CSE3013 - ARTIFICIAL INTELLIGENCE

# MOUSE CURSOR & KEYBOARD CONTROL USING FACIAL MOVEMENTS


**TEAM MEMBERS**:

Shingala Tirth Narendrakumar (19BCE0844)

Aryan Srivastava (20BCE0843)

Pallav Jyoti Buragohain (20BCE0866)

Vennela G (20BDS0146)

# ABSTRACT

To create software that will allow the user to perform mouse cursor control with facial movement with just a regular webcam. It will be hands-free, and also cuts down the cost of wearable hardware or sensors for mouse cursor control.

# KEYWORDS

- Mouse control
- Keyboard control
- Facial movements
- Hands-free device
- Face recognition
- Eye blink
- Touchless virtual keyboard
- Introduction

# INTRODUCTION

## Problem statement

People without hands can't use electronic gadgets. Face trackers with eye gaze communication have been proposed as a possible communication portal by researchers in the field. Some portals as such have been introduced, but they are quite expensive. In psychology, marketing, and user interfaces, Face trackers have become increasingly essential. They have been around for a while, but their use was mostly limited to laboratory investigations to examine the nature of human eye movements, and were too expensive to consider using in a genuine user-human interface a decade ago.

## Motivation

The number of assistive gadgets controlled by computers is growing every day. Physically challenged people can utilize these assistive devices to interact with a variety of applications such as communication with others, education, and entertainment. Vision-based systems have become more popular in recent years for creating and implementing various types of applications. The cursor movement in a computer application is controlled by the movement of lit markers using facial expression in this work. Face detection, eye extraction, and voice recognition are among the image processing techniques used. It captures an input image with a standard webcam. A computer would also be simple to use and learn for a physically challenged individual, which is another reason for constructing this project.

## Objectives

In the recent decade, there has been a huge growth in the technology sector. Mobile phones, tablets, and even laptops with Touch screens have been introduced and widely accepted. Eye gaze has been proposed as a possible communication portal by researchers in the field. Some portals as such have been introduced, but they are quite expensive. This system can make the lives of physically challenged individuals significantly easy by giving them access to computers without having to use their limbs. Scope and applications. The system can be used to access a computer without the usage of limbs - from eye movement alone. This can be used by physically challenged individuals. The scope of this paper includes reviewing existing work on the topic, developing and testing a system with a main interface, Mouse/keyboard simulation engine, User action detection module, Halt (sleep) module, mouse and keyboard function modules.

## Techniques Used

The mouse cursor can be controlled by facial movements such as moving up and down or left and right, blinking and/or voice. The keyboard can be controlled via keys displayed on the computer screen which get selected when the visual finger is "virtually" inside the area which corresponds to a particular key. A webcam with low resolution attached to a wearable glass frame can be utilised for the same. The iris centre is located using the images generated from the webcam and the movement of iris is coordinated to the movement of the mouse pointer.

## Literature Review

| Ref. No. | Paper Title | Journal name and year of publication | Work done | Techniques used | Gaps found |
|---|---|---|---|---|---|
| 1 | Eye Gaze controlled virtual keyboard | International Journal of Recent Technology and .Engineering (IJRTE), 2019 | Developed a virtual keyboard that works by detecting eye gaze and eye Blinking using video captured directly from a PC camera. | The system uses an approach that involves the 68 points of the face which is specific and must exist in every face is used | A maximum of 100 words only can be written using the keyboard. Results are not accurate for users wearing glasses. |
| 2 | Real-time eye blink detection | Center for Machine Perception, | Presented a literature review of | Estimated the landmark positions, | A fixed blink duration for all subjects |

