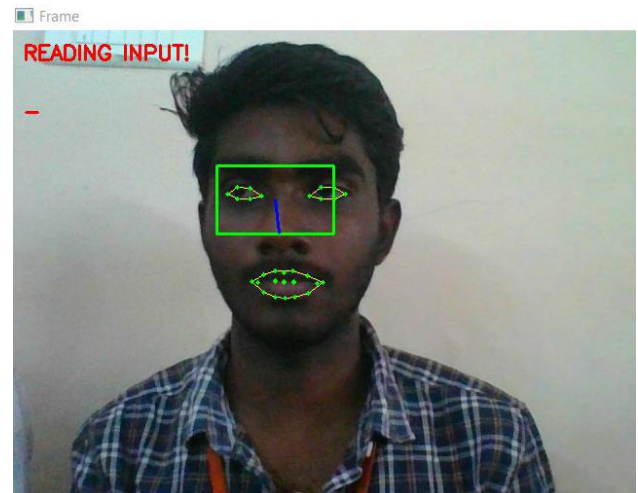


Figure 6: Reading input and Detected by Haar classifier

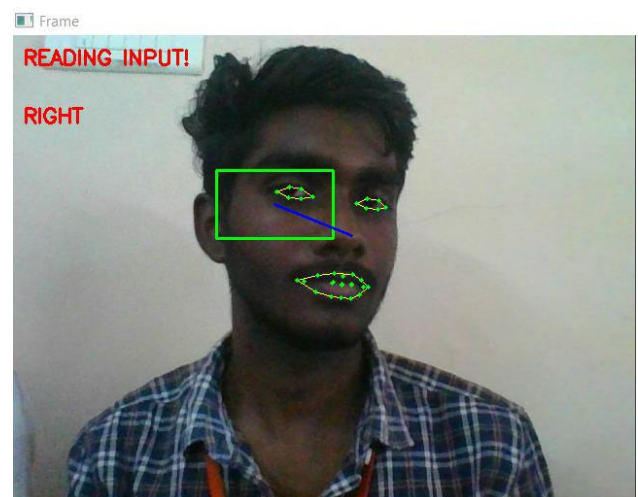


Figure 7: Mouse cursor moves to Right by eye movement

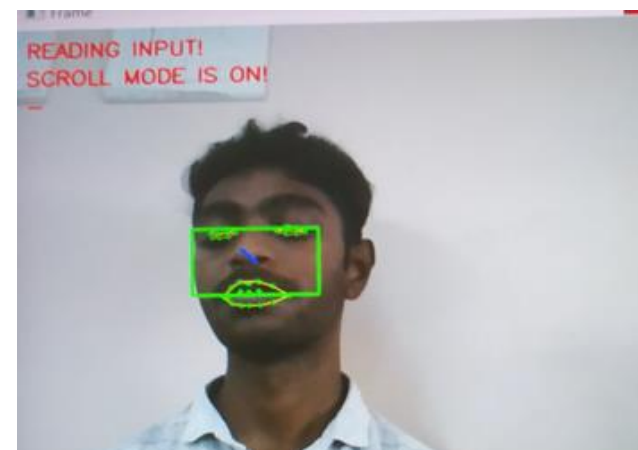


Figure 8: Scroll Mode is activated

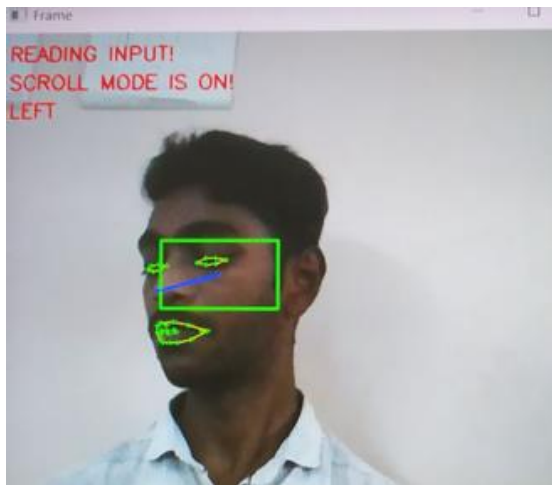


Figure 9: Page is Scrolled to Left

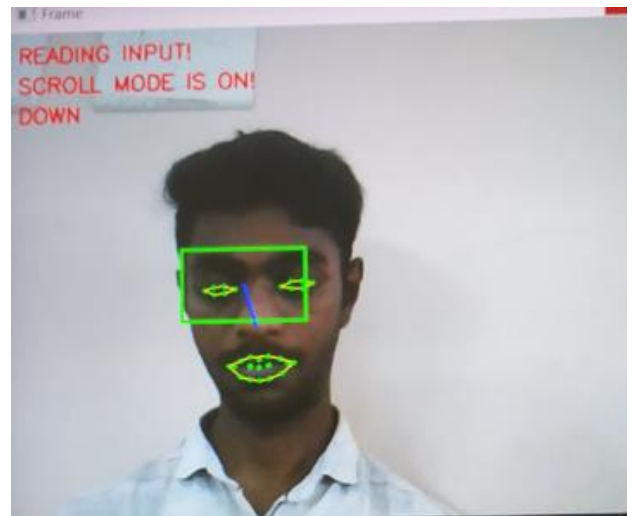


Figure 12: Scroll Mode turned on for downwards

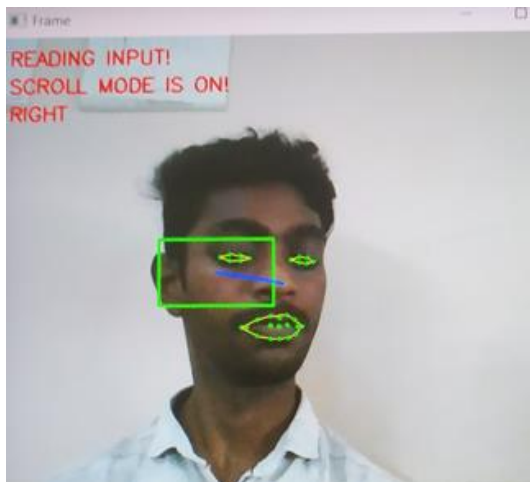


Figure 10: Page is Scrolled to Right



Figure 11: Scroll Mode turned on for upwards

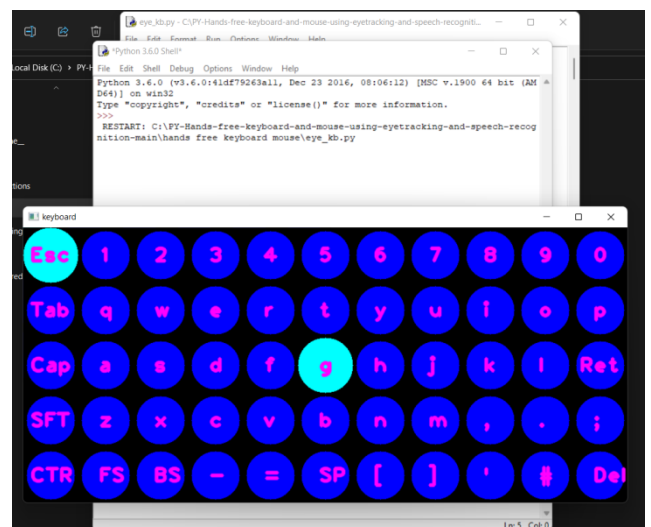


Figure 13: Implementation of virtual keyboard

From the Fig 6 to fig 13 shows the experimental outputs of mouse and virtual keyboard operations.

5.CONCLUSION

It is evident from the procedure used that the computer pointer may be moved by eyeball movement without the need of

hands. People with disabilities will find this useful when utilising the physical components of a system to control the pointer points and virtual keyboard operations. Because one's eyes alone can move the mouse and perform virtualkeyboard activities, without no outside assistance. The system works by facial expressions like movement of eye-ball and mouth movement. By applying Haar classifier the system identifies the region of the face, eyes, and mouth are detected and extracted for processing. It controls the action of the virtual mouse and virtual keyboard by providing a hands-free interaction between humans and computers. To achieve efficient and precise movement, this technology may be expanded to include eye movement and blinking.

REFERENCE

1. Anyway Juhong, NutthananWanluk, Pintavirooj, And SarinpornVisitsattapongse, "Smart Wheelchair Based On Eye Tracking", IEEE transaction, The 2016 Biomedical Engineering International Conference.
2. Balamurugan, P., J. Santhosh, and G. Arulkumaran. "HAND MOTION BASED MOUSE CURSOR CONTROL USING IMAGE PROCESSING TECHNIQUE." *Journal of Critical Reviews* 7.4(2020)
