International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 12 | Dec 2025 www.irjet.net p-ISSN: 2395-0072

Gesture Control System

Avula Swarna Rupa[1], Pantham Pallavi[2], Marri Bhavya Sri[3] , Ambati Tharun Reddy [4] ,Mrs. Swathi Sugur[5]

[1, 2, 3, 4] *Department of CSE(Data Science), AVN Institute of Engineering and Technology*
Under the guidance of
[5] *Asst. Professor, Department of CSE(Data Science), AVN Institute of Engineering and Technology*

Abstract*—Human–computer interaction has been mainly dependent on physical input devices like keyboards and mice. However, these devices may impose limitations on usability in certain environments and applications. With the development of computer vision techniques, touchless gesture-based interaction has become a natural and highly intuitive alternative. Gesture control systems are designed to understand human hand movements and then convert them into interacting commands for digital systems.*

The article describes a camera-based gesture control method that is capable of allowing a user to interact with a computer by utilizing a set of fixed hand gestures. Through the real-time detection of hand gestures, the system consequently carries out a multitude of interaction tasks including navigation, executing control operations, and issuing system-level commands. Hand gestures as an input mode not only promote accessibility but also represent a more instinctive way of interaction which is especially advantageous in cases when physical touch is either not preferred or not feasible.

Gesture-based interaction systems hold the promise of fundamentally changing assistive technologies, smart environments, presentation control, and interactive systems. The research paper thus puts the spotlight on gesture recognition being a crucial facilitator of contemporary human–computer interaction and points out its utility in the development of natural, efficient, and contactless user interfaces.

***Index Terms*— Gesture Recognition, Hand Tracking, Human–Computer Interaction, Touch less Interaction, Vision-Based Interaction**

I. INTRODUCTION

Human–computer interaction is a core element of how users relate to digital systems. However, the typically used methods of interaction, like keyboards and mice, have not changed significantly over the years and might sometimes even be unnatural and less efficient ways of controlling, especially in situations where touchless operation is required. As the need for more intuitive and contact-free interfaces grows, gesture-based interaction becomes a leading alternative which helps users communicate with computers in a natural way by hand movements.

Gesture control systems are designed to understand human hand gestures taken visually and thus turning them into system commands which can be understood by machines. Besides traditional input devices, people using gesture-based interfaces can carry out their work even without any physical contact making it possible to improve usability and accessibility. These systems get rinsed benefits, for example, in situations such as presentations, assistive technologies, smart environments, interactive applications where it is important to have very simple interaction and the user must have complete freedom of movement.

The technological progress in the field of computer vision and the development of hand tracking in real time have helped gesture recognition systems in gaining more and more accuracy and reliability. Thanks to these developments, the system can precisely locate the hands, identify the shapes of the fingers, and even the user's gesture pattern thus the users will be able to convey their intent efficiently which the systems react. Gesture-based systems, through the identification of predefined gestures and their association with control actions, deliver an interaction experience that is more natural and appealing to the users as compared to using traditional interfaces.

The ever-increasing curiosity about touchless interaction explains why it is necessary to create highly efficient and flexible gesture control systems. Gesture-based methods which allow natural and intuitive interaction with digital devices are the ones

that lead human–computer interaction to the next level. At the same time, these methods facilitate the development of more intelligent and easily accessible computing environments.

II. LITERATURE SURVEY

Gesture-based human-computer interaction (HCI) is an area that keeps changing to suit the user's needs for more natural and intuitive ways of getting input. Traditional methods of interaction/application are limited to physical devices, whereas, the vision-based methods give the users an opportunity to control the devices through body movement observation without physically touching the devices. Mackenzie has thoroughly analyzed the human-computer interaction (HCI) principles and among the points, he has brought out the main features of interactive systems including usability, intuitiveness, and efficiency. These are the same features that are at the core of gesture-based interaction systems.

Initially, the progress in vision-based ways of interaction was greatly influenced by the depth and motion-sensing technologies. Zhang [3] investigated how the depth-sensing tools like the Microsoft Kinect have revolutionized human motion tracking for gesture-driven applications. Inspired by this trend, Shotton et al. [4] developed state-of-the-art human pose recognition algorithms that operate in real-time, thus accurately detecting body and hand movements and enabling the creation of vision-based interactive systems.

Hands tracking and recognition of gestures are topics that have been deeply scrutinized by utilizing the traditional computer vision methods. Mittal et al. [5] came up with a strategy to recognize hands by integrating various visual cues. They identified the difficulties that arise from the fact that hands may have different shapes and the working condition may slightly vary like in the case of the light and background. Poppe [6] conducted an in-depth review of the literature on human action recognition based on the visual channel. He has classified different methods and singled out issues such as real-time performance and robustness limitations. These issues are still a challenge for the gesture control systems.

The current breakthroughs aim at creating perception pipelines which are not only extremely precise but at the same time they are capable of real-time operation. Lugaresi et al. [1] came up with a perception pipeline, which made it possible to detect hand landmarks and thus perform accurate gesture recognition using a normal camera. The utility of such perception pipelines is quite evident, considering that they still achieve a high level of accuracy but at the same time their computational cost is greatly diminished. Thus, such pipelines are the perfect fit for human-computer interaction applications that require real-time performance.

The usability together with user experience determine the success or failure of gesture-based systems in a big way. Greenberg and Buxton [7] elaborated on the idea that the usability evaluation methods for interactive systems must be appropriate. They pointed out that the design of a system should consider all functionalities but simultaneously a user’s comfort and ease of learning should be taken into account as well. The main goal of these considerations is to ensure that gesture-based interfaces are a convenient and friendly tool for users.

