IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

VIRTUAL- MOUSE CONTROL USING GESTURE RECOGNITION

Bhargavi Sajja[1], Dr. G. Srinivasa Rao[2], Sneha Bellamkonda[3], Lalitha Yagnasri Bollimuntha [4], Lakshmi Gayathri Goggi [5]

[1]Assistant Professor, Department of Electronics and Communication Engineering, Bapatla Women's Engineering College, Bapatla, A.P, INDIA- 522101
[2]Professor, Department of Electronics and Communication Engineering, Bapatla Women's Engineering College, Bapatla, A.P, INDIA- 522101
[3,4,5]U. G Student, Department of Electronics and Communication Engineering, Bapatla Women's Engineering College, Bapatla, A.P, INDIA- 522101

Abstract: In recent years, Human-Computer Interaction (HCI) has evolved significantly, with gesture-based interfaces offering an intuitive alternative to traditional input devices like keyboards and mice. This study presents a Virtual Mouse Control System utilizing Vision Transformer (ViT), Optical Flow, and Monocular Depth Estimation to achieve accurate gesture recognition and cursor control. The system captures real-time hand movements via a webcam, processes them using ViT for robust feature extraction, and applies Optical Flow to track motion trajectories dynamically. Additionally, Monocular Depth Estimation enhances spatial understanding by estimating the depth of gestures, enabling precise control over cursor movements and click actions. The proposed method eliminates the need for external sensors, making it cost-effective and scalable for real-world applications. Experimental results demonstrate high accuracy in gesture detection, seamless cursor navigation, and minimal latency, proving its potential for accessibility solutions and touchless computing environments.

IndexTerms **- Virtual Mouse, Gesture Recognition, Vision Transformer, Optical Flow, Monocular Depth Estimation, Human-Computer Interaction**

I.INTRODUCTION

Touchless interfaces have drawn a lot of interest in the age of sophisticated Human-Computer Interaction (HCI) because of their capacity to offer smooth and simple control mechanisms. Due to the physical limitations of traditional input devices like keyboards, mouse, and touchpads, there is an increasing need for alternate forms of engagement. Users may explore and interact with computers using natural hand movements thanks to a new and effective method called gesture recognition-based virtual mouse control. This study presents a Virtual Mouse Control System that achieves high-precision gesture identification by utilizing Vision Transformer (ViT), Optical Flow, and Monocular Depth Estimation. For reliable hand gesture identification and classification, ViT, a deep learning model created for visual tasks, is used, guaranteeing precise user input interpretation. The motion trajectory of hand gestures is tracked by optical flow, allowing for fluid click and cursor movements. Furthermore, the system is improved by Monocular Depth Estimation, which reduces false detections, improves spatial awareness, and determines the depth of movements. The suggested system is a scalable and affordable solution since it does not require external gear like gloves or specialized sensors thanks to the integration of these cutting-edge technology. It offers a user-friendly, contactless computing experience that is advantageous for a number of applications, including smart home interfaces, situations where hygiene is crucial, and accessibility solutions for people with disabilities. The technical details, experimental assessments, and possible uses of this gesture-based virtual mouse system are examined in this work.

II. LITERATURE SURVEY

- In their 1990 paper [1] titled *"Gesture Recognition with a Dataglove,"* Quam et al. explored the innovative use of a dataglove for recognizing hand gestures, marking one of the early efforts in human-computer interaction research. Presented at the IEEE Conference on Aerospace and Electronics, the study detailed how the dataglove captured hand movements and finger positions to interpret various gestures, enabling more natural and intuitive communication with machines. This research played a foundational role in the development of gesture-based control systems, which have since been applied in areas such as virtual reality, robotics, and assistive technologies. [Quam et al., 1990]
- In the [2] 2015 study titled *"Optical Mouse Sensor-Based Laser Spot Tracking for HCI Input,"* Guoli Wang and colleagues presented a novel approach to human-computer interaction (HCI) by utilizing an optical mouse sensor to track laser spot

movements. Featured in the *Proceedings of the Chinese Intelligent Systems Conference*, the research introduced a low-cost, efficient method for capturing user input through laser gestures projected onto a surface. By repurposing the widely available optical mouse sensor, the system offered a practical alternative for interactive applications, highlighting its potential in environments where traditional input devices may not be feasible. This work contributed to expanding the versatility of HCI technologies. [Guoli Wang et al., 2015]

