

Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer Using Machine Learning

Akashdip Mondal, Biplab Bhowmik, Bikram Debnath, Rakhi Dey, Smriti Dey

Department of Computer Science & Engineering, JIS College of Engineering, Kalyani, West Bengal, India

Abstract:

The rapid advancement of Human–Computer Interaction (HCI) technologies has significantly transformed the way users interact with digital systems. Traditional input devices such as the mouse and keyboard remain effective but present limitations in accessibility, hygiene, and flexibility, particularly for users with physical disabilities or in sterile environments. This research presents a Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer using Machine Learning and Computer Vision techniques as a touchless alternative to conventional input devices. The proposed system integrates hand gesture recognition and eye blink detection to perform real-time mouse operations without physical contact. A standard webcam is used to capture live video input, which is processed using OpenCV and MediaPipe frameworks. Hand landmark detection tracks fingertip movement for dynamic cursor navigation, while facial landmark analysis identifies intentional eye blinks to trigger mouse click actions. By combining these two interaction modalities, the system reduces false gesture activation, improves control accuracy, and minimizes user fatigue compared to gesture-only approaches. The implementation utilizes Python, OpenCV, MediaPipe, NumPy, and PyAutoGUI, making the framework cost-effective and accessible without specialized hardware. Experimental evaluation demonstrates gesture recognition accuracy between 92% and 95%, eye blink detection accuracy around 90%, and average system latency below 150 milliseconds under moderate lighting conditions. The system provides smooth cursor control, reliable click detection, and a short user adaptation period. The proposed framework offers significant potential in assistive computing, healthcare, education, and touchless industrial applications. Although performance may be affected by poor lighting, background clutter, and prolonged usage fatigue, the results demonstrate that multimodal vision-based virtual mouse systems can serve as practical and scalable next-generation human–computer interaction solutions.

Keywords: Human–Computer Interaction, Virtual Mouse, Computer Vision, MediaPipe, Eye Tracking, Gesture Recognition, Touchless Interface

I. INTRODUCTION

Human–Computer Interaction (HCI) has undergone significant transformation over the past few decades, evolving from command-line interfaces to graphical user interfaces and now towards more natural and intuitive interaction paradigms. Traditional input devices such as the keyboard and mouse have been the cornerstone of computer interaction since the early days of personal computing. While these devices offer precision and reliability, they pose limitations in terms of accessibility, hygiene, and naturalness of interaction, especially for individuals with physical disabilities or in environments where touch-based interaction is impractical.

With the rapid advancement of computer vision, machine learning, and artificial intelligence, researchers have increasingly focused on developing alternative interaction mechanisms that mimic natural human behavior. Gesture-based interaction and eye-tracking technologies have emerged as promising solutions to overcome the constraints of conventional input devices. These approaches allow users to interact with computing systems using hand gestures, facial movements, and eye blinks, enabling a touchless and intuitive mode of control [1].

Gesture recognition systems leverage visual input captured through cameras to interpret human hand movements and translate them into corresponding system commands. Similarly, eye-tracking systems analyze eye movements and blinking patterns to infer user intent. When combined, these technologies offer a powerful 9 multimodal interaction framework capable of replacing or supplementing traditional mouse operations such as cursor movement, clicking, dragging, and scrolling [2].

The motivation behind gesture- and eye-controlled virtual mouse systems is strongly tied to accessibility. Individuals suffering from motor impairments, repetitive strain

injuries, or temporary physical limitations often struggle to use standard input devices efficiently. Assistive technologies based on vision-based interaction can significantly enhance their ability to access digital resources and improve overall quality of life [3].

Moreover, in sterile environments such as hospitals, laboratories, and clean rooms, touchless interaction reduces the risk of contamination and enhances operational safety.

Recent developments in low-cost imaging hardware and open-source libraries such as OpenCV, MediaPipe, and Python-based automation tools have made it feasible to implement real-time gesture recognition systems using standard webcams. This democratization of technology has enabled researchers and students to develop robust interaction systems without the need for specialized hardware [4].

This project focuses on the design and implementation of a Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer using computer vision and machine learning techniques. The system aims to track hand gestures for cursor movement and detect eye blinks or gaze patterns for mouse click actions. By integrating these functionalities, the proposed system offers a cost-effective, touchless, and userfriendly alternative to conventional mouse devices.