3. C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. 300 Faces in the Wild Challenge: The first facial landmark localization Challenge. *Proceedings of IEEE Int'l Conf. on Computer Vision (ICCV-W)*, 300 Faces in the-Wild Challenge (300-W).
4. Hossain, Zakir, MdMaruf Hossain Shuvo, and PrionjitSarker. "Hardware and software implementation of real-time electrooculogram (EOG) acquisition system to control a computer cursor with eyeball movement." In *2017 4th International Conference on Advances in Electrical Engineering (ICAEE)*, pp. 132-137. IEEE, 2017.
5. IEEE Conference on Computer Vision by M.Pantic, S.Zafeiriou, G. Tzimiropoulos, C.Sagonas,
6. inSydney, Australia, December 2013
7. IoanaBacivarov, MirceaIonita, Peter Corcoran, "Statistical models of appearance for eye-tracking and eye-blink detection and measurement", Vol.54, No.3, August 2009. *IEEE transactions on consumer electronics*, pp.

8. Lee, Jun-Seok, Kyung-hwa Yu, Sang-won Leigh, Jin-Yong Chung, and Sung-Goo Cho. "Method for controlling device based on eyeball motion, and device therefor." U.S. Patent 9,864,429, issued January 9, 2018.
9. Mehta, Sukrit, et al. "Real-Time Driver Drowsiness Detection System Using Eye Aspect Ratio and Eye Closure Ratio." Available at SSRN 3356401(2019).
10. Ramesh R and Rishikesh M, "Eye Ball Movement to Control Computer Screen", Vol 6, Issue 3 February 2015 ISSN: 2155-6210 Journal of Biosensors & Bioelectronics pp.
11. Sharma, N. Jain and P. K. Pal, "Detection of eye closing/opening from EOG and its application in robotic arm control", Biocybern. Biomed. Eng, vol. 40, pp. 1-14, Jan.2020.
12. Swaroop Joshi, Vinay, Vasisht and C. Gururaj. HCI based eye controlled mouse. 3rd International conference on Electronics, Comm and Aerospace Technology (ICECA). IEEE, 2019. Sydney, Australia, December 2013.
13. Tereza Soukupova' and Jan C'ech. Real-Time Eye Blink Detection using Facial Landmarks. In 21st Computer Vision Winter Workshop, February 2016.
14. Vahid Kazemi, Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In CVPR, 2014.
15. Yu-Tzu Lin Ruei-Yan Lin Yu-Chih Lin Greg C Lee "Real-time eye-gaze estimation using a low-resolution webcam", Springer, pp.543-568, Aug (2012).