| | | | | | |
|---|---|---|---|---|---|
| | using facial landmarks | Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University in Prague, 2016 | driver drowsiness detection based on behavioural measures using machine learning techniques | extracted a single scalar quantity eye aspect ratio and characterized the eye-opening in each frame. Finally, an SVM classifier detects eye blinks as a pattern of EAR values in a short temporal window | was assumed, although everyone's blink lasts differently. |
| 3 | Touchless virtual keyboard controlled by eye blinking and EEG signals | Man-Machine Interactions 5.ICMMI 2017. Advances in Intelligent Systems and Computing,2017 | Developed a virtual keyboard where each key can be selected by three double-eye blinks registered by EMG sensor | An online desktop-based tool was created on which EEG signals are used as support that allows the user to change the input mode of single characters to the mode of predicted word selection. | Obtained result(WPM= 1.11) only partially confirms the calculated typing speed(WPM= 1.23) |
| 4 | A 3D Approach to Facial Landmarks: Detection, Refinement, and Tracking | ICPR 2014, IEEE 2014 | A real-time algorithm for accurate localization of facial landmarks in a single monocular image is proposed. | The algorithm is formulated as an optimization problem, in which the sum of responses of local classifiers is maximized concerning the camera pose. The algorithm simultaneously estimates a head position and orientation and detects the facial landmarks in the image. | None were identified |
| 5 | Eye Tracking Based Control System for Natural Human-Computer Interaction | Hindawi Computational Intelligence and Neuroscience,2017 | Developed an eye-tracking-based control system for user-computer dialogue which combines | A fixation function is defined to calculate the gaze time of the user and determines the mouse coordinates | None were identified |

| | | | | | |
|---|---|---|---|---|---|
| | | | both mouse functions and keyboard functions. | directed by eye movement. The system then performs directly from the viewpoint of the user. | |
| 6 | Real-Time Eye Tracking and Blink Detection with USB Cameras | Boston University Computer Science Technical Report No. 2005-12 | A human-computer interface (HCI) system designed for use by people with severe disabilities are presented. | The algorithm used by the system for detecting and analysing blinks is initialized automatically, dependent only upon the inevitability of the involuntary blinking of the user. | The model does not yield accurate results for people wearing glasses. |
| 7 | Low-cost natural interface based on head movements | Procedia Computer Science, 2015 | Developed a human-computer interface that allows for touchless computer control, avoiding the use of a keyboard, mouse, or touchscreen. | An interface that is based on the Kinect sensor, and allows computer control only by the movement of the head, eyebrows (up/down), and the mouth (opening/closing), in addition to available voice commands. | Voice commands were not covered in the initial application goal |
| 8 | Gesture Control Using Single Camera For PC | Procedia Computer Science, 2015 | Proposed a basic model of integrating Face Recognition method with Hand Tracking where the authenticated user alone can access their system without using a mouse thereby providing access a | The Face Recognition model uses Viola and Jones method for detection of the face and PCA (Principal Component Analysis) for recognition and identification of algorithms. | The application however fails clearly in darkness which is not the case with a real mouse, so the illumination is always provided. |

| | | | | | |
|---|---|---|---|---|---|
| | | | privilege to the user. | | |
| 9 | An empirical evaluation of hands-free computer interaction for users with motor disabilities | Journal of Biomedical Informatics, 2016 | To design and evaluate a universal solution for a hands-free HCI, based on the Emotiv EPOC+ device, which, among other capabilities, also enables controlling the computer with facial expressions and motion sensors. | The proposed solution is based on the Emotiv EPOC+ device. By measuring and analyzing EMG signals, facial expressions can be identified, which, in turn, can be used for the implementation of computer commands. | All participants had no experience in using the EPOC+ device, and everyone received the same level of training in using the device. |
| 10 | Locally Linear Regression for Pose-Invariant Face Recognition | IEEE Transactions on Image Processing, 2007 | Proposed a novel locally linear regression (LLR) method, which generates the virtual frontal view from a given non-frontal face image | Formulated the estimation of the linear mapping as a prediction problem and presented the regression-based solution. To improve the prediction accuracy for coarse alignment, LLR is further proposed. | The pose of the input non-frontal face image is assumed to be known. This implies that one has to use a front-end procedure to estimate the pose of the input image. |
| 11 | Incremental Face Alignment in the Wild | Proceedings of the IEEE conference on computer vision and pattern recognition, 2014 | proposed the possibility to automatically construct robust discriminative person and imaging condition-specific models in the wild that outperform state-of-the-art generic face | The Par-CLR method is the foundation for the proposed incremental face alignment framework, which has an exact incremental solution per level. | The system is not real-time |

