AI Enabled Mouse Automated With Hand Gestures

PRABHAS KUMAR REDDY BHUMIREDDY
1st Student
Computer Science and Design
Vel Tech University
Chennai, India
iamprabhas21@gmail.com

POOJASREE GANDIKOTA
2nd Student
Computer Science and Design
Veltech Univeristy
Chennai, India
poojagandikota341@gmail.com

BHAIRAVESHWAR REDDY MACHIREDDY
3nd Student
Computer Science and Design
Veltech Univeristy
Chennai, India
m.bhairav3366@gmail.com

Mr. V. Karthikeyan
Assistant Professor
Computer Science and Engineering
Veltech Univeristy
Chennai, India
vkinfo19@gmail.com

***Abstract*—An innovative device referred to as a virtual hand gesture mouse seeks to enable individuals to operate computers using more ease than before without any conventional hardware. Advanced computer vision strategies are explored to identify movements of the human hand from a camera shot including shape analysis, object detection, and machine learning. Specific hand movements, including finger movements for moving the cursor, clicking, scrolling, and dragging objects, are commonly recognized by the system. The Hygienic, Simple and Convenient Touchless Technology Touchless technology is a hygienic, simple, and convenient way of approaching different activities that are executed on a daily basis. The touchless technology is characterized by the absence of any physical contact with the device that one is using to execute a certain application or activity (Wolfe, 2005). It is mostly used in places such as the hospitals where there are high chances of contracting various diseases through physical contact with the devices. According to researchers, the amount of bacteria in a single computer mouse is more than that of a human toilet (Brunn and Wolf, 2007). In order to reduce the chances of contracting diseases, the touchless technology is applied in places such as the hospitals to prevent any form of physical contact with the devices. The touchless technology is also used in the public places where touching is not allowed. In such places, the human beings are expected to perform various activities based on the touchless technology. The touchless technology is not only applied in public places but also in the case of virtual reality where one is expected to perform various activities in accordance to the touchless technology. The touchless technology enables the user to be able to use a device without necessarily having any physical contact with the device (Bain, 2011). The hand gesture control can also be used to aid the physically challenged users. The use of the touchless technology will assist the physically challenge users to be able to control the device.**

Keywords: Touchless technology, Virtual hand gesture mouse, Hand gesture control, Physical contact, Virtual reality, Computer vision

I. Introduction

The virtual mouse is exciting and new. It tops the list of the most natural ways to use a computer, and it doesn't need a mouse or keyboard. This invention uses computer vision, machine learning, and object detection to understand the hand movements of the user in real time. But there is a twist. Instead of a sensor, the computer camera is used, so it recognizes the user's gestures to move the cursor, click, drag an object, or scroll down the page.

The principal merit of this technology is its ability to offer a contactless solution. This trait is not only convenient, but it is also viable due to hygiene and security reasons. That makes it an excellent solution for places where hygiene is critical, such as health-care facilities and public kiosks. Besides, it can be an appropriate solution if a virtual reality setup is being implemented, and the traditional hardware cannot be used. The gesture recognition technology is getting quicker and smarter with the passage of time. This technology is acting as a bridge between the physical world and digital world. It offers a very smooth and user-friendly way to control the device. It doesn't matter whether the user is performing basic or complex operations hand gestures are capable of doing it. No complex set of gears are required for the completion of these tasks.

The essence of the setup is computer vision, which enhances and enriches the user experience, making it more streamlined and dynamic. As technology improves, it starts to read other gestures more and more accurately, and it adapts to the way people behave. It makes a huge difference not only for the average user, but also for people with physical disabilities, because it gives them more ways to interact with devices which at some point used to be hard for them to use.

It's also ideal in settings where touch is either impractical or unsafe—think hospitals, food prep areas, or busy public environments. And as it continues to improve, gesture control is becoming increasingly relevant in industries like gaming, robotics, and smart home tech, where hands-free operation is not just useful, but often necessary.

