International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI Based Mouse Using Hand Gesture – The Review

Akshay Kumar[*1], Jasleen Kaur[*2], Vansh Taneja[*3], Aditi Jha[*4], Shreya Pant[*5], Swati Bhardwaj[*6]

[*1] UG Scholar, Department of CSE, Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, India

[*2] PG Scholar, Department of Biotechnology, Chandigarh group of colleges, Punjab Technical University, Mohali, India

[*3] UG Scholar, Department of CSE, Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, India

[*4] UG Scholar, Department of CSE, Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, India

[*5] UG Scholar, Department of CSE, Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, India

[*6] UG Scholar, B.DS, Bhojia Dental College and Hospital, Himachal Pradesh University, Baddi, India

ABSTRACT

As computers become more ubiquitous in society, traditional modes of interaction like the mouse and keyboard may become limiting. A potential approach to establishing natural and intuitive human-computer interaction is visual interpretation of hand motions. The literature on visual interpretation of hand gestures in relation to its function in human-computer interaction is reviewed in this article, with a focus on key contributions made by researchers in this vibrant topic. The objective of this review is to introduce gesture recognition as a mechanism for interaction with computers, exploring its potential to improve the efficiency and ease of information flow between humans and machines. The paper concludes with a discussion of the challenges and future directions for research in this field.

Keywords: Computer vision, Face recognition, Hand recognition, Artificial Intelligence, Human Computer Interaction.

I. INTRODUCTION

The field of human-computer interaction (HCI) has witnessed significant growth and development in recent years, driven by the need to create more intuitive and efficient ways for individuals to interact with computers. One approach to achieving this is through hand gesture recognition, which allows users to control computers or other devices using predictable hand movements. However, the challenge lies in ensuring that the resulting gestures are easily understood and interpreted by the computer, which requires the use of advanced technologies and techniques. Artificial intelligence (AI) has played a significant role in the development of hand gesture recognition systems [1,2]. This problem has been tackled using several AI techniques, including neural networks, natural language processing, pattern matching, and fuzzy systems, with variable degrees of success. These techniques rely on the analysis and interpretation of visual data captured by cameras or other sensors, and they can be used to extract meaningful features from the input data, such as hand position, orientation, and movement.

Face recognition is another area of AI that has been extensively studied and applied in HCI. Facial recognition systems typically involve the use of machine learning algorithms to identify and extract facial features from input images or video streams. To identify the person, these traits are then compared to a database of recognized faces. There are several uses for facial recognition, including in access control, security systems, and customized user interfaces. Hand gesture recognition systems typically involve three main stages: extraction, feature evaluation and extraction, and classification or recognition. In the extraction stage, the input data, which can include images, video, or sensor data, is processed to extract relevant information about the hand or hands in the scene [3]. This involves techniques such as background subtraction, thresholding, or edge detection. The feature evaluation and extraction stage involve the analysis of the extracted data to identify relevant features, such as hand position, orientation, and movement. This stage involves more advanced AI techniques, such as neural networks or pattern recognition algorithms. In the final classification or recognition stage, the extracted features are compared against a database of known gestures to identify the gesture being performed by the user. Several challenges exist in the development of hand gesture recognition systems, including the need for accurate and robust algorithms that can handle variations in lighting, background, and user pose [4]. The creation of more precise and effective systems that can recognize and decipher a variety of facial expressions and gestures is made possible by developments in computer vision and machine learning techniques [5]. While challenges remain, such as ensuring the security and privacy of user data, the potential benefits of these technologies make them an exciting area for continued research and development.

