iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52854

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Eye/Facial Gesture System Controller Based on AI/ML

Karamjeet Kaur[1], Nandini[2], Swati Raj[3], Shrishty Shakya[4], Shiv Pratap Singh Sengar[5]
[1]*School of Engineering & Technology, Sharda University, Greater Noida, U.P., 201306, India*
[1, 2, 3, 4, 5]*MGM COET, Noida, U.P., 201301, India*

Abstract: ***This paper presents a ground-breaking real-time algorithm for accurately detecting eye blinks in video sequences captured using standard cameras. The algorithm utilizes advanced landmark detectors trained on diverse datasets to ensure robustness against variations in head orientation, illumination, and facial expressions. By employing precise landmark detection techniques, the algorithm estimates the level of eye opening using the eye aspect ratio (EAR) in each frame. An SVM classifier analyses the patterns of EAR values within a short temporal window to identify eye blinks. Comparative evaluations on popular datasets demonstrate the superior performance of the proposed algorithm compared to existing methods. Additionally, the research explores the utilization of eye movements for controlling computer programs, evaluating four distinct approaches in a user-friendly photo viewer. The evaluation process considers various factors such as component sizes, execution time, unintended selections, and gesture repetitions. User experiments reveal that component sizes of 200px provide a convenient and efficient means of application control. The gaze-based method and gestures based on joining points receive positive feedback and exhibit satisfactory performance. This paper contributes significantly to the field of eye-based interaction by introducing an innovative blink detection algorithm and conducting a comprehensive evaluation of eye movement approaches for application control.***

Keywords: Eye / Facial gesture system controller based on AI / ML, Eye controller, Facial landmarks, Eye tracking system, Eye mouse cursor

I. INTRODUCTION

The ability to detect eye blinks is crucial in various applications, such as driver drowsiness monitoring, preventing eye-related syndromes in computer users, aiding disabled individuals in communication, and enhancing anti-spoofing measures in face recognition systems. Existing methods for blink detection can be categorized as active or passive, with active methods relying on specialized hardware and passive methods utilizing standard remote cameras. However, these approaches often have limitations and sensitivity issues. In contrast, recent advancements in robust facial landmark detection offer a promising solution for reliable and real-time blink detection. Eye blink detection technology has diverse applications, spanning from assisting disabled individuals in interacting with computers to drowsiness detection and cognitive load assessment. Numerous approaches have been proposed, including computer vision techniques utilizing facial expression analysis and classifiers like Viola and Jones algorithms, SVM, and Ad boost. Factors like head movements, eye detection, emotion recognition, and blink counts are considered for driver drowsiness detection. Real-time facial landmark detectors capture eye-related features, and the Eye Aspect Ratio (EAR) is widely used as a blink detection metric. Challenges such as lighting conditions and individual variations have led to the development of new techniques incorporating facial landmark detection for improved accuracy. These advancements hold promise for enhancing eye blink detection systems across various practical scenarios. This paper presents a simple yet efficient algorithm that leverages state-of-the art landmark detectors to accurately detect eye blinks. The proposed approach demonstrates superior performance, surpassing existing methods, and contributes to the evaluation of landmark detectors and the development of a novel real-time blink detection algorithm.

II. LITERATURE REVIEW

In [1], M. Betke, J. Gips and P. Fleming (2002). Proposed "The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities."

In [2], C. Morimoto and M. Mimica (2005). Suggested an "Eye gaze tracking techniques for interactive applications Computer Vision and Image Understanding 98".