That said, the technology still faces a few hurdles. Gesture recognition systems can struggle with background noise, inconsistent lighting, or the wide variety of possible hand movements—all of which can impact accuracy and responsiveness. But with the help of a webcam or a built-in camera, the system captures video frames of your hand in action, processes the motion, and translates it into the appropriate mouse function.

In the big picture, while the traditional mouse has been a cornerstone of computing for decades, it has its limitations—especially in environments where space is tight or for users with mobility challenges. Gesture recognition aims to solve those problems, and the future looks bright for this innovative way to interact with technology.

II. Literature Review

Jayashree Katti and others (2021) [1] conducted a review to make human life easier and more straight forward. Hand movements can be used to control electrical appliances. This will also be helpful in the present Covid-19 pandemic. In situations where social distancing should be at the top, contactless appliances offer additional advantages in public areas. Hand gestures are the most common and Main mode of communication in the present world. They can make it easier to create safe and user-friendly interfaces for many applications. Depth and color cameras have been employed in several computer vision hand gesture algorithms identification, but effective classification of movements from various individuals remains a problem.

Akanksha Kulkarni, et al. (2021) [2] has explained the CNN model in his paper to describe gesture and control devices which utilize precise gestures. Gesture recognition technology will be simple, potent and secure under this covid-19 pandemic environment.

Abdullah Mujahid, et al. (2021) [3] research aimed to Observing gestures can assist persons with some types of disabilities when interacting socially with other individuals. This paper suggests a light-weight YOLO (You Only Look Once) v3 and DarkNet-53 convolutional gesture recognition neural networks without additional preprocessing, image filtration, and image enhancement. The suggested model achieved exhibits extremely high accuracy even in a complicated environment, and it was able to recognize gestures even in low-resolution. The model suggested was tested in the image mode annotated dataset of hand gestures in Pascal VOC and YOLO format. We had better results by using hand-based feature extraction and common hand gestures of our proposed YOLOv3 based model with accuracy, precision, recall, and F-1 score of 97.68, 94.88, 98.66 and 96.70Kinetic movements on a video.

Uma N M et al. (2021) [4] proposed their work on a new approach that recognizes the hand gesture by Indian Sign Language and express them into written and speech output. The system utilizes the vision-based method by which the facial and hand movement expressions are captured by web-camera and the multiple technologies. The images captured are processed with image processing and recognized with A neural network is utilized to recognize hand gestures with OpenCV. In addition, facial expressions can be transformed into textual representation and speech using hardware microcontroller based (raspberry pi). Sachin Devangan et al. (2020) [5] have discussed in-depth. In this research, different ways through which we can execute computer vision-based hand gesture natively on mobile phones are discussed. If hand gestures can be detected by mobile phones it can offer a novel form of communication using mobiles and in avoiding the problems associated with voice and touch interfaces. while improving the user experience in the process also supports. other gesture-based applications. The methods that we learn are vision-based because The majority of the smartphones possess a camera module. Also, it does not necessitate additional sensors or other equipment hardware. We compared the various available methods based on algorithms used and respective Accuracy.

In their 2019 survey, D. H. Park et al.[6] provide a comprehensive overview of how hand gesture recognition has evolved—from basic computer vision techniques to today's more advanced machine learning-based approaches. The paper highlights how gesture control is being used across various interactive systems, such as virtual mouse controls, gaming interfaces, and interactive screens. The authors also discuss how gesture recognition is making an impact in emerging technologies like augmented reality and wearable devices. Importantly, they point to key areas for future research, such as improving recognition accuracy, enhancing user experience, and expanding gesture libraries for broader applications. This survey serves as a valuable starting point for researchers and developers working in HCI.

A. Elmezain et al. (2020) [7] dives deeper into the techniques behind hand gesture recognition. The authors compare different methods—including template matching, machine learning, and deep learning—while outlining the strengths and limitations of each. They discuss real-world applications, from gesture-based authentication to sign language recognition and virtual mouse systems. The review also tackles common challenges like occlusion (when one part of the hand covers another), inconsistent lighting, and user differences. Notably, the authors suggest that future research should focus on multimodal gesture recognition systems—tools that can work reliably across various settings and for different users. The paper is especially useful for those looking to develop more accurate and adaptable HCI technologies.