II.NECESSITY

The application domain of an AI-based mouse using face recognition and hand gesture recognition is vast and diverse. It can be applied in various industries and fields, such as gaming, healthcare, automotive, education, and entertainment, to name a few [6]. In the gaming industry, this technology can revolutionize the way players interact with the game. Instead of using a traditional mouse and keyboard, players can use hand gestures and facial expressions to control the game. As an example, hand gestures could be used to direct a character's movement, whereas facial expressions are useful to communicate feelings and initiate gameplay. This technology can be applied to the healthcare industry to create assistive gadgets for those with disabilities. People with motor impairments can use

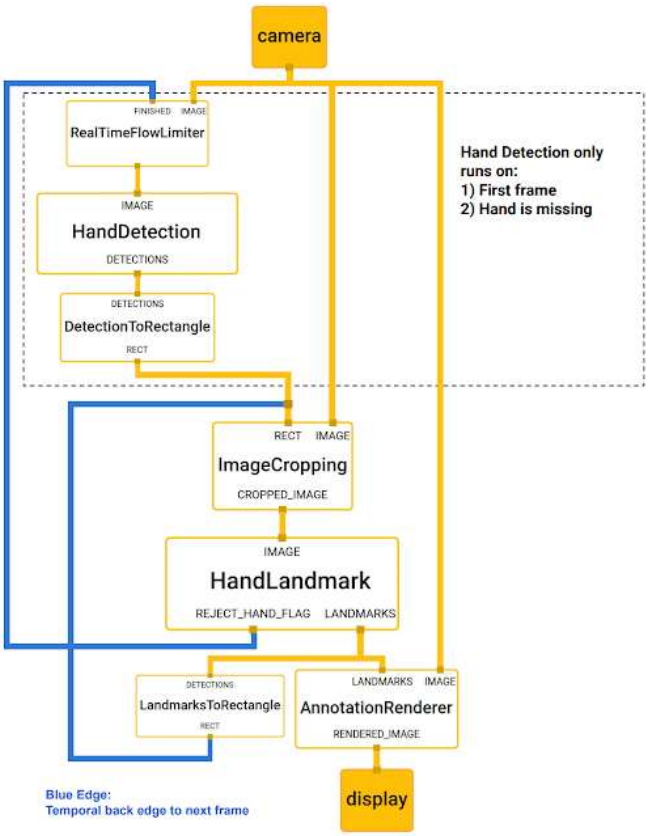
hand gestures to control their computers or smart devices, while people with speech impairments can use facial expressions to convey their thoughts and emotions. In education, this technology can be used to develop interactive and immersive learning environments. For instance, students can use hand gestures to manipulate 3D objects or navigate virtual environments, while teachers can use facial expressions to provide feedback and engage students. In entertainment, this technology can be used to create immersive and interactive experiences for audiences. performers can use facial expressions to control the lighting, sound, and visual effects, while audiences can use hand gestures to interact with the performers or the environment. Additionally, people can communicate with hologram displays in an even more natural and easy way, without the necessity of physical mouse. This can be particularly useful in situations where physical controllers may be impractical or inconvenient, such as in virtual meetings, presentations, or training simulations. Sign language recognition can be integrated with hand gesture recognition to provide a more comprehensive communication system [7]. This will allow users to use hand gestures and sign language interchangeably to control the computer and communicate with others [8].

III.METHODOLOGY

Vision-based interactions is a complicated area of interdisciplinary research that spans many disciplines, including bioinformatics, psychology, machine learning, graphics, image processing, and computer vision. Building a successful working system for vision-based interaction requires meeting specific requirements. Robustness is one of the key needs since in the actual world, there are many variables that might cause visual information to be noisy, insufficient, and rich, including shifting lighting, clutter, occlusion, dynamical backgrounds, etc. Therefore, vision-based systems must be user-independent and resistant to these influences. Computational effectiveness is a crucial requirement since vision-based interaction frequently calls for real-time systems. As a result, vision-based interface algorithms and methods should be both efficient and affordable. The system should also be tolerant of user errors or malfunctions, and the effects of faults should be kept to a minimum. Instead, then letting the machine make poor judgements, users can be requested to perform certain tasks again. Additionally, the system must be adaptable to a variety of applications, including desktop environments, virtual environments (VE), robot navigation, and sign language recognition. Consequently, the foundation of the system for vision-based interaction should be the same for all these applications. Most current systems rely on either motion recognition or skin color to identify and separate the gesturing hand from the rest of the background [9]. However, the right choice of features or cues, together with sophisticated recognition algorithms, will determine whether any current or future research in the discipline of Human Computer Interaction (HCI) employing hand gestures is successful or unsuccessful. Therefore, future research in this field should concentrate on creating more sophisticated methods to enhance the functionality of vision-based interface systems.