©IJRASET: All Rights are Reserved

In [3], Kazemi & Sullivan (2014) Kazemi V, Sullivan J. Provided a "One millisecond face alignment with an ensemble of regression trees".
In [4], J. Cech, V. Franc, and J. Matas (2014). Provided "A 3D approach to facial landmarks: Detection, refinement, and tracking".
In [5], Z. Xuebai, L. Xiaolong, Y. Shyan-Ming, L. Shu-Fan (2017), Proposed an "Eye tracking based control system for natural human-computer interaction".
In [6], Chen et al. (2019) Chen D, Chen Q, Wu J, Yu X, Jia T. Suggested a "Face swapping: realistic image synthesis based on facial landmarks alignment".
In [7], Navastara, Putra & Fatichah (2020) Navastara DA, Putra WYM, Fatichah C. Proposed a "Drowsiness detection based on facial landmark and uniform local binary pattern".

III. THE STATE OF ART

A. Interaction by Eye Blinking

The alternative methods for controlling applications using eye movements, specifically focusing on blinking, or closing a single eye. These movements, typically involuntary, can be utilized in a controlled and intentional manner. Examples include using eye blinks to confirm choices, operating virtual on-screen keyboards, initiating mouse actions based on closed eyes, and distinguishing voluntary blinks from involuntary ones. Methods such as setting a duration threshold for eye closure (200ms) help differentiate intentional actions from unconscious blinks. These approaches offer potential for enhanced interaction and control in various systems and interfaces.

B. Interaction Based on Eye Movements

An alternative form of interaction involves executing specific operations based on correctly repeating predefined eye movement sequences. This approach has been used for tasks like switching between control modes in computer games. Research indicates that these gesture-based interactions, while not overly complex, can prevent inadvertent object selection and potentially be faster than gaze-based methods. However, users must memorize the sequences associated with each action. Although promising, these eye-controlled systems require further refinement and testing, motivating ongoing research to expand the range of viable solutions.

IV. METHODOLOGY

The initial step involved using a face detection algorithm to locate the face in an image captured by a webcam. Then, the eyes were detected, focusing on tracking the movement of a single eye for faster processing. The iris movement was specifically tracked by exploiting its lower intensity compared to the rest of the eye. By measuring the shift of the iris as the person changed their gaze, the cursor location on a graphical user interface (GUI) was mapped accordingly. The cursor movement on the graphical user interface (GUI) will be controlled by both head poses, specifically right and left movements, as well as eye movements. Eye blinks will be utilized to perform left and right clicks on the GUI. Additionally, the openness of the mouth will be used to activate and deactivate the eye mouse control feature. This integrated system allows for a comprehensive control mechanism where various facial expressions and movements are utilized to interact with the GUI.

A. Eye Blink Detection with facial Landmarks

Eye blinking is a rapid and involuntary process involving the closure and reopening of the eyelids. It is controlled by multiple muscles, with the orbicularis oculi and levator palpebrae superioris playing crucial roles in regulating eye closure and opening. Blinking serves important functions such as moisturizing the eyes' corners and cleansing the cornea from dust and debris. It also helps spread tears across the eyeball's surface, particularly the cornea, and acts as a reflex to protect the eyes from foreign objects. Facial landmark identification aims to locate and track significant facial landmarks, enabling applications such as face alignment, head pose estimation, face swapping, and blink detection.

Kim et al. (2020) utilized semantic segmentation to accurately extract facial landmarks, emphasizing the even distribution of pixels based on landmark positions for improved classification performance. Utaminingrum et al. (2021) proposed a segmentation and probability calculation approach using facial landmarks to detect initial eye movement positions. Blinking eyes can be detected by analysing the difference between horizontal and vertical lines in the eye area. Another study by Navastara, Putra & Fatichah (2020) focused on eye features extracted using Uniform Local Binary Pattern (ULBP) and the Eye Aspect Ratio (EAR).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