Fig. 1. Transition from traditional computing to human-computer interaction (HCI)



Fig. 2. Evolution of user interface technologies (1950s–present)

II. LITERATURE REVIEW

Extensive research has been conducted in the domain of gesture recognition, eye tracking, and alternative human–computer interaction methods. Early gesture-based systems relied on hardware sensors such as data gloves and infrared markers, which, although accurate, were expensive and cumbersome to use [5]. The shift towards vision-based techniques marked a significant milestone,

enabling gesture recognition using cameras and image processing algorithms.

One of the foundational works in vision-based hand gesture recognition was presented by Freeman and Weissman, who explored the use of hand orientation and motion cues for interaction [6]. However, these early approaches were limited by lighting conditions and computational constraints. With the advent of machine learning, particularly convolutional neural networks (CNNs), gesture recognition accuracy improved significantly.

Mitra and Acharya provided a comprehensive survey on gesture recognition techniques, categorizing them into vision-based and sensor-based approaches [7]. Their study highlighted that vision-based methods offer greater user convenience but require robust algorithms to handle background clutter and occlusions. This insight influenced later works that focused on background subtraction and feature extraction techniques.

Eye-tracking technology has also evolved substantially over the years. Early eyetracking systems required specialized infrared cameras and head-mounted devices, making them unsuitable for everyday use. Jacob’s pioneering research demonstrated the potential of eye movements as an interaction modality, particularly for pointing and selection tasks [8]. Subsequent studies explored blink detection as a reliable mechanism for click events.

Zhu and Ji proposed a real-time eye gaze detection system using facial landmarks, which laid the groundwork for modern webcam-based eye tracking [9]. Their approach demonstrated that accurate eye tracking could be achieved without dedicated hardware, paving the way for low-cost assistive systems.

In the context of virtual mouse control, several researchers have combined hand gesture recognition with cursor movement. Pradeep et al. developed a hand gesturecontrolled mouse using color markers and image segmentation techniques [10]. While effective, the reliance on markers limited usability. Later works eliminated the need for markers by employing skin color detection and deep learning models.

Khan and Ibraheem introduced a vision-based virtual mouse system using fingertip tracking and gesture classification [11]. Their system supported basic mouse operations and demonstrated high accuracy under controlled conditions. However, the absence of eye-based interaction restricted its usability for users with limited hand mobility.

Multimodal interaction systems combining eye and hand inputs have gained attention in recent years. Chen et al. proposed an HCI framework integrating gaze tracking and

gesture recognition to enhance interaction robustness [12]. Their study concluded that multimodal systems reduce user fatigue and improve interaction accuracy.

More recent studies have leveraged frameworks like MediaPipe for real-time hand and face landmark detection. Lugaresi et al. introduced MediaPipe Hands, a high-fidelity hand tracking solution capable of detecting 21 hand landmarks in real time [13]. This advancement significantly simplified the implementation of gesture-based systems and improved performance on consumer-grade hardware.

Despite these advancements, challenges remain in terms of robustness, lighting variations, user adaptability, and false gesture detection. The existing literature highlights the need for integrated systems that are accurate, affordable, and adaptable to diverse users and environments. The proposed system builds upon these research efforts by combining hand gesture tracking and eye blink detection into a unified virtual mouse framework.

III. PROBLEM STATEMENT

Traditional computer input devices such as the mouse and keyboard have remained the primary tools for Human–Computer Interaction for many years due to their precision and ease of use. However, these devices require continuous physical contact and fine motor control, making them difficult to use for individuals with physical disabilities, motor impairments, or temporary movement limitations. In addition, in environments such as hospitals, laboratories, industrial workplaces, and public systems, touch-based interaction may be inconvenient, unhygienic, or impractical. Existing gesture-based virtual mouse systems attempt to solve these challenges, but many rely solely on hand gesture recognition, which often leads to false gesture detection, accidental clicks, background sensitivity, and user fatigue during prolonged use.

Similarly, eye-tracking systems alone may face limitations such as calibration complexity, inaccurate blink interpretation, and discomfort during long-term usage. Therefore, there is a need for an efficient, low-cost, touchless interaction system that combines multiple input modalities to improve reliability and accessibility. The proposed Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer addresses these issues by integrating hand gesture recognition for cursor movement and eye blink detection for click actions, creating a more intuitive, accurate, and practical alternative to conventional input devices for modern Human–Computer Interaction applications.

IV. OBJECTIVES

The primary objective of this research is to design and develop a touchless virtual mouse system that enables users to control computer mouse operations through hand gestures and eye blink detection using computer vision and machine learning techniques. The system aims to provide an alternative Human–Computer Interaction method that reduces dependence on conventional hardware devices such as the mouse and keyboard.