- In the [3] 2013 paper *"Supporting Hand Gesture Manipulation of Projected Content with Mobile Phones,"* Baldauf and Frohlich explored the integration of mobile phones with hand gesture recognition for manipulating projected content. Presented at the European Conference on Computer Vision, the study focused on enhancing user interaction by combining the portability of mobile devices with intuitive hand gestures, allowing users to interact with virtual content in a more natural and flexible manner. The research highlighted the potential of mobile phones as both a tool for gesture recognition and a medium for controlling projected displays, paving the way for new types of interactive experiences in augmented reality and interactive multimedia. [Baldauf & Frohlich, 2013]
- In the [4] 2021 paper *"Virtual Mouse Hand Gestures,"* Roshnee Matlani, Roshan Dadlani, Sharv Dumbre, Shruti Mishra, and Abha Tewari explored the development of a system that uses hand gestures to control a virtual mouse, aiming to provide a more intuitive and accessible method for interacting with computers. Presented at the International Conference on Technology Advancements and Innovations, the study focused on recognizing and interpreting hand movements to perform standard mouse functions such as clicking, scrolling, and dragging. This innovative approach enhances user experience by eliminating the need for physical input devices, making it especially beneficial in environments where traditional peripherals might be impractical. The research demonstrates the growing potential of gesture-based interfaces in improving human-computer interaction. [Matlani et al., 2021]
- In the [5] 2016 paper *"Hand Gesture Recognition for Human-Computer Interaction,"* Mayur Yeshi, Pradeep Kale, Bhushan Yeshi, and Vinod Sonawane discussed the implementation of hand gesture recognition systems as a method for improving human-computer interaction (HCI). Published in the *International Journal of Scientific Development and Research*, the study highlighted various techniques and algorithms used to identify and interpret hand gestures, enabling more natural and intuitive user interactions with computing devices. The authors explored the challenges and potential applications of gesture-based interfaces, emphasizing their role in areas such as virtual reality, gaming, and assistive technologies. This research underscored the importance of gesture recognition as a key advancement in the field of HCI, offering users a more immersive and hands-free way to interact with technology. [Yeshi et al., 2016]
- In the [6] 2021 paper *"Deep Learning Based Real-Time AI Virtual Mouse System Using Computer Vision to Avoid COVID-19 Spread,"* Shriram, Nagaraj, Jaya, Sankar, and Ajay proposed an innovative solution to mitigate the spread of COVID-19 by using a deep learning-based virtual mouse system. Published in the *Journal of Healthcare Engineering*, the study focused on leveraging computer vision and AI to recognize hand gestures, enabling users to control a virtual mouse without physical contact with any hardware. This touchless interaction method not only addresses health concerns related to the pandemic but also enhances user convenience in environments where hygiene is a priority. The system demonstrated the potential of AI and computer vision in facilitating safer and more efficient human-computer interactions. [Shriram et al., 2021]
- In the [7] 2020 paper *"Implementing Hand Gesture Mouse Using OpenCV,"* Steven Raj, Veeresh Gobbur, Praveen, Rahul Patil, and Veerendra Naik explored the use of hand gesture recognition to control a virtual mouse, utilizing the OpenCV library for image processing. Published in the *International Research Journal of Engineering and Technology*, the study focused on developing a system that tracks hand movements and translates them into mouse actions like pointing, clicking, and scrolling. By employing OpenCV, the authors were able to create an efficient and cost-effective solution for touchless interaction with computers. This research contributes to the growing field of gesture-based interfaces, offering practical applications in various sectors, including accessibility and user-friendly computing. [Raj et al., 2020]
- In the [8] 2013 paper *"Cursor Control System Using Hand Gesture Recognition,"* Sneha U., Monika B., and Ashwini M. presented a system for controlling a computer cursor through hand gestures, aiming to provide a more intuitive and hands-free method of interaction. Published in the *International Journal of Advanced Research in Computer and Communication Engineering*, the study focused on using image processing techniques to recognize and interpret hand movements, translating them into cursor movements on a screen. This gesture-based approach to cursor control not only enhances user experience but also contributes to the development of accessible technologies, especially for individuals with physical disabilities. The research highlights the potential of gesture recognition systems in simplifying human-computer interaction. [U et al., 2013]
- In the [9] 2022 paper *"Virtual Mouse Using YOLO,"* Krishnamoorthi, Gowtham, Sanjeevi, and Revanth Vishnu introduced a novel approach for controlling a virtual mouse using the YOLO (You Only Look Once) object detection algorithm. Presented at the *International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems*, the study focused on leveraging YOLO's real-time object detection capabilities to recognize hand gestures and translate them into mouse movements, clicks, and scrolling. This approach allows for a touchless, efficient method of interacting with computers, demonstrating the potential of advanced machine learning techniques like YOLO in enhancing user interfaces, especially in scenarios requiring hands-free operation. [Krishnamoorthi et al., 2022]
- In the [10] 2019 paper *"Virtual Mouse Implementation Using OpenCV,"* Varun K.S., Puneeth I., and Jacob T.P. explored the development of a virtual mouse system that utilizes hand gesture recognition through the OpenCV library. Presented at the *International Conference on Trends in Electronics and Informatics*, the study aimed to create a touchless and intuitive method for controlling a computer's mouse cursor. By using OpenCV for real-time image processing, the authors developed a system that tracks hand movements and translates them into corresponding actions, such as moving the cursor, clicking, and scrolling. This innovative approach has the potential to enhance user interaction, particularly in environments where touchless or hands-free interfaces are required. [Varun et al., 2019]
- In the [11] 1994 paper *"Towards a Vision-Based Hand Gesture Interface,"* Quek et al. introduced a pioneering vision-based system for interpreting hand gestures as a means of human-computer interaction. Presented in the *Proceedings of Virtual Reality Software and Technology*, the study explored the development of a gesture recognition interface using computer vision