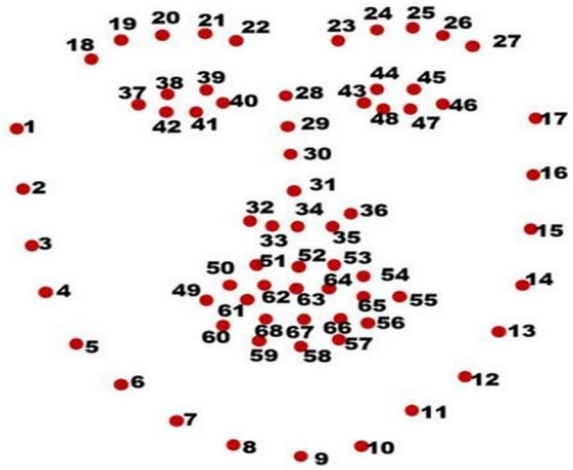


Figure 4.1: 68 Facial Landmarks

In our research, we employ Dlib's 68 Facial Landmark model (Kazemi & Sullivan, 2014). The pre-trained facial landmark detector in the Dlib library estimates 68 (x, y)-coordinates corresponding to facial structures. These coordinates are categorized into different regions, including jaw points, brow points, nose points, eye points, mouth points, and lip points. The process of facial landmark identification involves two steps: face detection, which locates a human face and returns its rectangular position, and face landmark estimation, where the facial landmarks are identified within the detected face region. The Dlib face landmark predictor is trained on the 68-point iBUG 300-W dataset or other selected datasets. The Dlib framework facilitates training form predictors based on input training data.

B. Eye Aspect Ratio (EAR)

In order to detect eye blinks in each video frame, the algorithm first detects the landmarks of the eyes. The eye aspect ratio (EAR) is then computed based on the height and width of the eye. The EAR is a reliable indicator of eye openness, remaining relatively constant when the eye is open and approaching zero as the eye closes. It is robust to variations in person and head pose, as well as uniform scaling and in-plane rotation of the face. Since both eyes blink simultaneously, the EAR values of both eyes are averaged. This approach provides a consistent measure of eye opening.

The targeted area of the face is identified by applying the index criteria. The point index for the two eyes is as follows: (1) Left eye: (37, 38, 39, 40, 41, 42), (2) Right eye: (43, 44, 45, 46, 47, 48). Once the eye region is extracted, it undergoes processing to detect eye blinks. The discovery of the eye region occurs at the initial stage of the system.

The eye landmarks are detected for every video frame. The eye aspect ratio (EAR) between height and width of the eye is computed.

$$EAR = \frac{||p2 - p6|| + ||p3 - p5||}{2\,||p1 - p4||}$$

where p1... p6 are the 2D landmark locations.

C. Mouth Aspect Ratio (MAR)

In order to detect whether a person is yawning or not in each video frame, the algorithm first detects the landmarks of the mouth. The mouth aspect ratio (MAR) is then computed based on the height and width of the mouth. The MAR is a reliable indicator of mouth openness, remaining relatively constant when the mouth is open and approaching zero as the mouth closes. This approach provides a consistent measure of mouth opening by using facial points.

Using the facial points, we can identify that points 49 to 68 represent the coordinates of the mouth. However, we will only utilize eight points, specifically points 61 to 68. The formula for calculating the Mouth Aspect Ratio (MAR) is as follows.

The mouth landmarks are detected for every video frame. The mouth aspect ratio (MAR) between height and width of the mouth is computed.

$$MAR = \frac{||p2 - p8|| + ||p3 - p7|| + ||p4 - p6||}{2\,||p1 - p5||}$$

Where p1….p8 are 2D landmark locations.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

