

Augmented Virtual Mouse System with Enhanced Gesture Recognition

Gopinadh Kotari, Keerthi Jangam, Kolluri Sri Sai, Mamillapalli Aparna

Department of ECE, R.V.R. & J.C. College of Engineering

Abstract - This paper presents the development of a contactless system designed to serve as an alternative to the traditional physical input device, such as a computer mouse or touchpad, utilized in human-computer interaction. The proposed system aims to offer a more convenient and hygienic way of interacting with computer devices by using hand gestures instead of physical contact with the mouse system or machine interface. The system has potential applications in various fields including healthcare, public interfaces, and gaming. Existing gesture-controlled input systems lack a comprehensive touchless and accurate system capable of performing almost every operation of a physical mouse. The proposed approach involves recognising the hand and its gestures. Hand tracking and dynamic gesture recognition are incorporated to ensure a smooth and hassle-free user experience during interaction. The efficient Mediapipe framework is used to detect and track hand movements from video frames captured through a computing device Camera. Computer vision algorithms are employed for gesture recognition within each frame, along with Python libraries for processing gestures. Sufficient gestures are added into the system, enabling it to execute a wide range of mouse functions with substantial precision metrics. This paper provides details regarding the development and overall performance of the system.

Key Words: *hands, mouse, contactless, gesture, OpenCV, mediapipe*

1. INTRODUCTION

Communication methods have undergone significant changes in the field of computers. The mouse was used for a long time to click on things and move around on screens. Nowadays, there is a growing interest in finding new ways to interact with computers without touching them. This paper explores the idea of using hand movements instead of a mouse to control computers. The proposed device was made possible through the use of computer vision and deep learning techniques. The developed system is able to perform all the functions of a normal physical mouse with significant accuracy.

After reviewing various methodologies, it was found that the mediapipe framework is known for achieving high accuracy in object detection and tracking. The hands module of this framework has been utilized for hand gesture recognition and tracking. Also, python libraries like opencv, and numpy were used for image processing, gesture recognition. Finally, pyautogui library is implemented for mapping of gesture into the actions. This paper includes essential information about the methodology for implementing the proposed system along with the gestures utilized within, and concludes by discussing its performance metrics.

2. LITERATURE REVIEW

Previous research on gesture-based systems and methodology implement in those systems has been studied as part of the literature review. This section presents an overview of the findings from the study.

[1] demonstrated gesture recognition through the use of colour distinction. The system is operated using coloured caps and five distinct hand gestures have been implemented to correspond to five different operations. A series of three images is used for each mouse operation, capturing different stages of the gesture recognition process. The system produces inconsistent results with different brightness levels. The accuracy of the system dropped from above 90% to below 75% when the brightness levels are outside the range of 400 – 750 lux. The identified problem arises from the use of colour caps. This issue is addressed in the system discussed, as it only requires bare hands, which are less affected by varying brightness levels.

[2] a gesture-controlled mouse has been created, with all the system's gestures being static. Each finger is mapped to a different mouse operation, and combining various finger indications triggers specific operations. To enhance user interaction with the computer, dynamic gesture recognition could be incorporated into the design. The proposed design primarily focuses on supporting dynamic gestures, aiming to provide users with a seamless experience.

[3] the Histogram of oriented gradient with relative distance features was implemented through Validation Hidden Markov Model, Convolutional Neural Network, Principal Component Analysis, and Support Vector Machine. The identification accuracy range for the signer-dependent variable varied from 69 to 98 percent, averaging at 88.8percent. For the signer-independent variable, the identification accuracy ranged from 48 percent to 97.2 percent, with an average recognition accuracy of 78.2 percent as per the selected study findings. Limited progress in continuous gesture recognition indicates a need for further development in order to create a vision-based gesture recognition system that is both practical and effective an improvement over current standards within this field. To achieve higher precision levels requires an active system rather than just a single camera but entails increased processing time and necessitates high-performance computing devices which are impractical for public systems.