techniques, allowing users to interact with virtual environments through hand movements. The authors highlighted the challenges and potential of creating intuitive and natural interfaces that rely on visual input, laying the groundwork for future advancements in gesture-based control systems. This research contributed to the early development of vision-based interfaces, which would later find applications in virtual reality, augmented reality, and interactive media. [Quek et al., 1994]

- In the [12] 2021 paper *"Deep Convolutional Neural Network-Based Image Classification for COVID-19 Diagnosis,"* Tharsanee R.M., Soundariya R.S., Kumar A.S., Karthiga M., and Sountharrajan S. explored the use of deep learning, specifically convolutional neural networks (CNNs), to assist in the diagnosis of COVID-19 through medical imaging. Published in *Data Science for COVID-19* by Academic Press, the study focused on applying CNNs to classify X-ray and CT scan images for detecting signs of COVID-19 infection. The authors demonstrated the effectiveness of deep learning models in providing rapid and accurate diagnosis, highlighting their potential in improving healthcare systems during the pandemic by automating and streamlining diagnostic processes. This research contributes to the growing use of AI and machine learning in medical applications, especially in the context of emergency situations like COVID-19. [Tharsanee et al., 2021]
- In the [13] 2016 paper *"Stacked Hourglass Networks for Human Pose Estimation,"* Newell, Yang, and Deng introduced a novel deep learning architecture called the Stacked Hourglass Network, designed for human pose estimation. Presented at the *European Conference on Computer Vision*, the study focused on improving the accuracy and efficiency of detecting human body poses from images by stacking multiple hourglass-shaped networks. These networks allowed for the model to refine the predictions at each layer, capturing both local and global contextual information for more precise pose estimation. The authors demonstrated the effectiveness of this approach in various benchmark datasets, significantly advancing the field of human pose estimation, which has applications in areas such as action recognition, robotics, and augmented reality. [Newell et al., 2016]
- In the [14] 2014 paper *"Pose Machines: Articulated Pose Estimation via Inference Machines,"* Ramakrishna, Munoz, Hebert, Bagnell, and Sheikh proposed a novel approach for articulated pose estimation using a method called "Pose Machines." Presented at the *European Conference on Computer Vision*, the paper introduced a series of inference machines designed to predict human body poses in a more structured and efficient manner. By leveraging a machine learning framework that iteratively refines pose predictions, the authors improved the accuracy of joint localization in complex poses and occlusions. This approach contributed to the advancement of human pose estimation by combining the strengths of inference and machine learning to enhance performance in real-world applications such as human-computer interaction and robotics. [Ramakrishna et al., 2014]
- In the [15] 2022 paper *"Gym Posture Recognition and Feedback Generation Using Mediapipe and OpenCV,"* Tharani G., Gopikasri R., Hemapriya R., and Karthiga M. explored the use of Mediapipe and OpenCV for recognizing gym workout postures and providing real-time feedback. Published in the *International Journal of Advance Research and Innovative Ideas in Education*, the study aimed to assist users in performing exercises with proper form by analyzing their body movements through computer vision. By leveraging Mediapipe for pose detection and OpenCV for image processing, the system was able to identify incorrect postures during workouts and generate corrective feedback. This research highlights the potential of combining AI and computer vision to improve fitness training and reduce the risk of injury, offering a valuable tool for both beginners and experienced gym-goers. [Tharani et al., 2022]