D. Eye detection Flow

The eye blink detection process begins by dividing the video into frames. Using Dlib and the Facial landmarks feature, the face is detected, leveraging a classic Histogram of Oriented Gradients (HOG) feature and linear classifier. Dlib's facial landmarks detector identifies specific facial features like eyes, ears, and nose. The eye area is then determined by utilizing the facial landmarks dataset, which consists of 68 landmarks on the face. The index criteria help identify the targeted area for the eyes. Once the eye region is extracted, the blink detection process takes place. Two lines, drawn horizontally and vertically, are used to detect blinks. A blink is identified when the eyelids temporarily close, and the eyeball is not visible. The upper and lower eyelids must be connected for a blink to occur. By comparing the vertical line's size between open and closed eyes, blinks are detected. The detection is based on a threshold value derived from Modified EAR equations. If the Eye Aspect Ratio (EAR) remains below the Modified EAR Threshold for 3 seconds, an eye blink is considered. Multiple threshold values, including 0.2, 0.3, and modified EAR thresholds, are implemented in the experiment using different video datasets The assumption that a low Eye Aspect Ratio (EAR) always indicates blinking is not entirely accurate. There are instances when a low EAR value can occur due to intentional eye closure, facial expressions like yawning, or even random variations in facial landmarks. To overcome this limitation, we propose a classifier that considers a larger temporal window of frames as input. Through our experiments, we discovered that considering a range of ±6 frames significantly improve blink detection when the eye is most closed. Hence, for each frame, we collect a 13-dimensional feature by combining the EAR values of its ±6 neighbouring frames.

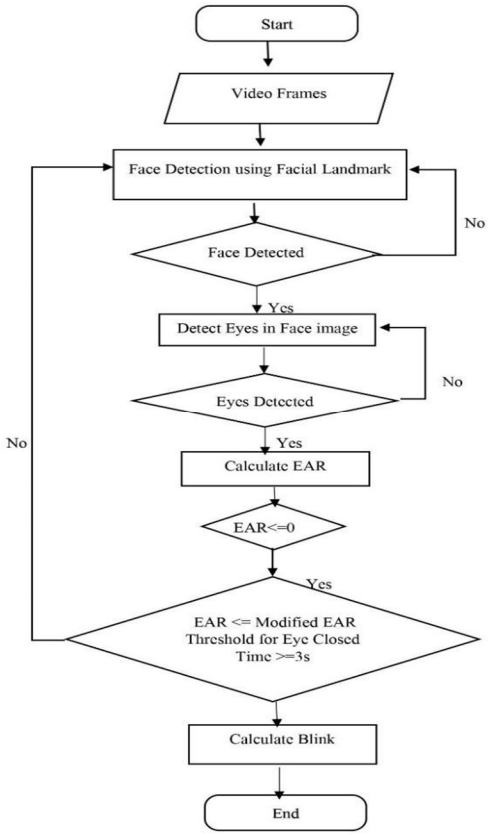


Figure 4.4: Eye Detection Flow Chart

E. Classification

SVM, or Support Vector Machine, is a powerful Machine Learning technique widely used in various domains, including face recognition. It excels in handling complex datasets by extracting intrinsic properties and is particularly effective in computer vision, biometrics, and speech recognition applications. In facial recognition, SVM outperforms other methods in terms of accuracy and robustness, thanks to its ability to handle small datasets, complex patterns, and noisy data. Its efficiency and adaptability make it suitable for real-time applications, such as security systems or image databases, where it can process large amounts of data quickly and adjust parameters effortlessly.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The interpretation that a low value of the Eye Aspect Ratio (EAR) indicates blinking is not always accurate. A low EAR value can occur when a person intentionally closes their eyes for an extended period or engages in facial expressions such as yawning. It can also capture short random fluctuations of the facial landmarks. To address this, we propose a classifier that considers a larger temporal window of frames as input. Through experimentation, we determined that a ±6 frame range has a significant impact on detecting blinks when an eye is most closed. Thus, for each frame, we gather a 13-dimensional feature by concatenating the EAR values of its ±6 neighbouring frames.

We implement a linear SVM classifier, named EAR SVM, which is trained using manually annotated sequences. Positive examples consist of ground-truth blinks, while negatives are sampled from non-blink sections of the videos with a spacing of 5 frames and a margin of 7 frames from the ground-truth blinks. During testing, the classifier scans the frames in a windowed manner. For each frame, except the beginning and ending of a video sequence, a 13-dimensional feature is computed and classified by the EAR SVM.