| | | | | | |
|---|---|---|---|---|---|
| | | | alignment strategies. | | |
| 12 | A System Identification Approach for Video-based Face Recognition | Proceedings of the 17th International Conference on Pattern Recognition, 2004 | Presented a structured approach to the problem of video-based face recognition that dealt with the problem of recognizing faces when both gallery and probe consists of face videos | A moving face is represented as a linear dynamical system whose appearance changes with time. Subspace angles based distance metrics are used to get the measure of similarity between ARMA models representing moving face sequences. | None identified |
| 13 | Automatic 3D Face Reconstruction from Single Images or Video | 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008 | Presented a fully automated algorithm for re-constructing a textured 3D model of a face from a single photograph or a raw video stream. | The algorithm is based on a combination of Support Vector Machines (SVMs) and a Morphable Model of 3D faces. | Simple ambient illumination is assumed in the images or videos |
| 14 | Video-Based Face Recognition Using Adaptive Hidden Markov Models | IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’03), 2003 | Proposed to use adaptive Hidden Markov Models (HMM) to perform video-based face recognition | During the recognition process, the temporal characteristics of the test video sequence are analysed over time by the HMM corresponding to each subject. The likelihood scores | Observation probabilities of Hidden Markov Models (HMM) are not taken into consideration. |

## **Related Work**

As we can see from the literature survey, Face Recognition is utilized in studying facial characteristics, storing features in a database, using them to discover users. It facilitates in encoding the relationship among the facial features, to discover structure features of face and set of photographs that capture the variants of faces that exist even if lightning situations vary. There are different approaches to recognize a face. But in this work, we extract Eigenvalues from a picture that is transformed to Eigenfaces which consists of Eigen features. The advantage of extracting Eigenvalues is that it offers much more correct outcomes than every other technique. This technique is known as Viola and Jones which is called after them. The advantage of using Viola and Jones is that it no longer requires a complete frontal pose of a face. It makes use of cascading items in place of measuring the ratio of eyes and nose.

## **Gaps identified**

As shown in the literature survey, certain models depicted inaccuracy while dealing with users wearing glasses- this occurred due to the glare caused and while working in darkness. Along with that, certain models had word limitations (up to 100 words only) and took generalized assumptions which may or may not work for all users (fixed blink duration).

Implementation of techniques

● We have used shape_predictor_68 which is already a trained model for facial detection and is available publicly.

● Detecting facial landmarks is a subset of the shape prediction problem. Given an input image (and normally an ROI that specifies the object of interest), a shape predictor attempts to localize key points of interest along with the shape.

● In the context of facial landmarks, our goal is to detect important facial structures on the face using shape prediction methods.

● Detecting facial landmarks is therefore a two-step process:

  ○ Step #1: Localize the face in the image.

  ○ Step #2: Detect the key facial structures on the face ROI.

● Face detection (Step #1) can be achieved in a number of ways.

● We could use OpenCV’s built-in Haar cascades.

● We might apply a pre-trained HOG + Linear SVM object detector specifically for the task of face detection.

● Or we might even use deep learning-based algorithms for face localization.

● In either case, the actual algorithm used to detect the face in the image doesn't matter. Instead, what's important is that through some method we obtain the face bounding box (i.e., the (x, y)-coordinates of the face in the image).

● Step #2: detecting key facial structures in the face region.