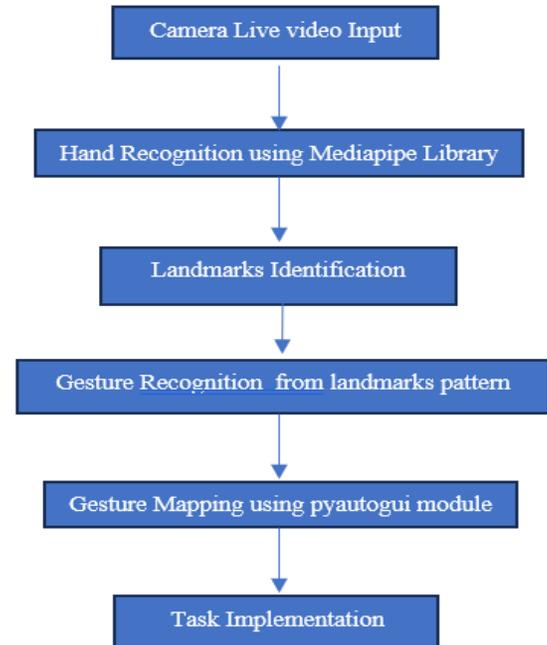
[4] have implemented a virtual mouse using the Mediapipe framework, which has been trained with ample datasets to achieve high accuracy. The Hand Landmark Identification Hands module of the Mediapipe framework is employed for this purpose, particularly designed for public computing devices. Although the system provides limited gestures, there is scope for incorporating many more gestures so that every operation performed by a physical mouse can be easily carried out by the virtual mouse. It successfully executes click and drag operations with an accuracy of 90 percent across various tasks.

[5] This research implemented an NUI for gesture recognition without relying on machine learning. The design involved a virtual monitor-based hand-mouse interface that captures the user's physical features through Kinect. Compared to traditional machine-learning-based gesture recognition, this approach was relatively simple to implement. Experiments were conducted to measure the intuitiveness and accuracy of the proposed hand-mouse interface which yielded excellent results; indicating its potential applicability in mouse control systems operated by hands. Additionally, it is anticipated that this system will be part of emerging new interfaces alongside other techniques.

3. METHODOLOGY

The project's operation consists of four main steps: capturing frames, detecting hand gestures, recognizing gestures, and mapping the gestures to specific operations. Overview of each step is as follows.

Fig-1: Flow diagram



1) *Capturing frames*: The camera of computer system captures live video streams. The cv2 module from the Python library is employed to access the camera and obtain input image frames from the video feed. At the same time, the captured frame is converted from the BGR image to the RGB image for easier processing. Additionally, the live video stream is displayed to enhance the user experience. Once the frames are captured, the next step is to locate hand within those frames.

2) *Hand landmarks detection*: In this step, the system searches for the presence of hand(s) in the captured frames using the mediapipe framework. The hands module in the mediapipe library contains pre-trained ML models and algorithms that generate 21 3-Dimensional coordinates representing the detected hand. These coordinates change as the hand moves around the frame. The details regarding the 21 landmarks are provided in Fig-2.

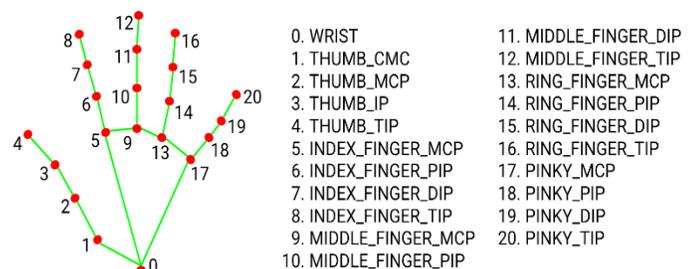


Fig-2: Hand landmarks

This framework is able to detect very accurately because it has been trained with a huge number of datasets. The proposed system is designed to operate with an 80 percent detection coefficient and a 75 percent tracking coefficient, thereby ensuring high throughput throughout the entire operation.

3) *Gesture Recognition*: The coordinates for each landmark are obtained when a hand is recognized in a particular image frame. Each set of coordinates contains the terms x, y, and z. Here, x and y indicate horizontal and vertical distances, while z indicates the depth of the pixel with respect to the central landmark, i.e., the wrist.