This approach enhances the accuracy of blink detection by considering a larger temporal context, reducing the likelihood of false positives.

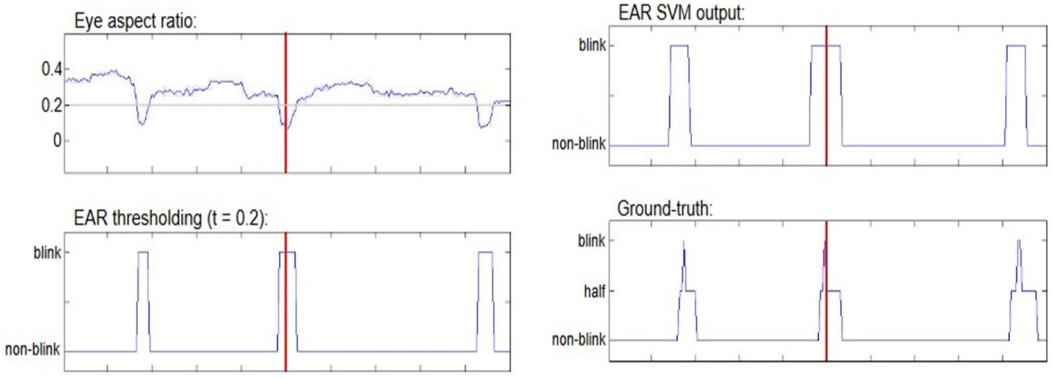


Figure 4.5: The Accuracy of Blink Detection

V. RESULTS

In this research, the objective is to develop an affordable eye-tracking system that enables users to control the mouse cursor on a computer. The system is designed to be cost-effective and efficient, utilizing only a camera and software modules written in the Python programming language. Additionally, if necessary, the system can display the spatial field of view history, providing insights into the user's eye movements and areas of prolonged attention. This information can be utilized to refine the interface or gather spatial attention data for future applications. It is worth noting that the system is capable of operating in various environmental conditions, requiring only minor adjustments to brightness and contrast settings to ensure its reliability. Overall, the achievement of creating such a low-cost eye-tracking system is commendable.

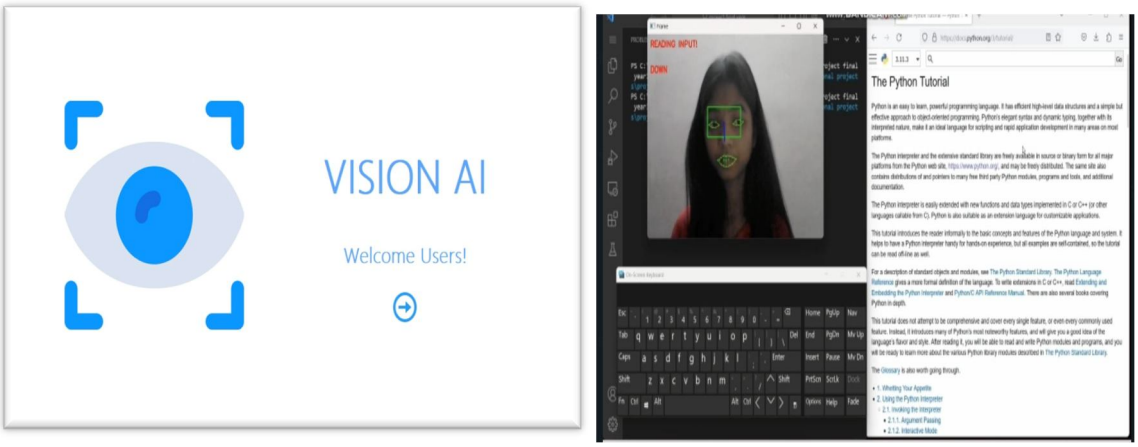


Figure 5: Vision AI software and resulting output

VI. CONCLUSION AND FUTURE SCOPE

In this research, a low-cost eye-tracking system was developed to enable users to control the computer mouse cursor. The system utilizes a camera and software modules coded in Python, offering efficiency and affordability. Additionally, the system has the capability to display eye movement history, providing valuable data for interface improvement and spatial attention analysis. Remarkably, the system demonstrates robustness in various environmental conditions, requiring only minor adjustments to brightness and contrast settings.