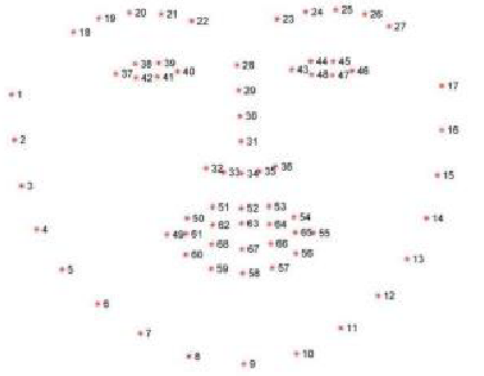
● There are a variety of facial landmark detectors, but all methods essentially try to localize and label the following facial regions: Mouth, Right eyebrow, Left eyebrow, Right eye, Left eye, Nose, Jaw

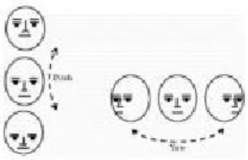
● The facial landmark detector included in the dlib library is an implementation of the One Millisecond Face Alignment with an Ensemble of Regression Trees paper by Kazemi and Sullivan (2014).

● This method starts by using:

● A training set of labeled facial landmarks on an image. These images are manually labeled, specifying specific (x, y)-coordinates of regions surrounding each facial structure.

● Priors, more specifically, the probability of distance between pairs of input pixels



| Action | Function |
|---|---|
|  Opening Mouth | Activate / Deactivate Mouse Control |
|  Right Eye Wink | Right Click |
|  Left Eye Wink | Left Click |
|  Squinting Eyes | Activate / Deactivate Scrolling |
|  Head Movements (Pitch and Yaw) | Scrolling / Cursor Movement |

## Algorithm used:

● Distance from head to eyes is 50%

● Eyes=50% from eyes to nose is 70%

● Nose=70%

● remaining is mouth, Mouth=100;

● Eyes= current_face (1:Eyes);

● Nose= current_face(Eyes:Nose);

● Mouth= = current_face (Nose:Mouth);

Viola-Jones: The characteristics of Viola–Jones algorithm which make it a good detection algorithm are:

● Robust – very high detection rate (true-positive rate) & very low false-positive rate always.

● Real time – For practical applications at least 2 frames per second must be processed.

● Face detection only (not recognition) - The goal is to distinguish faces from non-faces (detection is the first step in the recognition process).

The algorithm has four stages:

● Haar Feature Selection

● Creating an Integral Image

● Adaboost Training

● Cascading Classifiers

The algorithm has linear time complexity o(N) where N is no. of Pixels.