S. M. Hassan et al. (2021) [8] look at how hand gesture recognition continues to shape the future of HCI. They explore modern approaches involving machine learning, computer vision, and sensor-based technologies, emphasizing how these innovations are making gesture control more precise and practical. This work complements the findings of earlier studies and reinforces the growing significance of touchless interaction.

III. PROPOSED METHODOLOGY

A step by stepapproach of setting Virtual Mouse system along with third party tool integration into a single flow to get the expected result involving a complete and accurate end to end approach in integration for the areas of interest falling under Linear Algebra with example usage of Mediapipe and OpenCV with beautifully integrated Open source libraries of python. The combination of these technologies allows the system to process imagesin real-time, recognize gestures, and track hands. It approaches sculpturing of each component by looking at its interaction in the entire process of making the system responsive and accurate which in turnleads to controlling the computer just by waving of hands.

A. Data Collection and Preprocessing:

- Preparing the Virtual Mouse system for training involves starting with the data. That is the creation of a mixed dataset. The data in this context, involves different hand gestures and the necessary movements. This is to enable the system to learn and perform a variety of functions correctly.
- OpenCV, we access features like Video Capture are utilized to gather perfect snapshot frames direct from the camera. The snapshots occur as an input and are processed through a preprocessing stage to clean the images and increase their clarity before being fed into the model. Open CV has helpful tools like Gaussian Blur, and Histogram Equalization that ensure the data is more clear and consistent for training.
- One of the most essential parts of every supervised machine learning project is annotations of the data. You need to label the gestures with accuracy so the model can learn from them. Although it can be a really hard task, PyAutoGUI is here to simplify and make the annotation process faster and more accurate.

B. Hand Tracking and Pose Estimation using Mediapipe:

- Mediapipe has a fundamental place in the world of real-time hand tracking and pose estimation. The mechanisms and algorithms for hand tracking are very simple and efficient. Hand tracking is one of the leading modules of Mediapipe today. You can detect hand poses on the human body with ease. It is the best module for real-time interactive applications like virtual mouse
- Mediapipe uses techniques, graph models and convolutional neural networks (CNNs) to identify hand landmarks. It helps to pinpoint the exact positions of the hand. Landmarks enable the system to establish the hand's position, orientation and how it moves through space.
- It is possible to use Mediapipe to precisely track fingers. Most remarkable about this technology is that it recognizes a position and direction of each individual finger separately. Such details are critical to recognizing complex gestures and translating them into meaningful actions, which makes this technology especially useful for gesture control systems.

C. Gesture Recognition using Python Libraries:

- We use different Python libraries to create advanced custom algorithms for precise recognition of gestures. The developed algorithms are based on the methods of image processing: pattern recognition and feature extraction. They allow distinguishing different hand gestures.
- By designing these tailored algorithms, we ensure the system can label and identify a wide range of gestures effectively. Libraries like NumPy and SciPy are especially useful here—they help us perform the numerical and statistical computations needed to fine-tune the recognition model based on specific requirements.
- Once a gesture is detected, PyAutoGUI steps in to translate it into a real action on the screen—like moving the cursor or clicking—making the interaction feel smooth and responsive. It plays a crucial role in bridging the gap between gesture recognition and on-screen control.

IV. ALGORITHMS AND TOOLS USED

For the detection of fingers and hands, we utilized one of the helpful open-source libraries Mediapipe, it is a framework of cross-platform capabilities type which was developed by Google and OpenCV to achieve some Computer Vision-related operations. The algorithm utilizes machine learning-related concepts for hand gesture detection and tracking their movement.

A. Mediapipe:

"Mediapipe" is a sweeping and extensive open-source platform that has been carefully created by Google, and it is dedicated to an extensive variety of concerns with computer vision technology such as features including real-time pose estimation and hand tracking. For the case of our particular project, "Mediapipe" was a critical component and a key element that helped us in achieving very accurate and precise hand tracking, which is actually a critical and essential part of the overall functionality of the Virtual Mouse system that we are developing.