Another important objective is to improve accessibility for individuals with physical disabilities, limited motor control, or users who face difficulty operating traditional input devices. The proposed framework also seeks to provide a hygienic and contactless interaction solution suitable for environments such as hospitals, laboratories, educational institutions, and industrial workplaces where physical contact may be undesirable.

The research further aims to develop a low-cost and hardware-independent system using a standard webcam and open-source technologies such as OpenCV, MediaPipe, Python, and PyAutoGUI, ensuring affordability and easy implementation. Additional objectives include achieving accurate real-time gesture recognition, reliable eye blink detection, low response latency, and reduced false activations compared to gesture-only systems.

Finally, the study aims to explore the practical applications of multimodal touchless interaction systems and contribute to the advancement of intelligent, accessible, and user-friendly Human–Computer Interaction technologies for future assistive and automation-based applications.

V. METHODOLOGY

A. System Architecture:

The proposed Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer follows a modular methodology for enabling real-time touchless computer interaction using computer vision and machine learning techniques. The system is designed to capture live video input, detect hand and facial landmarks, interpret user actions, and convert them into mouse control commands.

B. Video Acquisition:

A standard webcam is used as the primary input device for continuous video acquisition. The captured frames are processed using OpenCV for image preprocessing tasks such as resizing, color conversion, and noise reduction to improve detection performance. MediaPipe is employed for real-time hand landmark detection, enabling the

identification of key points such as fingertips, joints, and palm positions.

C. Hand Gesture Recognition:

For click operations, facial landmark detection is used to monitor eye movement and blinking behavior. The Eye Aspect Ratio (EAR) method is applied to determine whether the eye is open or closed. Intentional blink patterns are interpreted as mouse click commands, reducing dependence on hand gestures for all interactions and minimizing accidental triggers.

D. Eye Blink Detection:

The interpreted gesture and blink data are integrated through PyAutoGUI, which translates the detected actions into operating system mouse events such as cursor movement and clicking. The system operates continuously in real time, ensuring responsive interaction with minimal latency.

E. Mouse Control:

Performance evaluation is conducted based on gesture recognition accuracy, blink detection reliability, response time, and usability under different environmental conditions. This modular methodology ensures affordability, scalability, and practical implementation using widely available hardware and open-source software tools.



Fig. 3. Factors contributing to the evolution of human-computer interaction (HCI)

VI. PROPOSED METHOD

The proposed Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer utilizes a multimodal interaction approach that combines hand gesture recognition and eye blink detection for touchless computer control. The system captures live video input through a standard webcam and processes it in real time using computer vision techniques. MediaPipe is used to detect hand landmarks, particularly fingertip positions, which are mapped to screen coordinates for smooth cursor movement. Different hand gestures are interpreted to support navigation and control operations.

For mouse click actions, the system employs facial landmark detection to monitor eye movement and blinking behavior. The Eye Aspect Ratio (EAR) technique is used to identify intentional eye blinks, which are converted into click commands.

This separation of cursor movement and click operations reduces accidental activations and improves interaction accuracy. The interpreted inputs are integrated using PyAutoGUI to execute mouse events within the operating system. This low-cost, touchless framework provides an efficient and accessible alternative to conventional input devices for real-time Human–Computer Interaction.

Algorithm flow:

1. Capture frame from webcam.
2. Detect hand landmarks.
3. Detect facial landmarks.
4. Compute cursor coordinates.
5. Evaluate blink threshold.
6. Trigger corresponding mouse event.
7. Repeat in real time.



Fig. 4. Facial Landmark Detection Process Showing Input and Output

VII. IMPLEMENTATION DETAILS

The implementation uses Python as the primary language.

A. Software stack:

Python, OpenCV, MediaPipe, NumPy, PyAutoGUI

B. Hardware requirements:

Standard webcam, Personal computer/laptop, Basic CPU processing capability.

The implementation emphasizes affordability and portability.

VIII. RESULTS

TABLE I. Performance Metrics:

PARAMETER	OBSERVED VALUE
GESTURE RECOGNITION ACCURACY	92-95%
EYE BLINK DETECTION ACCURACY	~90%
AVERAGE LATENCY	<150 MS
USER ADAPTATION TIME	5-10 MINUTES

HARDWARE USED	STANDARD WEBCAM
---------------	--------------------

The system performed well under moderate lighting conditions. Cursor response remained smooth during normal hand movement. Accuracy decreased under poor illumination and cluttered backgrounds.