## Working Algorithm:

● First the camera access will be given when we run the program

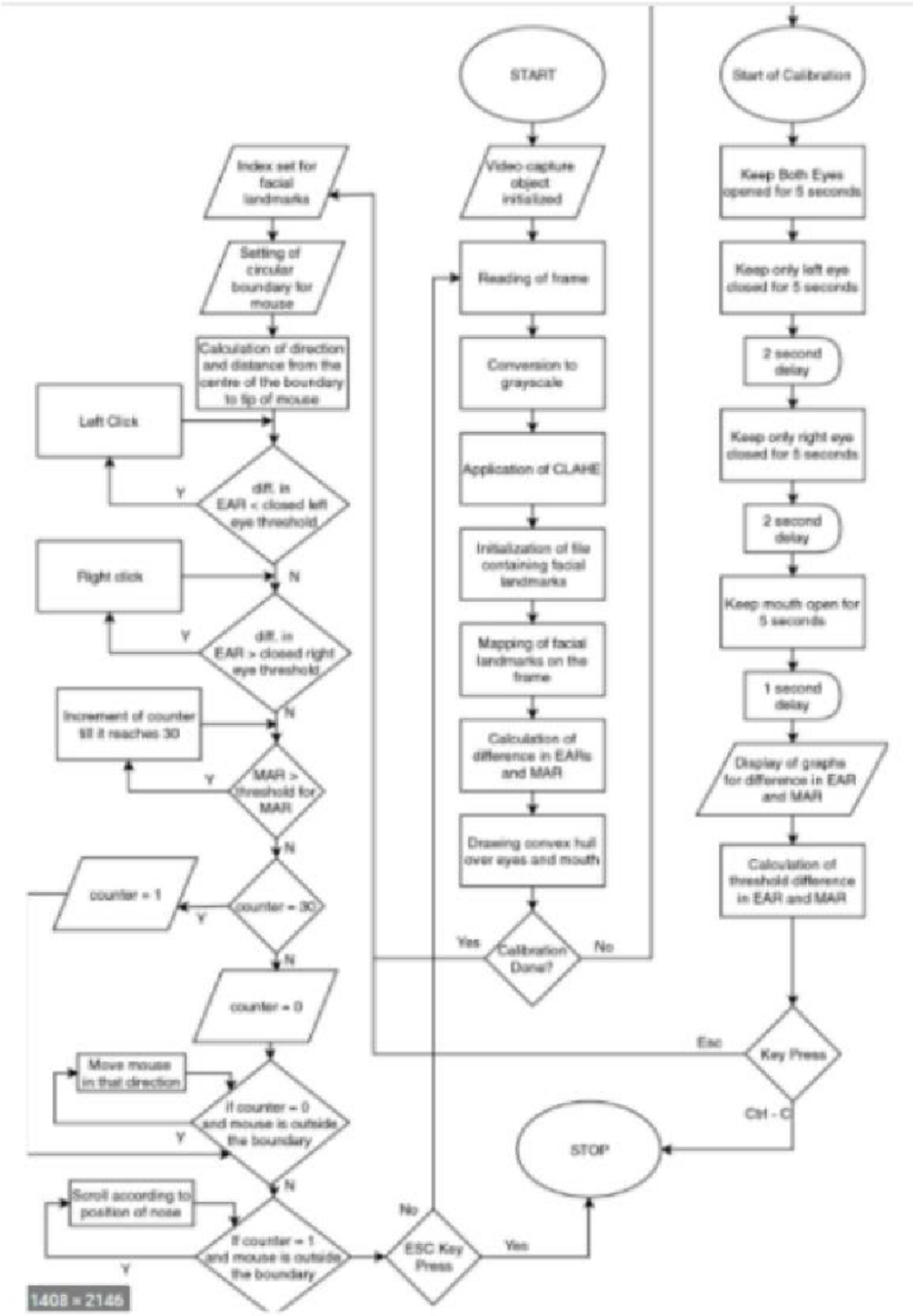
● Second the facial marks will be analysed and then it will be appointed to eyes, mouth, Nose.

● The eyes will have landmarks from 37-48. Nose will have 28-36. mouth 49-68. for our implementation we will be appointing them to the camera.

● Then next we will be taking real time input according to the distance between the coordinates appointed for each part.

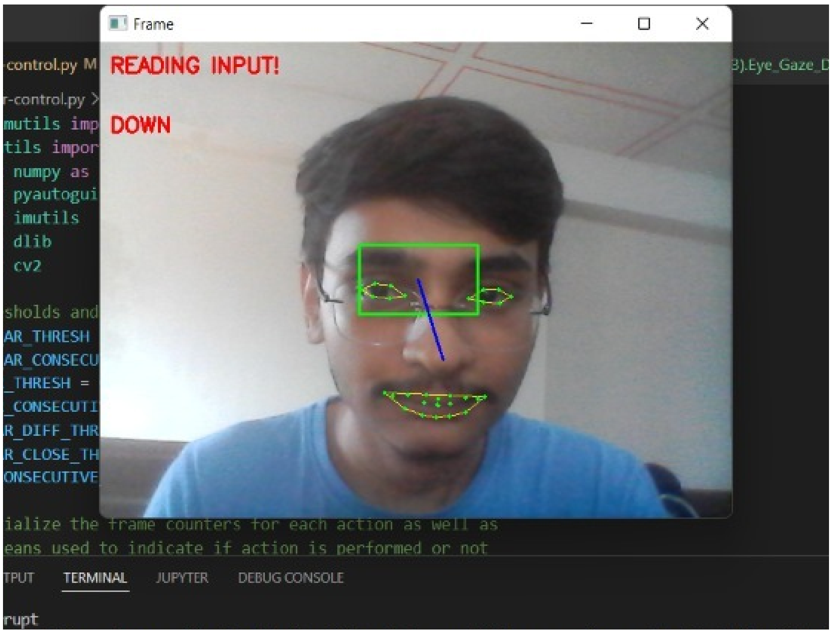
● SO for eyes when the eye aspect ratio is lesser than 0.02 it will be counted as wink. Similarly for each part we have assigned a threshold and according to that all the functionalities will be done.