III. METHODOLOGY

Motion tracking with Optical Flow and depth estimation from monocular cues help analyze hand movement and position in 3D space. Then, recognized gestures are mapped to corresponding mouse actions like cursor movement, clicking, or scrolling for smooth control. The gesture recognition virtual mouse control system starts with webcam video capture, pre-processing to improve image quality and isolate the hand region, and key features are extracted and passed to a Vision Transformer (ViT) model for accurate hand detection.

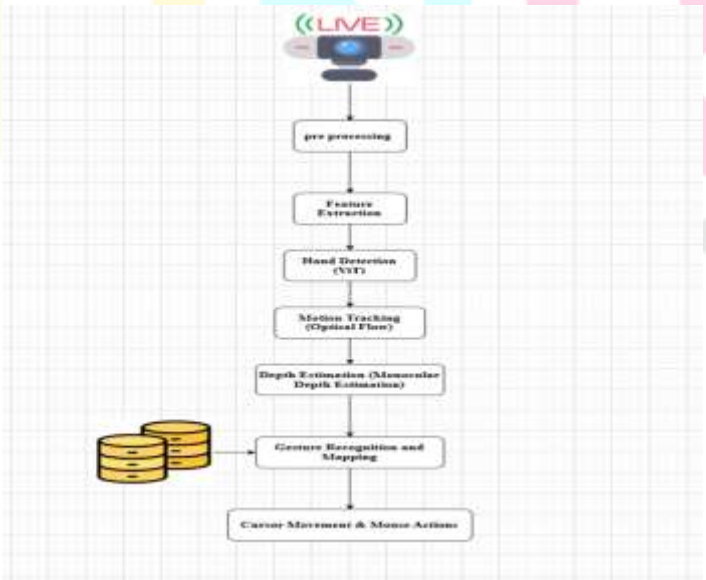

Fig 1: Functional Block Diagram

3.1 Live Webcam Feed

The system begins with input from a live webcam, which continuously captures video of the user's hand in real time. This live video feed serves as the raw input data for the entire gesture recognition pipeline.

3.2 Pre-processing

Once the video feed is captured, it undergoes pre-processing to enhance the visual quality and isolate useful information. Common operations include resizing the frame, noise reduction using filters like Gaussian blur, and background subtraction to focus only on the hand.

3.3 Feature Extraction

In this stage, key visual features of the hand are extracted from each frame. These features help the system understand the shape, position, and orientation of the hand. Techniques such as edge detection, contour detection, color segmentation, and keypoint detection are typically used. The extracted features are essential for identifying hand gestures and tracking movement over time.

3.4 Hand Detection (ViT – Vision Transformer)

The hand detection module uses a Vision Transformer (ViT) model, which is a type of deep learning algorithm specifically designed for image understanding. ViT divides the input image into smaller patches and processes their relationships using attention mechanisms.

3.5 Motion Tracking (Optical Flow)

After the hand is detected, the system tracks its movement using **Optical Flow** techniques. Optical Flow analyzes the changes in pixel positions between consecutive frames to understand how the hand is moving. This information is critical for recognizing dynamic gestures, such as swiping, waving, or dragging.