In this study, a real-time eye blink detection algorithm was presented and evaluated. The research demonstrated that regression-based facial landmark detectors offer precise estimations of eye openness, even in challenging conditions such as low image quality and real-world scenarios. The algorithm achieved state-of-the-art performance on standard datasets by employing a robust landmark detector followed by a simple eye blink detection based on Support Vector Machines (SVM). The computational overhead of the eye blink detection process was found to be negligible, allowing the algorithm to run in real-time alongside the landmark detectors.

The proposed SVM method, utilizing a temporal window of the eye aspect ratio (EAR), outperformed traditional EAR thresholding techniques. However, the thresholding approach still proved useful as a single image classifier for eye state detection when longer sequences of data were not available.

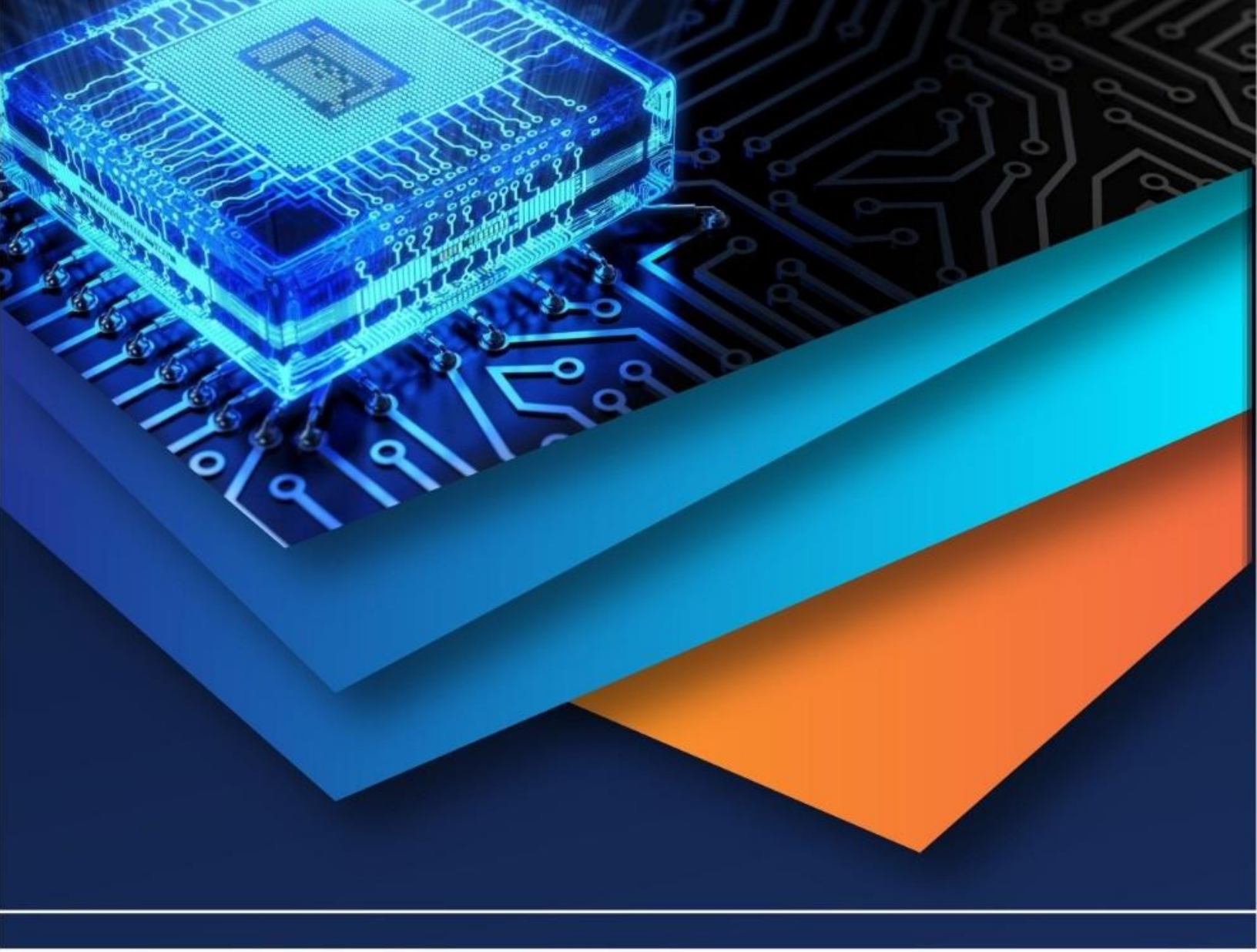
Two limitations of the study were identified. Firstly, a fixed blink duration was assumed for all subjects, despite variations among individuals. An adaptive approach could enhance the results by accounting for these differences. Secondly, the estimation of eye-opening using EAR, derived from 2D images, may lose discriminability for out-of-plane rotations. Introducing a 3D definition of EAR, utilizing landmark detectors that estimate a 3D pose of facial landmarks, could potentially address this limitation.