Distance calculation methods, such as Euclidean and Manhattan distances, are employed within these coordinates to analyse spatial relationships between different landmarks and accurately track hand gestures. This analysis helps in identifying the indicating gesture based on the observed relationships. The system includes a total of 7 gestures, including the neutral gesture. The gestures and their corresponding operations are mentioned in the Table-1.

4) *Gesture mapping*: The coordinates for each landmark are obtained when a hand is recognized in a particular image frame. Each set of coordinates contains the terms x, y, and z.

Here, x and y indicate horizontal and vertical distances, while z indicates the depth of the pixel with respect to the central landmark, i.e., the wrist. Distance calculation methods, such as Euclidean and Manhattan distances, are employed within these coordinates to analyse spatial relationships between different landmarks and accurately track hand gestures.

This analysis helps in identifying the indicating gesture based on the observed relationships. The system includes a total of 7 gestures, including the neutral gesture. The gestures and their corresponding operations are mentioned in the Fig-3.

3. WORK DONE

In the proposed system a total of seven gestures are included system. Except neutral and cursor movement gestures remaining all are dynamic gestures. Hand gesture images of each operation is provided further sections.

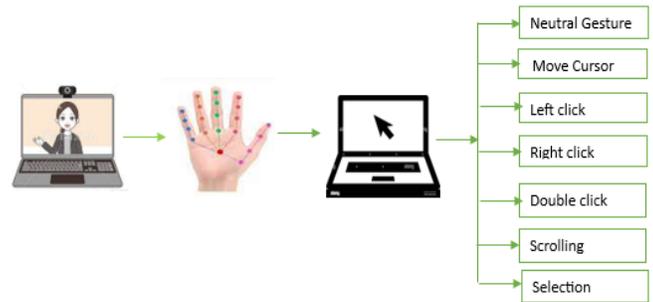


Fig-3: Operations of the proposed system

There are some key features regarding the system that are described as follows:

a) *Dynamic gesture recognition*: Recognition of dynamic hand gestures in virtual mouse systems controlled by hand gestures depends on ongoing tracking and extraction of features to understand hand movements. These motions are divided into specific gestures and categorized using machine learning techniques. Users receive immediate feedback, with adaptive learning methods enhancing recognition precision over time. Robustness is ensured through noise filtering methods, while user calibration customizes recognition for individual users. Together, these components improve the system's precision, responsiveness, and overall user satisfaction.

Fig-4: Gestures



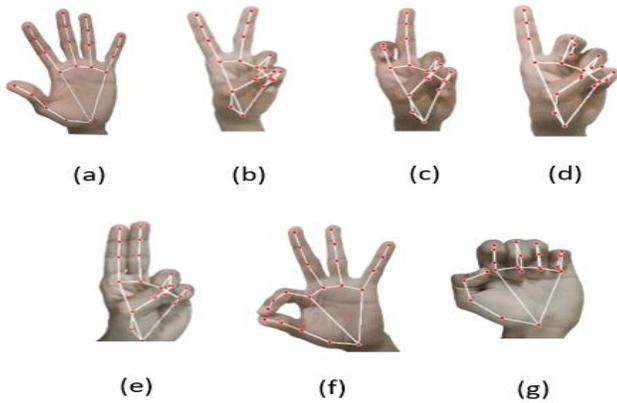


Fig-5: Example for dynamic gesture recognition

b) *Cursor control and tracking*: The system allows cursor control through the use of gesture (b) shown in Fig-5. By placing the hand in that position, the system executes cursor movement operations. The on-screen cursor will move based on the relative difference obtained when the 3D coordinates of landmarks are changed, providing highly accurate gesture recognition and a smooth tracking experience

4. RESULTS AND ANALYSIS

Details about the performance of the proposed system are discussed in this section to assess this, the accuracy percentage of gesture recognition is computed. Each gesture is recorded 200 times and then the accuracy percentage is determined. The results are outlined in the table. The achieved accuracy percentages for each gesture are significantly higher compared to results provided in papers that are mentioned as part of the Literature review [1] & [3].