Fig a show the working of Google's Mediapipe module for hand geture recognition whose working is explained below.

1. Camera: The process starts with a camera capturing a frame.
2. RealTimeFlowLimiter: This component restricts the hand detection to run only on the first frame and when a hand is missing from the scene. This helps to optimize processing and conserve resources.
3. Hand Detection: This is the core module that is responsible for detecting hands in the frame. It can employ various techniques such as skin color detection, template matching, or machine learning algorithms.
4. Detections: The hand detection module outputs a list of detections, which are essentially bounding boxes around the hands that it identified in the frame.
5. Detection ToRectangle: This component converts the detections (bounding boxes) into a format that can be understood by the subsequent steps in the process.
6. Image Cropping: The image is cropped based on the bounding boxes obtained in the previous step. This helps to focus on the regions of interest (the hands) and discard irrelevant background information.
7. HandLandmark: This module extracts hand landmarks from the cropped image. Hand landmarks are a set of key points that define the structure of the hand, such as the base of the palm, the tips of the fingers, and the knuckles.
8. REJECT HAND_FLAG: This step checks for the presence of a 'REJECT_HAND_FLAG'. It is unclear from the diagram what this flag signifies, but it is likely used to exclude hand detections that do not meet certain criteria (e.g., low confidence score, wrong size or aspect ratio).
9. Landmarks ToRectangle: Similar to step 5, this component converts the extracted hand landmarks into a format that can be used by the following steps.
10. AnnotationRenderer: This module overlays the detected hand landmarks on the original image, effectively visualizing the hand pose.
11. Rendered Image: This is the final output of the process, which is an image with the detected hand landmarks visualized on top.
12. Blue Edge: This indicates the path that the process continues to follow for subsequent frames, presumably repeating steps 2 through 11.
13. Temporal back edge to next frame: This indicates that the process can jump back to the beginning if a hand is not detected in the current frame.



(Fig a.) Hand recognition algorithm [10]

3.1 Feature for gesture recognition

Recognizing hand gestures can be challenging due to their varying shape, motion, and texture. However, because of self-occlusion and lighting conditions, geometric elements like fingertips, finger orientations, and hand shapes may not always be accurate in identifying static hand postures. Color, texture, and silhouette are examples of non-geometric traits that are insufficient for recognition. As a result, the recognizer automatically selects features from the input image. Hand features can be derived through three approaches.

- First approach for deriving hand features is the model-based or kinematic model approach. This technique uses a three-dimensional representation of the hand to infer the position of the palm and joint angles by searching for the kinematic parameters that corresponds with an edge-based image of hand. However, this method is limited by the difficulty in extracting features from the texture less human hand, leading to unreliable edge extraction from occluding boundaries. To address this issue, high-contrast and homogeneous backgrounds relative to the hand can be used to facilitate unambiguous correspondence between the edges and the model edges.
- Second is the view-based approach, that simulates the hand using a set of 2-dimensional intensity images, is the second method. A series of perspectives are used to model gestures. This alternative approach has received a lot of attention recently since kinematic model-based approaches have fitting issues.
- The third approach, known as the low-level features-based method, quickly and noise-resistantly extracts low-level picture measurements. The center of the hand, the main axes that describe an elliptical confining region of a hand, as well as the optical flow/affine motion of the hand region in a picture are a few low-level properties that are suggested by this method. Many gesture applications merely need to link the input video to the gesture; full hand reconstruction is not always essential for gesture detection.