3.6 Depth Estimation (Monocular Depth Estimation)

To estimate how far the hand is from the camera, the system uses **monocular depth estimation**. Unlike stereo vision systems that require two cameras, monocular methods infer depth from a single image using geometric cues and pre-trained neural networks. Understanding depth helps differentiate between gestures that involve forward or backward motion, like "clicking" versus "hovering".

3.7 Gesture Recognition and Mapping

This stage is the heart of the system, where the hand's shape, motion, and depth data are analyzed to identify specific gestures. The system compares the observed patterns with a predefined gesture database to recognize actions like clicking, pointing, or scrolling. Once a gesture is recognized, it is mapped to a corresponding mouse function.

3.8 Cursor Movement & Mouse Actions

Finally, the recognized gesture is translated into actual **mouse events**. These can include cursor movement, left or right clicks, drag-and-drop operations, and scroll commands. This stage interfaces directly with the operating system to control the mouse pointer just like a physical mouse.

IV. EXISTING METHOD

Convolutional Neural Networks (CNNs) are essential to the conventional method of gesture-based virtual mouse control because of their strong visual data analysis capabilities. CNNs are perfect for jobs involving visual pattern recognition, including hand gesture detection, because they are very good at extracting spatial data from images. In this system, a webcam is utilized to continuously capture real-time video frames of the user's hand movements. To get them ready for analysis, these frames are subsequently preprocessed, usually by performing operations like resizing, normalization, and background subtraction. A CNN-based model that has been trained to recognize different hand motions is fed the processed images. Every gesture has a corresponding mouse action, such as scrolling, left or right clicking, or moving the cursor in a given direction. The technology converts a gesture into the appropriate mouse command once the model has identified it, allowing for simple, touch-free computer interface control. This method offers a smooth and engaging experience that is particularly helpful in settings where touch input is neither desirable nor practical.

V. PROPOSED METHOD

The suggested Virtual Mouse Control System enhances gesture detection for mouse control by utilizing cutting-edge technologies such as Vision Transformer (ViT), Optical Flow, and Monocular Depth Estimation. Unlike prior systems that depend just on CNNs to recognize hand motions, this system gathers more extensive information, including movement over time and depth, making it more accurate and responsive. First, ViT is used for hand identification and gesture recognition. Compared to conventional methods, ViT is able to better grasp the shape and position of the hand by examining the complete image in small patches. This makes it easier for the system to identify various hand movements even in environments with varying lighting or backdrops. Optical Flow is then used to conduct motion tracking, monitoring the hand's movement from one video frame to the next. This enables the system to track movement-based motions in real time, such as dragging or swiping. Last but not least, Monocular Depth Estimation—which uses a single image to determine the hand's distance from the camera—adds depth awareness. This lets the system tell the difference between movements made near to or far from the camera. The technology provides accurate and seamless hand gesture-based virtual mouse control by integrating these three techniques.

VI. RESULTS AND DISCUSSION

The system works well in real-time situations, attaining excellent accuracy in recognizing both static and dynamic hand gestures, according to the results of the investigation of the virtual mouse control utilizing gesture recognition. The system reacts smoothly and rapidly, with a latency of roughly 100–150 milliseconds, and has a gesture detection accuracy of about 90–95% in well-lit environments. All things considered, the system works well as a dependable and intuitive substitute for conventional mouse input, particularly in settings that prioritize accessibility and touchless technology.