1) Hand Tracking Module::

- Through accepting input images or video frames, the "Mediapipe" software offers a highly specialized and advanced hand tracking module with the special ability of real-time detection and tracking of hands. This sophisticated module identifies and detects hands accurately by utilizing the strength of machine learning models that have been thoroughly trained on a wide range of large and varied datasets. Consequently, a system that can track and follow the user's hand in real-time throughout the interaction is provided.
- "Mediapipe" hand tracking technology is able to do more than simply identify whether or not a hand is present. Rather, it does much more by indicating significant landmarks and significant details on the hand itself. The system can effectively chart and monitor the complex hand movements due to this advanced ability. As the system is capable of monitoring the relative positions of distinct fingers to the palm, this capability is particularly useful in monitoring a high number of gestures. An accurate

and complete knowledge of different hand positions and configurations defines this overall approach.

2) Finger Tracking and Pose Estimation::

- Mediapipe also comes equipped with a very impressive finger tracking feature and very high accuracy for hand pose estimation. The sophisticated system is meant to precisely detect and identify every position and orientation of each finger. This allows the Virtual Mouse system to easily identify and establish complex hand motions as well as motions that have specific finger alignment configurations in various groups.
- The pose estimation functionality of the framework is especially useful in determining the orientation of the hand in three dimensions. The system requires this information so that it can accurately interpret hand movement and map it to actions on the computer interface. To interpret hand movement correctly and translate it into actions on the computer interface, this information is crucial.

B. OpenCV:

Library is an extremely robust and efficient open-source library that has been widely employed in numerous domains for computer vision and image processing applications.

OpenCV was significantly depended upon for our project and was the foundation for most of the most essential features and functionality employed to create the Virtual Mouse system. OpenCV was used by us for implementing the features and functions wherever necessary as well.

1) Data Preprocessing and Collection::

- OpenCV is one of the necessary and basic data collection and preprocessing tools. It allows easy grabbing of picture frames directly from cameras, and it is an important first step to achieve the large dataset required to train the innovative Virtual Mouse system.
- OpenCV is a robust library that uses a simple yet highly effective mechanism of collecting visual data, thereby ensuring that the resulting training set is not only comprehensive but also highly representative of a vast array of real-world situations encountered in daily life.
- Besides its basic functionality, OpenCV provides a rich collection of functions that are specialized in the complicated task of image preprocessing. These are image improvement to emphasize some features, contrast adjustment to enhance the quality of the images visually, and noise reduction to eliminate unwanted disturbance. These are the starting preprocessing steps that are necessary to ensure better-quality data collection, i.e., there must be minimal to zero noise or disturbance that may interfere with further analysis.

2) Hand Detection::

- The first method of hand detection utilizes OpenCV's specially implemented Haar Cascade Classifier. A critical element to the precise calculation of whether or not hands appear in different image frames is the classifier, which is a powerful object detection tool within machine learning. Through accurate detection of the precise identification of the exact features of the human hand, the Haar Cascade Classifier enables the system to detect a specific area, or region of interest (ROI), that in this instance is the hand so that the system can further process it in detail.
- Haar Cascade Classifier is a fundamental and essential component as a basic building block of the Virtual Mouse system's data collection process. It is the underlying foundation on which everything else in gesture recognition and hand tracking is established and built upon.

3) Real-time Image Processing::

- OpenCV is unmatched in real-time image processing expertise, and therefore it is a highly valuable as well as enormously helpful tool for processing video frames and camera input in real time. Such remarkable capability is of utmost importance to the Virtual Mouse system since it needs to display smooth, continuous, and highly responsive interaction that is essential to usability as well as user experience.
- Real-time image processing offered by the library helps the system maintain pace with hand movements and gestures in real-time. By providing a stable and consistent platform for image analysis and manipulation in detail, OpenCV is thus responsible for assisting the system in identifying various gestures and converting them into corresponding onscreen activities with little lag in between.