Table-1: Gestures

Operation	Gesture Number as per Fig-4	Gesture Detection Accuracy
Neutral	(a)	99.5%
Cursor Movement	(b)	99%
Left click	(c)	92%
Right click	(d)	92%
Double click	(e)	84%
Scrolling	(f)	76.5%
Selection	(h)	80%

Gestures (a), (b), (c) & (d) depicted in Fig-5 have achieved an accuracy of over 90 percent, while gestures (e), (f), & (g) exhibited a relatively lower level of accuracy due to the complexity in gesture positioning. Nevertheless, overall system performance signifies notable progress in gestured-controlled input devices.

The performance analysis of the proposed system also involved assessing its accuracy under different lighting conditions. The system's ability to perform consistently across varying levels of brightness was observed, and it is capable of recognizing and performing operations even with changing light levels.

The mediapipe framework is a comprehensive library for detecting and tracking various real-time objects. Incorporating such library modules into this system was a key aspect of its design. The detection accuracy is remarkable, and the results are presented in Table-1

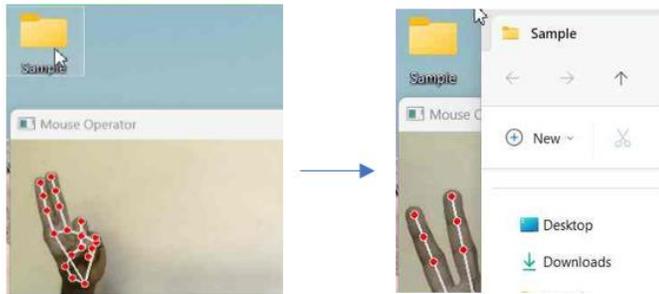


Fig-6: Implementation of double click gesture

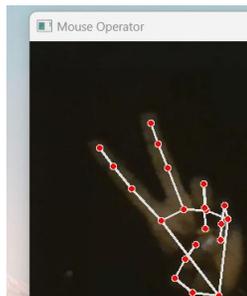


Fig-7: Low light operation of the system

5. Conclusion

A technique is proposed for controlling the on-screen cursor without relying on any external physical device. The system utilizes a live camera to control the mouse pointer and perform its tasks. All mouse operations can be performed using just a single hand, providing a more convenient and hygienic way for users to interact with systems or computing devices. This system is particularly useful in public computing environments such as boarding pass printers at airports or e-book stations at libraries. Overall, the development of these types of systems is helpful in meeting ongoing and upcoming technological advancements.

References

1. V. V. Reddy, T. Dhyanchand, G. V. Krishna and S. Maheshwaram, "Virtual Mouse Control Using colored Finger Tips and Hand Gesture Recognition," 2020 IEEE-HYDCON, Hyderabad, India, 2020, pp. 1-5, doi: 10.1109/HYDCON48903.2020.9242677.
2. CHAVALI, E Sankar. (2023). Virtual Mouse Using Hand Gesture. 7. 7. 10.55041/IJSREM21501.
3. N. Mohamed, M. B. Mustafa and N. Jomhari, "A Review of the Hand Gesture Recognition System: Current Progress and Future Directions," in IEEE Access, vol. 9, pp. 157422-157436, 2021, doi: 10.1109/ACCESS.2021.3129650.

10.1109/ACCESS.2021.3129650.

4. A. Mall, "An Improved Virtual Mouse for Public Computing Systems using Computer Vision," 2022 *International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, Bengaluru, India, 2022, pp. 107-112, doi: 10.1109/CENTCON56610.2022.10051583.

5. C. Jeon, O. -J. Kwon, D. Shin and D. Shin, "Hand-Mouse Interface Using Virtual Monitor Concept for Natural Interaction," in *IEEE Access*, vol. 5, pp. 25181-25188, 2017, doi: 10.1109/ACCESS.2017.2768405.

6. Lugaresi, Camillo, et al. "MediaPipe: A Framework for Perceiving and Processing Reality." *Proceedings of the Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR)*, 2019.

7. Hemalatha, M. & Sreeja, V. & Aswathi, S.. (2024). Hand Gesture Controlled Virtual Mouse. *IJARCCCE*. 13. 10.17148/IJARCCCE.2024.133143.