3.2 Methods for gesture recognition

Gesture recognition delves into the realm of creating systems that can decipher and interpret human hands, turning them into meaningful interactions. Unlike simple movement tracking, gesture recognition unlocks a deeper layer, transforming motions into specific commands that computers can understand. Two primary approaches dominate the field of gesture recognition:

1. Data Gloves: This method equips users with wearable sensors, often in the form of gloves, that track hand and finger movements with high precision. For example, surgeons might use them to manipulate virtual instruments during training, or artists could paint intricate details in 3D sculpting software.
2. Computer Vision: This approach harnesses the power of cameras and computer vision algorithms to analyze visual cues without requiring any physical sensors worn by the user. Like waving your hand to change slides in a presentation or interacting with virtual objects by simply reaching out. Computer vision offers freedom of movement and wider accessibility, making it ideal for everyday interactions with devices and smart environments.

3.2.1 Hand gestures based on instrumented glove approach

Data gloves equipped with sensors have the unique ability to capture detailed hand movements and positions. They readily provide specific coordinates for palm and finger locations, along with their orientation and configurations using attached sensors. However, this approach comes with drawbacks. One key limitation is the physical connection to a computer, hindering the natural flow of interaction. Additionally, the cost of these devices can be quite high. Looking forward, advancements in touch technology promise a more engaging experience. Modern glove-based systems incorporate industrial-grade haptic technology, allowing users to "feel" the shape, texture, and weight of virtual objects through microfluidic technology. This opens exciting possibilities for more immersive and intuitive interactions. While offering valuable capabilities, data gloves currently face limitations in terms of user mobility and affordability. Future developments in touch technology have the potential to address these challenges, paving the way for even more sophisticated and accessible glove-based interactions.

3.2.2 Hand gestures based on computer vision approach

Computer vision offer a popular, convenient, and effective approach for contactless communication between humans and computers. This method leverages various camera configurations, including monocular, fisheye, time-of-flight (TOF), and infrared (IR) cameras. However, it involves various challenges. These include variations in lighting, complexities with the background, the impact of occlusions (objects blocking the view), the presence of intricate backgrounds, trade-offs between processing time and resolution/frame rate, and the potential for foreground or background objects to mimic skin tones or hand appearances, leading to misidentification. The following sections will delve deeper into these challenges.

3.2.2.1 Color-based recognition

This method employs a camera to track hand movement using a glove adorned with distinct color markers. It facilitates interaction with 3D models, enabling tasks like zooming, moving, drawing, and even writing on a virtual keyboard with remarkable flexibility. The glove's colors allow the camera sensor to pinpoint the palm and finger locations, facilitating the extraction of a geometric hand shape model. This approach boasts advantages like ease of use and affordability compared to sensor-based data gloves. However, wearing colored gloves remains necessary, potentially hindering natural and spontaneous interaction within the human-computer interface (HCI).

3.2.2.2 Appearance based recognition

This method, relying solely on 2D image features like hand shapes and pixel intensities, offers several advantages for real-time gesture recognition. Compared to 3D models, it's simpler to implement and computationally efficient, making it well-suited for quick processing. Additionally, its ability to detect diverse skin tones enhances its adaptability. Notably, the AdaBoost algorithm plays a key role in handling occlusion by utilizing fixed key points, even on partial hand images. This allows the system to learn both static hand postures and dynamic motions through separate models, leading to more robust gesture recognition.

3.2.2.3 Motion based recognition

Motion-based recognition holds promise for detecting and extracting objects from image sequences. However, achieving reliable gesture recognition presents several challenges. The AdaBoost algorithm, commonly used for object detection and motion modeling, faces limitations when incorporating multiple gestures into the recognition process or dealing with dynamic backgrounds. Additionally, occlusions among tracked hand gestures or inaccuracies in region extraction can lead to the loss of information, further impacted by long-distance interactions that affect the appearance of the tracked region. Overcoming these obstacles is crucial for realizing the full potential of motion-based gesture recognition technology.

3.2.2.4 Skeleton based recognition

Skeleton-based recognition stands out by tailoring its model parameters to excel at detecting intricate hand features. This approach leverages various representations of skeletal data to categorize hand gestures. Its strength lies in capturing geometric properties, limitations, and readily translatable data features and correlations. By focusing on these aspects, it effectively extracts both geometric and statistical features crucial for recognition. Key features used include joint orientation, distances between joints, the spatial location of each joint, the angle between joints, and even their trajectories and curvatures. This comprehensive set of features empowers the system to accurately decipher complex hand movements, paving the way for rich and intuitive human-computer interaction.