V. DIAGRAMS

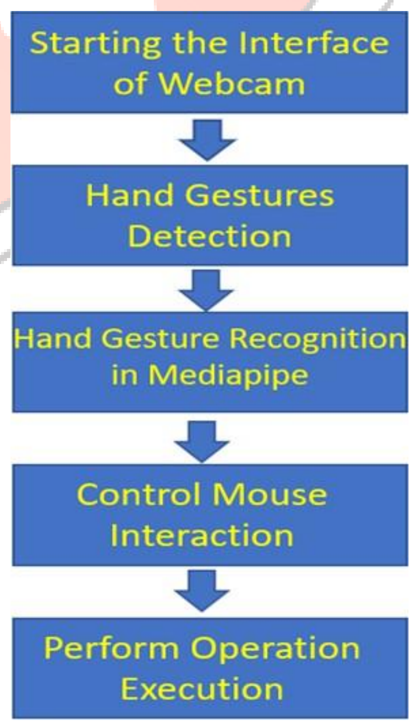

Fig. 1. Flow graph for Hand Gestures Recognition

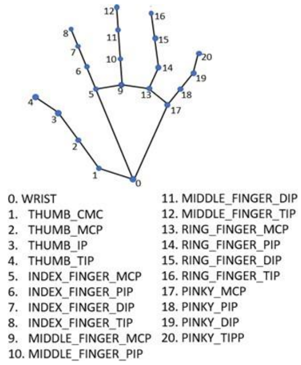

Fig. 2. Model Graph of Mediapipe

Fig. 3. Work Flow of Hand Gestures Recognition

VI. RESULT

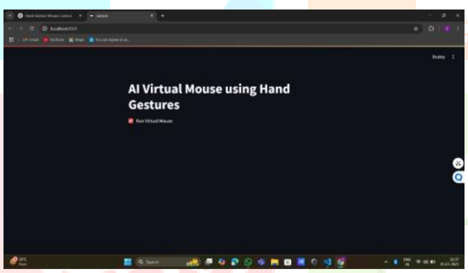

Fig. 4. Window to start virtual mouse

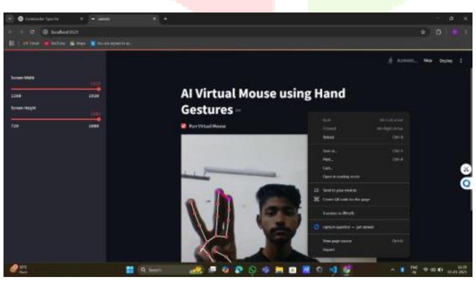

Fig. 5. Mouse Operation-Right Click

VII. CONCLUSION

In summary, virtual hand-gesture-controlled mouse is a revolutionary advancement in human communication with digital technology. Through the application of computer vision