TABLE II. Comparative Analysis:

PARAMETER	TREDITIONAL MOUSE	PROPOSED SYSTEM
PHYSICAL CONTACT	REQUIRED	NOT REQUIRED
PORTABILITY	MODERATE	HIGH
PRECISION	VERY HIGH	MODERATE
HYGIENE	LOW	HIGH
COST	LOW	VERY LOW

IX. DISCUSSION

The experimental evaluation of the proposed Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer demonstrates that the system is capable of providing effective real-time touchless computer interaction. The integration of hand gesture recognition and eye blink detection improves overall usability by distributing interaction tasks across two different input modalities. Hand gestures provide intuitive cursor navigation, while eye blink detection enables reliable click operations, reducing dependence on a single interaction method. This multimodal design helps minimize accidental activations commonly observed in gesture-only systems.

The system achieved high gesture recognition accuracy and acceptable blink detection performance under moderate lighting conditions, with low response latency suitable for real-time applications. Users were able to adapt to the system within a short period, indicating a low learning curve and practical usability. However, certain limitations were observed during testing. Performance decreased in poor lighting environments, cluttered backgrounds, and situations involving rapid hand movement or head motion. Extended use may also result in user fatigue due to continuous gesture interaction.

Despite these challenges, the proposed framework demonstrates strong potential as a cost-effective alternative to conventional input devices, especially in applications where touchless interaction is preferred. Future enhancements using deep learning, adaptive calibration, and sensor fusion techniques could further

improve system robustness, accuracy, and long-term usability.

X. APPLICATIONS

A. *Healthcare:*

In healthcare environments, the touchless interaction capability can reduce contamination risks in hospitals, operation theaters, laboratories, and sterile medical spaces where physical contact with shared devices should be minimized. Educational institutions can also benefit from this technology, as teachers and presenters can control slides and digital content without using traditional peripherals, enabling more dynamic presentations.

B. *Accessibility:*

Additionally, the framework has potential applications in virtual reality, augmented reality, smart home automation, gaming, and next-generation intelligent interfaces. Its low-cost implementation using standard webcams and open-source software makes it practical for widespread adoption in both research and real-world touchless computing applications.

C. *Education:*

The proposed Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer has wide-ranging applications across multiple domains where touchless and intuitive human-computer interaction is beneficial. One major application area is assistive technology, where the system can help individuals with physical disabilities, limited motor control, or mobility impairments interact with computers more independently. By reducing reliance on conventional input devices, the system improves accessibility and digital inclusion.

D. *Industrial Control:*

Industrial workplaces and manufacturing environments represent another important application area, particularly where workers use gloves, handle machinery, or require hands-free interaction for safety and efficiency. Public kiosks and smart interactive systems can use this technology to provide hygienic user interaction in shared spaces.

XI. LIMITATIONS

1. *Lighting Dependency:*

The system performance is highly affected by poor or inconsistent lighting conditions, which can reduce gesture and eye detection accuracy.

2. Background Sensitivity:

Complex or cluttered backgrounds may interfere with hand landmark detection, causing false recognition or unstable cursor movement.

3. Limited Precision:

The virtual mouse provides moderate accuracy but may not be suitable for tasks requiring high precision, such as graphic design or detailed editing.

5. Webcam Quality Dependency:

System performance depends on camera resolution and frame quality, with low-quality webcams reducing detection reliability.

6. False Blink Detection:

Natural involuntary blinking or sudden head movement may occasionally trigger unintended mouse click actions.

XII. FUTURE WORK

Future improvements may include:

1. Deep Learning Integration:

Advanced deep learning models can be implemented to improve gesture recognition accuracy under varying environmental conditions.

2. Adaptive Calibration: Personalized calibration techniques can enhance performance based on individual user behavior and physical capabilities.

3. Improved Eye Tracking:

More accurate gaze and blink detection methods can reduce false click activations and improve usability.

4. Sensor Fusion:

Combining webcams with depth sensors or motion sensors can improve robustness and detection stability.

5. Multi-User Support:

Future versions can support multiple users simultaneously for collaborative interaction environments.

6. AR/VR Integration:

The framework can be extended for use in augmented reality and virtual reality applications for immersive interaction.

7. Cross-Platform Optimization:

Optimizing the system for mobile devices, embedded systems, and smart devices can increase practical adoption.