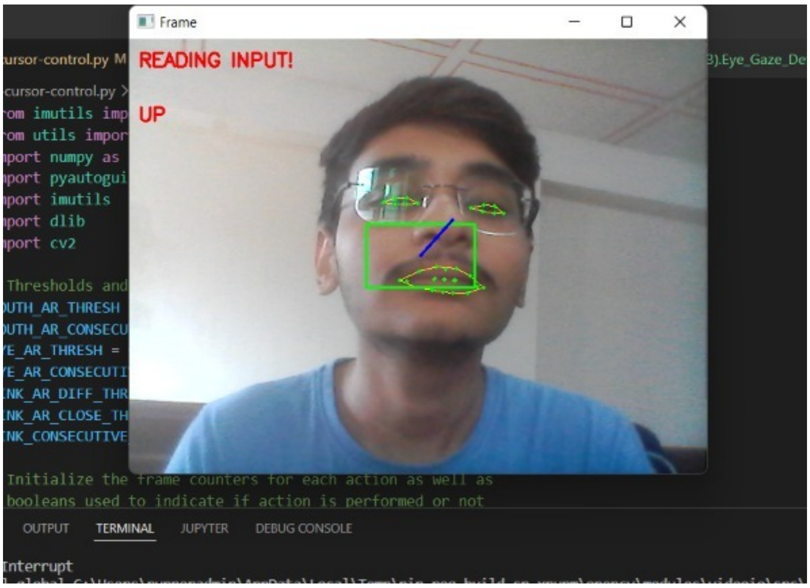
## Architecture diagram

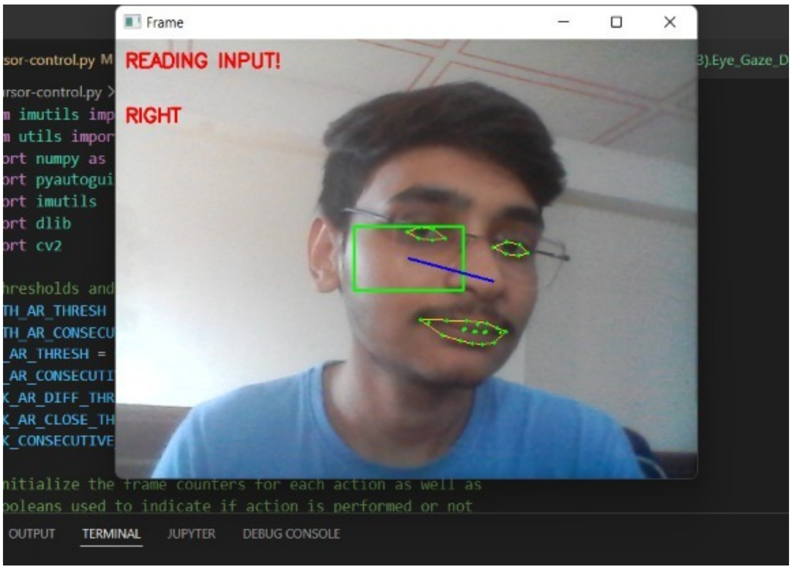


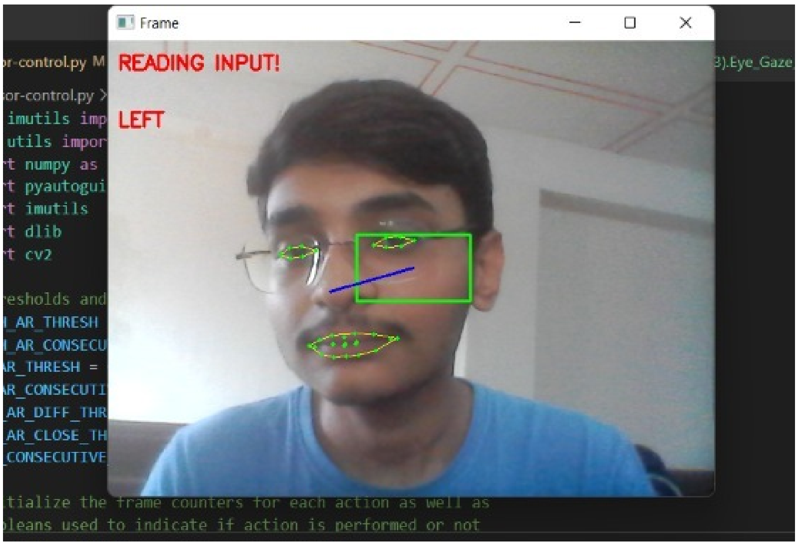
# IMPLEMENTATION

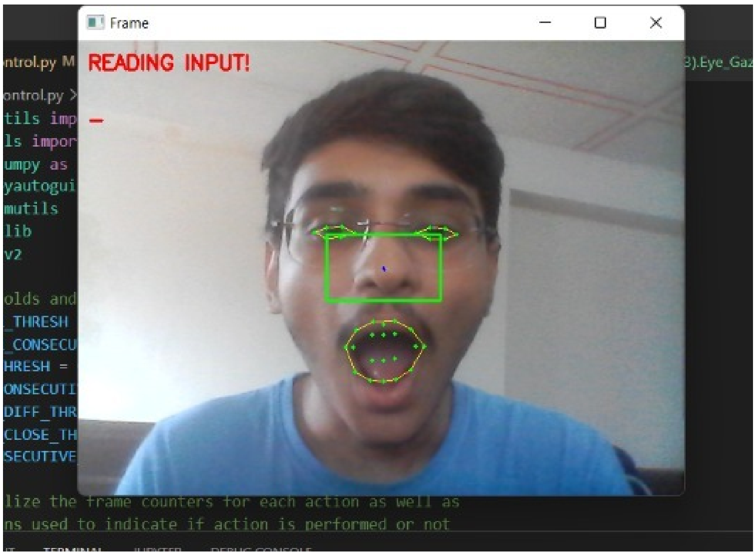
## Results:




Frame
READING INPUT!
RIGHT
OUTPUT
TERMINAL
JUPYTER
DEBUG CONSOLE


Frame
READING INPUT!
LEFT


Frame
READING INPUT!

## Advantages:

● All the actions that are shown work according to their desired function.

● The Project also clearly depicts the face and its facial movements and the functions are performed accordingly.

● The detection of eye gaze and the nose also works with good accuracy.

● The application is user-friendly, and users are given clear instructions, feedback, and system visibility status.

## Disadvantages:

● The accuracy decreases in darkness.

● Users might need some time to get accustomed to the facial controls.

**Link of the code:**
https://drive.google.com/file/d/1v8_GMqjp0DuI2RsJV_2pZPAZm_NSn5Hy/view?usp=sharing

## References:

[1] Cech, J. and Soukupova, T., 2016. Real-time eye blink detection using facial landmarks.