3.2.2.5 Depth based recognition

Researchers have explored various camera types for hand gesture recognition, with depth cameras offering unique advantages. These cameras capture 3D information (depth), unlike regular color cameras that capture 2D projections. This depth data is less affected by factors like lighting, shadows, and color variations, potentially improving recognition accuracy. However, the cost, size, and limited availability of depth cameras currently pose challenges for widespread adoption.

3.2.2.6 3D Model based recognition

The 3D hand model hinges on a highly flexible 3D Kinematic model. Key hand parameters are estimated by comparing the real image with the projected 2D appearance of this 3D model. Essentially, the 3D model acts as a template for human hand features, like pose estimation. It can take the form of a volumetric, skeletal, or detailed 3D model closely resembling the user's hand. During an iterative process, the parameters of the 3D model are refined to better match the real hand. To further enhance accuracy, depth information is also incorporated into the model. This approach allows the 3D model to continuously adapt and accurately represent the user's hand and its movements.

3.2.2.7 Deep learning based recognition

While deep learning boasts significant potential in modern applications due to its powerful learning capabilities, it isn't without its challenges. By leveraging multi-layered networks to analyze data, it delivers impressive predictions. However, a major hurdle lies in the substantial datasets required to train these algorithms, potentially leading to lengthy processing times. Therefore, optimizing data acquisition and processing methods remains crucial for unlocking deep learning's full potential in real-world applications.

3.3 Applications of hand gesture based mouse

Common applications of Hand Gesture based mouse are:

1. Medical- Surgeons often require intricate details of a patient's anatomy or detailed organ models during surgery. Gesture recognition can be a powerful tool in such situations, allowing surgeons to manipulate these virtual representations hands-free. Surgeons can perform operations like zooming, rotating, and cropping medical images or navigating through different views simply by gesturing – all without needing a mouse, keyboard, or touchscreen.
2. Sign Language Recognition- For individuals who face challenges with spoken communication, sign language serves as a vital bridge. To bridge the gap between sign language and the wider world, numerous research efforts have explored sign language recognition technologies. One approach involves glove-mounted sensors that track hand movements and translate them into corresponding outputs, aiming to offer a seamless communication channel for those who rely on sign language.
3. Automation- Hand recognition can be used for home automation by creating a specific gesture to control light or fan speed.
4. Gaming- It can also be used to play games wherein hand gesture are used to control the game

IV.RESEARCH GAPS AND CHALLENGES

A review of existing research reveals a gap in applying hand gesture recognition specifically to healthcare, despite strong focus on computer applications, sign language, and virtual object interaction. While numerous studies focus on improving recognition frameworks or developing new algorithms, real-world healthcare applications remain underexplored. The key challenge lies in building a robust and accurate system that overcomes common issues like limited environments and background noise. Existing proposals primarily fall into two categories of computer vision techniques, offering a foundation for further development in healthcare-oriented gesture recognition. This gap presents an exciting opportunity to translate the potential of hand gestures into tangible benefits for medical care and patient interaction. A straightforward approach to gesture recognition utilizes image processing techniques through libraries like Open-NI or OpenCV. This allows for real-time interaction, but comes with trade-offs due to the computational needs of processing continuous video streams. Some drawbacks include limitations in handling complex backgrounds, varying lighting conditions, limitations in recognizing gestures at different distances, and challenges in differentiating between multiple objects or simultaneous gestures. While convenient and readily available, these methods necessitate careful consideration of their inherent limitations to ensure their suitability for a particular application. Second approach gesture recognition aims to match human movements against a pre-existing dataset to interpret commands. Simpler gestures require less complex algorithms, but intricate patterns demand sophisticated approaches like deep learning and AI. These techniques excel at real-time gesture identification by comparing user input to stored postures and commands within the dataset. However, limitations arise from potential inaccuracies in classification algorithms, leading to missed gestures. Additionally, matching against a large dataset takes longer compared to simpler, dedicated algorithms. The trade-off between complexity and accuracy remains a key challenge in gesture recognition technology.