XIII. CONCLUSION

This research presents the design and implementation of a Gesture-Based Eye and Hand Controlled Virtual Mouse Pointer as an innovative touchless Human–Computer Interaction system using computer vision and machine learning techniques. The proposed framework successfully demonstrates that conventional input devices such as the mouse can be partially replaced with a multimodal interaction approach that combines hand gesture recognition and eye blink detection for real-time computer control. By utilizing a standard webcam along with open-source technologies including OpenCV, MediaPipe, Python, NumPy, and PyAutoGUI, the system provides a cost-effective and accessible alternative without requiring specialized hardware.

The developed system effectively enables cursor movement through hand gesture tracking while using eye blink detection for mouse click operations, thereby improving interaction efficiency and reducing accidental activations commonly found in gesture-only systems. Experimental evaluation indicates that the framework achieves high gesture recognition accuracy, reliable blink detection performance, and low response latency under moderate environmental conditions. These results confirm the practical feasibility of the proposed system for real-time applications.

One of the key strengths of this research lies in its accessibility and versatility. The proposed solution can significantly benefit individuals with physical disabilities, users with limited motor control, and environments where touch-based interaction is inconvenient or unhygienic, such as healthcare institutions, laboratories, educational spaces, and industrial settings. The integration of multimodal interaction enhances usability by reducing dependency on a single control mechanism.

However, certain limitations remain, including sensitivity to lighting conditions, background complexity, reduced precision for high-detail tasks, and possible user fatigue during prolonged usage. Despite these challenges, the system establishes a strong foundation for future advancements in touchless interaction technology.

Future improvements may involve deep learning-based gesture classification, adaptive calibration techniques, improved eye-tracking precision, sensor fusion, and compatibility with augmented or virtual reality systems. Overall, this research demonstrates that multimodal gesture- and eye-controlled interfaces represent a promising step toward intelligent, inclusive, and next-generation Human–Computer Interaction technologies.

REFERENCES:

1. Pavlovic, V. I., Sharma, R., & Huang, T. S., "Visual interpretation of hand gestures for human– computer interaction: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677–695, 1997.
2. Mitra, S., & Acharya, T., "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 37, no. 3, pp. 311–324, 2007.
3. Turk, M., & Pentland, A., "Face recognition using eigenfaces," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, 1991.
4. Viola, P., & Jones, M., "Rapid object detection using a boosted cascade of simple features," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518, 2001.
5. Dalal, N., & Triggs, B., "Histograms of oriented gradients for human detection," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–893, 2005.
6. Zhang, Z., "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
7. Shotton, J., et al., "Real-time human pose recognition in parts from single depth images," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1297–1304, 2011.
8. Starner, T., & Pentland, A., "Real-time American Sign Language recognition from video using hidden Markov models," *International Symposium on Computer Vision*, pp. 265–270, 1995.
9. Wachs, J. P., Kölsch, M., Stern, H., & Edan, Y., "Vision-based hand-gesture applications," *Communications of the ACM*, vol. 54, no. 2, pp. 60–71, 2011.
10. Szeliski, R., *Computer Vision: Algorithms and Applications*, Springer, 2010.
11. Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 2006.
12. Goodfellow, I., Bengio, Y., & Courville, A., *Deep Learning*, MIT Press, 2016.
13. Bradski, G., "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
14. Zhang, X., Sugano, Y., Fritz, M., & Bulling, A., "Appearance-based gaze estimation in the wild," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4511–4520, 2015.
15. Hansen, D. W., & Ji, Q., "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, 2010.
16. Kumar, P., & Kumar, R., "Hand gesture recognition using computer vision: A survey," *International Journal of Advanced Research in Computer Science*, vol. 7, no. 3, pp. 203–207, 2016.
17. Shaikh, S., & Raut, P., "Virtual mouse using hand gesture," *International Journal of Engineering Research & Technology (IJERT)*, vol. 8, no. 4, pp. 227–231, 2019.
18. Gupta, H., Sharma, S., & Verma, R., "Eye blink based mouse control system," *International Journal of Computer Applications*, vol. 79, no. 2, pp. 12–16, 2013.
19. Jain, A., & Singh, A., "Touchless computer control using hand gesture recognition," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 6, no. 5, pp. 4563–4568, 2018.
20. Chaudhary, A., & Raheja, J. L., "A survey on gesture recognition for human–machine interaction," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 11, pp. 567–575, 2016.