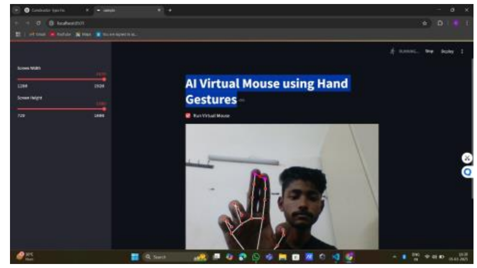

Fig. 6. Mouse Operation-Right Click

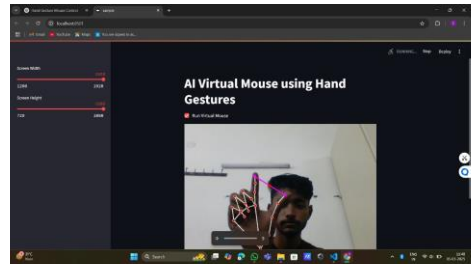

Fig. 7. Mouse Operation-Volume Up

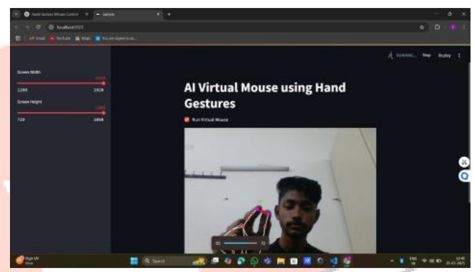

Fig. 8. Mouse Operation-Volume Down

technology and machine learning, the interface is a spontaneous, natural, and sanitary complement to traditional inputs, with huge increases in functionality, user accessibility, and end-to-end productivity. Since advancement in forthcoming years will elevate a broader array of gestures, combine more accuracy with less lag, and amplify software realization, users can expect broad applicability beyond its current applications in medical settings, public spaces, virtual worlds, and robots. On account of favorable prospects, then, the virtual user interface with gestures is destined to transform human-technology interaction by naturalizing, democratizing, and enveloping computing.

REFERENCES

[1] C. Cao, Y. Gao, and S. Zhang, ”Real-Time Hand Gesture Detection and Recognition Using Convolutional Neural Networks,” IEEE Transactions on Image Processing, vol. 26, no. 12, pp. 5946-5956, 2017.
[2] S. S. Kulkarni and S. M. Patil conducted a survey titled ”Hand Gesture Recognition for Human-Computer Interaction” and published in International Journal of Computer Applications, vol. 179, iss. 5, pp. 9-16 in 2018.
[3] D. H. Park and J. H. Choi, ”A Survey on Hand Gesture Recognition Systems,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 3, pp. 362-370, 2019.

[4] Elmezain and M. Hussain, "A Review of Hand Gesture Recognition Techniques for Human-Computer Interaction," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 11, pp. 4649-4671, 2020.S. M. Hassan and M. A. Hannan, "Hand Gesture Recognition: A Literature Review," International Journal of Human-Computer Interaction, vol. 37, no. 4, pp. 308-328, 2021.

[5] N. Sharma and A. Sharma, "A Survey on Hand Gesture Recognition Techniques," International Journal of Computer Applications, vol. 170, no. 6, pp. 1-6, 2017.

[6] Y. Kim and J. Choi, "A Deep Learning Approach to Hand Gesture Recognition for Human-Computer Interaction," Sensors, vol. 18, no. 11, p. 3766, 2018.

[7] M. Yang, Y. Du, and X. Zhang, "Hand Gesture Recognition for Human-Computer Interaction Based on Convolutional Neural Networks," International Journal of Advanced Computer Science and Applications, vol. 9, no. 10, pp. 405-411, 2018.

[8] Wang and Y. Yang, "Hand Gesture Recognition System for Human-Computer Interaction Based on Machine Learning," IEEE Access, vol. 8, pp. 78153-78162, 2020.

[9] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," in Advances in Neural Information Processing Systems, 2014, pp. 568-576.

[10] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," in Advances in Neural Information Processing Systems, 2014, pp. 568-576.

[11] S. U. Dudhane, "Cursor control system using hand gesture recognition, IJARCCE, vol. 2, no. 5, 2013.".

[12] P. Nandhini, J. Jaya, and J. George, "Computer vision system for food quality evaluation a review, in Proceedings of the 2013 International Conference on Current Trends in Engineering and Technology (ICCTET), pp. 85–87, Coimbatore, India, July 2013. ISSN: 2455-2631 November 2024 IJSDR — Volume 9 Issue 11 IJSDR2411058 International Journal of Scientific Development and Research (IJSDR) www.ijsdr.org 456.

[13] Dynamic hand gesture recognition system based on hidden Markov model using OpenCV Published in 2013 3rd IEEE International Advance Computing Conference (IACC), Dr. Rajat Shrivastava.

[14] Hand Gesture Recognition using Depth Images: A Review Jesus Suarez and Robin R. Murphy, Member, IEEE, 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. September 9-13, 2012, Paris, France.

[15] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov, "Realtime computer vision with opencv, Queue, Vol. 10, no. 4, pp.40-56, 2012".