V.CONCLUSION

Traditionally, interacting with technology often feels unnatural and restrictive. Hand gesture recognition steps in to remedy this, offering a more intuitive and versatile alternative. This technology eliminates the need for bulky hardware, promoting naturalness and ease of use. It also circumvents potential issues associated with traditional devices, like button malfunctions or limited functionality. However, achieving this seamless interaction requires robust algorithms. Previous discussions highlighted the challenges of developing reliable systems utilizing camera sensors. These algorithms must contend with common issues like varying lighting conditions and complex backgrounds to deliver accurate results. It's important to remember that each gesture recognition technique, be it glove-based or vision-based, has its own strengths and weaknesses. Some excel in situations requiring high precision, while others favor ease of use and accessibility. Selecting the right approach depends on the specific demands of the application. In Conclusion, hand gesture recognition holds immense potential to revolutionize how we interact with technology. By unlocking the power of natural human gestures, we pave the way for a future of intuitive, accessible, and empowering human-computer interaction.

VI.REFERENCES

[1] A. Kumar, N. Pathak, M. Kirola, N. Sharma, B. Rajakumar and K. Joshi, "AI based mouse using Face Recognition and Hand Gesture Recognition," 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/ICAIA57370.2023.10169243.

[2] Alessandro Floris, Simone Porcu, and Luigi Atzori. 2024. Controlling Media Player with Hands: A Transformer Approach and a Quality of Experience Assessment. ACM Trans. Multimedia Comput. Commun. Appl. 20, 5, Article 132 (May 2024), 22 pages. https://doi.org/10.1145/3638560

[3] F. A. Farid et al., "Single Shot Detector CNN and Deep Dilated Masks for Vision-Based Hand Gesture Recognition from Video Sequences," in IEEE Access, doi: 10.1109/ACCESS.2024.3360857.

[4] Zhang, H., Yin, L. & Zhang, H. A real-time camera-based gaze-tracking system involving dual interactive modes and its application in gaming. Multimedia Systems 30, 15 (2024). https://doi.org/10.1007/s00530-023-01204-9.

[5] Y. G. Pame and V. G. Kottawar, "A Novel Approach to Improve User Experience of Mouse Control using CNN Based Hand Gesture Recognition," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-6, doi: 10.1109/ICCUBEA58933.2023.10392164.

[6] R. N. Phursule, G. Y. Kakade, A. Koul and S. Bhasin, "Virtual Mouse and Gesture Based Keyboard," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-4, doi: 10.1109/ICCUBEA58933.2023.10392123.

[7] Palivela, L.H., Premanand, V., Begum, A. (2023). Hand Gesture-Based AI System for Accessing Windows Applications. In: Shakya, S., Tavares, J.M.R.S., Fernández-Caballero, A., Papakostas, G. (eds) Fourth International Conference on Image Processing and Capsule Networks. ICIPCN 2023. Lecture Notes in Networks and Systems, vol 798. Springer, Singapore. https://doi.org/10.1007/978-981-99-7093-3_43

[8] Hoppe, A.H., Klooz, D., van de Camp, F., Stiefelhagen, R. (2023). Mouse-Based Hand Gesture Interaction in Virtual Reality. In: Stephanidis, C., Antona, M., Ntoa, S., Salvendy, G. (eds) HCI International 2023 Posters. HCII 2023. Communications in Computer and Information Science, vol 1836. Springer, Cham. https://doi.org/10.1007/978-3-031-36004-6_26

[9] A. Kumar, R. Pandey, K. Alam, H. Anandaram, K. Joshi and S. Chaudhary, "Hand Gesture based AI Controller for Presentation, Virtual Drawing and System Volume Management," 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), Dhulikhel, Nepal, 2024, pp. 620-624, doi: 10.1109/ICIPCN63822.2024.00107.

[10] https://blog.research.google/2019/08/on-device-real-time-hand-tracking-with.html