Cent. Mach. Perception, Dep. Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague, pp.1-8.

[2] Chau, M. and Betke, M., 2005. Real time eye tracking and blink detection with usb cameras.

Boston University Computer Science Department.

[3] Zhang, X., Liu, X., Yuan, S.M. and Lin, S.F., 2017. Eye tracking based control system for

natural human-computer interaction. Computational intelligence and neuroscience, 2017.

[4] Chakraborty, P., Roy, D., Zahid, Z.R. and Rahman, S., 2019. Eye gaze controlled virtual

keyboard. Int. J. Rec. Technol. Eng.(IJRTE), 8(4), pp.3264-3269.

[5] Martins, J.M., Rodrigues, J.M. and Martins, J.A., 2015. Low-cost natural interface based on

head movements. Procedia Computer Science, 67, pp.312-321.

[6] Lalithamani, N., 2016. Gesture control using single camera for PC. Procedia Computer

Science, 78, pp.146-152.

[7] Šumak, B., Špindler, M., Debeljak, M., Heričko, M. and Pušnik, M., 2019. An empirical

evaluation of a hands-free computer interaction for users with motor disabilities. Journal of biomedical informatics, 96, p.103249.

[8] Dobosz, K. and Stawski, K., 2017, October. Touchless virtual keyboard controlled by eye blinking and EEG signals. In International Conference on Man–Machine Interactions (pp. 52-61). Springer, Cham.

[9] Chai, X., Shan, S., Chen, X. and Gao, W., 2007. Locally linear regression for pose-invariant face recognition. IEEE Transactions on image processing, 16(7), pp.1716-1725.

[10] Asthana, A., Zafeiriou, S., Cheng, S. and Pantic, M., 2014. Incremental face alignment in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1859-1866).

[11] Aggarwal, G., Chowdhury, A.R. and Chellappa, R., 2004, August. A system identification approach for video-based face recognition. In Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. (Vol. 4, pp. 175-178). IEEE.

[12] Breuer, P., Kim, K.I., Kienzle, W., Scholkopf, B. and Blanz, V., 2008, September. Automatic 3D face reconstruction from single images or video. In 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition (pp. 1-8). IEEE.

[13] Cech, J., Franc, V. and Matas, J., 2014, August. A 3D approach to facial landmarks: Detection, refinement, and tracking. In 2014 22nd International Conference on Pattern Recognition (pp. 2173-2178). IEEE.

[14] Liu, X. and Cheng, T., 2003, June. Video-based face recognition using adaptive hidden markov models. In 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. (Vol. 1, pp. I-I). IEEE.

[15] Li, S., Liu, X., Chai, X., Zhang, H., Lao, S. and Shan, S., 2012, October. Morphable displacement field based image matching for face recognition across pose. In European conference on computer vision (pp. 102-115). Springer, Berlin, Heidelberg.

[16] Aggarwal, G., Chowdhury, A.R. and Chellappa, R., 2004, August. A system identification approach for video-based face recognition. In Proceedings of the 17th International Conference

on Pattern Recognition, 2004. ICPR 2004. (Vol. 4, pp. 175-178). IEEE.

[17] Pentland, A., Moghaddam, B. and Starner, T., 1994. View-based and modular eigenspaces for face recognition.

[18] Cai, Q., Sankaranarayanan, A., Zhang, Q., Zhang, Z. and Liu, Z., 2010, June. Real time head pose tracking from multiple cameras with a generic model. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops (pp. 25-32). IEEE.

[19] Bergasa, L.M., Nuevo, J., Sotelo, M.A., Barea, R. and Lopez, M.E., 2006. Real-time system for monitoring driver vigilance. IEEE Transactions on Intelligent Transportation Systems, 7(1), pp.63-77.

[20] Lee, H.S. and Kim, D., 2006. Generating frontal view face image for pose invariant face recognition. Pattern recognition letters, 27(7), pp.747-754.