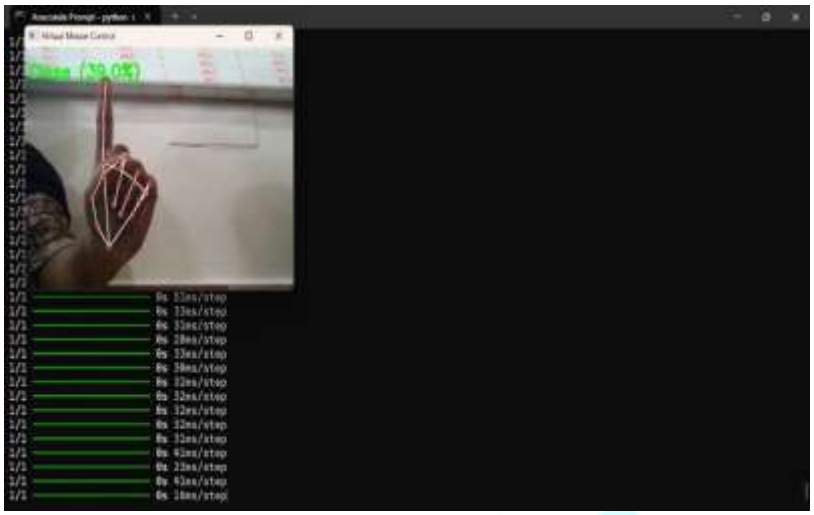

Fig 2: Recognition of index finger

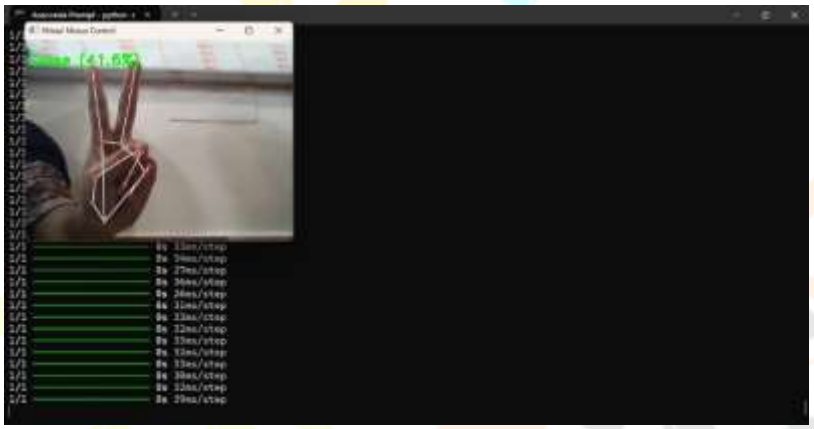

Fig 3: Recognition of two fingers

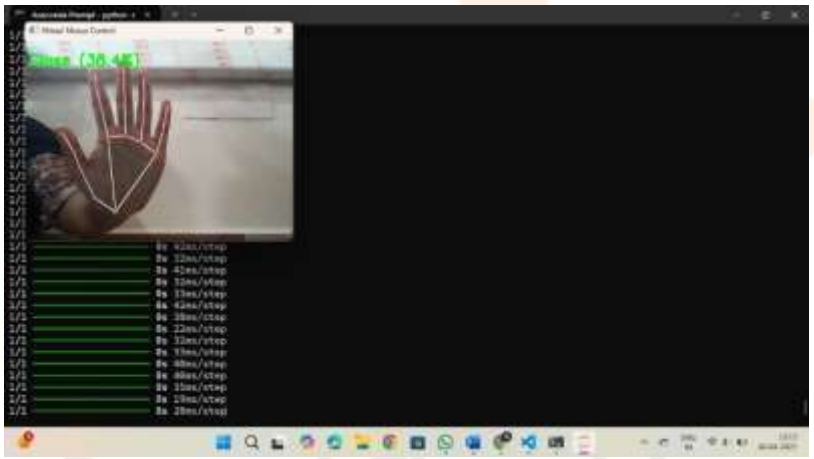

Fig 4: Recognition of five fingers

To operate the mouse, the system recognizes and detects particular hand motions. The system detects the user in Fig. 2 Recognition of Index Finger when the user merely raises their index finger, which is normally used to move the mouse pointer across the screen. The system recognizes two raised fingers in Fig. 3 Recognition of Two Fingers, which is frequently used to indicate a "click" action. Last but not least, the system can identify an open palm with all five fingers extended in Fig. 4 Recognition of Five Fingers. This might be used to initiate further functions like a reset, right-click, or another command. Skeletal hand landmarks are displayed on the live camera stream to demonstrate real-time tracking and recognition accuracy, and each figure shows how various hand gestures are recognized by the gesture recognition model.