The research aims to develop a hands-free computing system by controlling the computer using eye movements. It explores movement-based human-computer interaction techniques, specifically using the Viola-Jones algorithm to track eye movements for operating the mouse cursor and performing clicks. The paper highlights the significant future potential of this project, suggesting applications such as driving cars using eye movements and operating digital appliances through body movements.

REFERENCES

[1] M. Betke, J. Gips and P. Fleming. "The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities." IEEE Transactions on Neural Systems and Rehabilitation Engineering, 10:1, pages 1-10, March 2002.
[2] C. Morimoto and M. Mimica. "Eye gaze tracking techniques for interactive applications." *Computer Vision and Image Understanding 98"*, pages 4-24, 2005.
[3] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2879–2886. IEEE, 2012.
[4] Kazemi & Sullivan (2014) Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. Proceedings of the IEEE computer society conference on computer vision and pattern recognition; Piscataway. 2014.[CrossRef] [Google Scholar].
[5] J. Cech, V. Franc, and J. Matas. A 3D approach to facial landmarks: Detection, refinement, and tracking. In Proc. International Conference on Pattern Recognition, 2014.
[6] Z. Xuebai, L. Xiaolong, Y. Shyan-Ming, L. Shu-Fan, Eye tracking based control system for natural human-computer interaction, Computational Intelligence and Neuroscience 2017 (2017) .
[7] Chen et al. (2019) Chen D, Chen Q, Wu J, Yu X, Jia T. Face swapping: realistic image synthesis based on facial landmarks alignment. Mathematical Problems in Engineering. [CrossRef] [Google Scholar].
[8] Kim et al. (2020) Kim H, Kim H, Rew J, Hwang E. FLSNet: robust facial landmark semantic segmentation. IEEE Access. [CrossRef] [Google Scholar].
[9] Navastara, Putra & Fatichah (2020) Navastara DA, Putra WYM, Fatichah C. Drowsiness detection based on facial landmark and uniform local binary pattern. Journal of Physics. [CrossRef] [Google Scholar].
[10] Utaminingrum et al. (2021) Utaminingrum F, Purwanto AD, Masruri MRR, Ogata K, Somawirata IK. Eye movement and blink detection for selecting menu on-screen display using probability analysis based on facial landmark. International Journal of Innovative Computing, Information and Control. [CrossRef] [Google Scholar].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)