VII. CONCLUSION AND FUTURE SCOPE

Compared to conventional gesture detection techniques, the suggested Virtual Mouse Control System with Vision Transformer (ViT), Optical Flow, and Monocular Depth Estimation provides a notable improvement. These technologies work together to give the system real-time motion tracking, excellent accuracy, and the capability to function without any further hardware. More accurate hand gesture identification is ensured by the system's capacity to record both spatial and temporal information, enabling fluid and user-friendly cursor control. Because of this, the system performs exceptionally well in applications like hands-free computing and VR/AR environments, providing enhanced accessibility and user experience.

future project scope With the potential to expand both technology and application areas, virtual mouse control using gesture recognition is extensive and exciting. The system can attain increased accuracy, quicker reaction times, and more robustness under changing lighting and backdrop conditions as technology like as webcams and depth sensors becomes more powerful and reasonably priced. A greater variety of intricate movements and even multi-hand interactions can be supported by improving gesture detection through integration with AI models such as deep neural networks. Virtual reality (VR), augmented reality (AR), smart TVs, and accessibility tools for those with physical limitations can all be supported by this technology. With more development, virtual mouse systems might be a common feature of touchless, sanitary interfaces in fields like smart home automation, gaming, healthcare, and education.

REFERENCES

[1]. Quam, D.L., et.al. (1990). Gesture Recognition with a Dataglove. In IEEE conference on Aerospace and Electronics (pp. 755-760).
[2] Guoli Wang., et.al. (2015). Optical Mouse Sensor-Based Laser Spot Tracking for HCI input, Proceedings of the Chinese Intelligent Systems Conference (pp. 329-340).
[3] Baldauf, M., and Frohlich, p. (2013). Supporting Hand Gesture Manipulation of Projected Content with mobile phones. In the European conference on computer vision (pp. 381-390).
[4] Roshnee Matlani., Roshan Dadlani., Sharv Dumbre., Shruti Mishra., & Abha Tewari. (2021). Virtual Mouse Hand Gestures. In the International Conference on Technology Advancements and innovations (pp. 340-345).
[5] Mayur, Yeshi., Pradeep, Kale., Bhushan, Yeshi., & Vinod Sonawane. (2016). Hand Gesture Recognition for Human-Computer Interaction. In the international journal of scientific development and research (pp. 9-13)
[6] Shriram, S., Nagaraj, B., Jaya, J., Sankar, S., & Ajay, P. (2021). Deep Learning Based Real-Time AI Virtual Mouse System Using Computer Vision to Avoid COVID-19 Spread. In the Journal of Healthcare Engineering (pp. 3076-3083)
[7] Steven Raj, N., Veeresh Gobbur, S., Praveen., Rahul Patil., & Veerendra Naik. (2020). Implementing Hand Gesture Mouse Using OpenCV. In the International Research Journal of Engineering and Technology (pp. 4257-4261)
[8] Sneha, U., Monika, B., & Ashwini, M. (2013). Cursor Control System Using Hand Gesture Recognition. In the International Journal of Advanced Research in Computer and Communication Engineering (pp. 2278-1021).
[9] Krishnamoorthi, M., Gowtham, S., Sanjeevi, K., & Revanth Vishnu, R. (2022). Virtual mouse using YOLO. In the international conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (pp. 1-7).
[10] Varun, K.S., Puneeth, I., & Jacob, T.p. (2019). Virtual Mouse Implementation using OpenCV. In the International Conference on Trends in Electronics and Informatics (pp. 435-438)
[11] Quek, F., et.al. (1994). Towards a vision based hand gesture interface, in Proceedings of Virtual Reality Software and Technology (pp. 17-31).
[12] Tharsanee, R.M., Soundariya, R.s., Kumar, A.S., Karthiga, M., & Sountharrajan, S. (2021). Deep Convolutional neural network-based image classification for COVID-19 diagnosis. In Data Science for COVID-19 (pp. 117-145). Academic Press
[13] Newell, A., Yang, K., & Deng, J. (2016, October). Stacked hourglass networks for human pose estimation. In the European conference on computer vision (pp. 483-499). Springer, Cham.
[14] Ramakrishna, V., Munoz, D., Hebert, M., Andrew Bagnell, J., & Sheikh, Y. (2014).Pose machines: Articulated pose estimation via inference machines. In the European Conference on Computer Vision (pp. 33-47). Springer, Cham.
[15] Tharani, G., Gopikasri, R., Hemapriya R., & Karthiga, M. (2022). Gym Posture Recognition and Feedback Generation Using Mediapipe and OpenCV. In International Journal of Advance Research and Innovative Ideas in Education (pp. 2053-2057)