There should definitely be additional options to the gestures recognition that facilitate interaction clarity, which is multimodal feedback. There is an emerging trend of additionally introducing interaction modalities such as speech and audio feedback. To give an example, Taylor et al. [10] came up with speech synthesis systems that can provide auditory feedback. On the other hand, Young et al. [8] established the basic theories of speech processing that lead to the development of the voice-based interaction. Interestingly, Jain et al. [9] elaborated on the biometrics characteristics that are connected to the understanding of both gesture and pattern recognition as behavioral identifiers.

In general, the literature has shown that there have been substantial developments in the field of vision-based gesture recognition, HCI, and multimodal feedback systems. Despite that, some of these works are focusing on the singular components, for example, hand detection or pose estimation and do not provide the big picture. This, therefore, indicates that the field is still yearning for the kind of holistic systems that will sweep not only the gesture recognition but also the usability issues and feedback mechanisms so as to fully enable effective and natural touch-free interaction.

III. PROPOSED SYSTEM

The system that is being proposed is a gesture recognition system based on a camera and computer vision technology that allows users to perform various tasks on a computer just by using their hands in the air without physically touching the computer or mouse or keyboard. The concept is to capture spontaneous hand movements through the camera and translate them into meaningful commands that get executed on the system which, in turn, makes it less dependent on the conventional standby input devices like keyboard and mouse. The whole idea of this system is to support the user interaction with the computer by hand gestures that it will be possible to increase the accessibility, usability, and the overall user experience at any kind of computing environment.

The system architecture includes the user interface and control layer and the processing and decision-making layer, which unitedly work to provide uninterrupted and dependable functionality. The user interface and control layer constantly grabs live video feed and simultaneously shows the results of the image processing in a manner of recognized gestures and hand landmarks to the user. Therefore, the user has not only been helped to perform the gestures more visually but also to quickly adapt to the system which leads to a more pleasant user experience.

A system control layer is the brain of the system behind the scenes. It is in charge of real-time hand landmark detection, gesture classification, mode management, and execution of system-level actions.

By processing the hand landmarks extracted, the system recognizes a set of predefined gesture patterns and carries out the corresponding control operation like moving the cursor, click actions ticking, scrolling, controlling presentation flow, launching applications, and adjusting system settings. To avoid unintended gesture activation, the system control layer also has the capability of managing the different modes of operation depending on the context of the actions which are going to be carried out.

For the purpose of further providing a good experience and diversification of the users, the system control layer facilitates not only multi-user profiles but also custom gesture calibration. Through maintaining the user-specific gesture setups and the preferences, the system control layer makes it possible for various users to interact with the system in accordance with their unique requirements. This functionality has given the system a higher level of reliability because it is, to some extent, capable of handling changes in hand dimension, gesture manner, and user behavior.

Moreover, the system implements some other features such as voice feedback controlled by the system control layer which serves as a guide in recognizing the hand gestures, confirming the change in modes and completing the executed actions by the system. Such a feedback system improves the quality, smoothness, and clarity of the interaction ensuring that the user is at any time aware of the system's responses during operation. The system control layer has been optimized to be able to perform the tasks swiftly and efficiently even under resource limitations which guarantee the continuous flow of the interaction without any irritating time lag.

The system as described here is an integration of several features such as real-time hand gesture recognition, a structured mode-based user interface, control and decision-making logic directed by the system control layer, personalization by user and feedback methods in a single framework. On the one hand, it can provide the users with a closed and well-organized interaction system by natural and hygienic means without compromising on the control and personalization being able to offer an efficient, scalable, and practical method of interacting with machines by gesture.

IV. SYSTEM ARCHITECTURE

It incorporates vision-based gesture detection, backend processing, and action execution in a neat and efficient way. It is represented by three complementary views: a functional block diagram (Figure 1), a use-case perspective (Figure 2), and a detailed processing workflow (Figure 3).

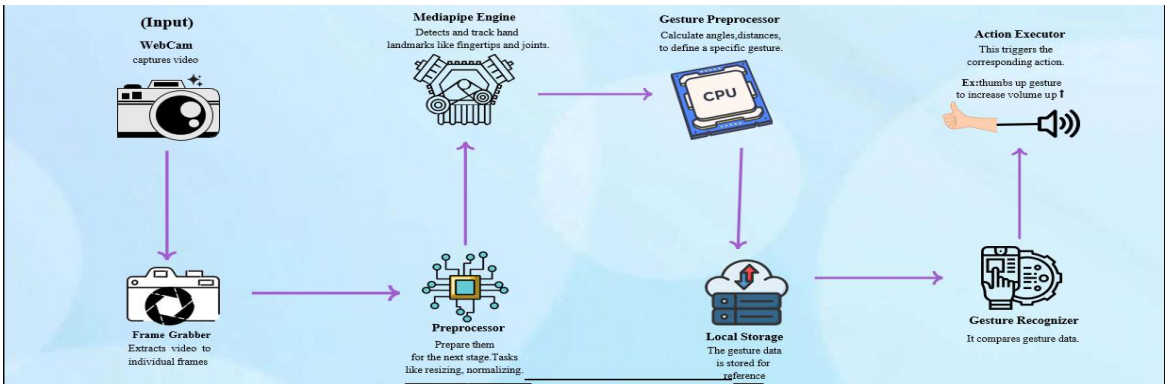

Figure 1 – System Architecture

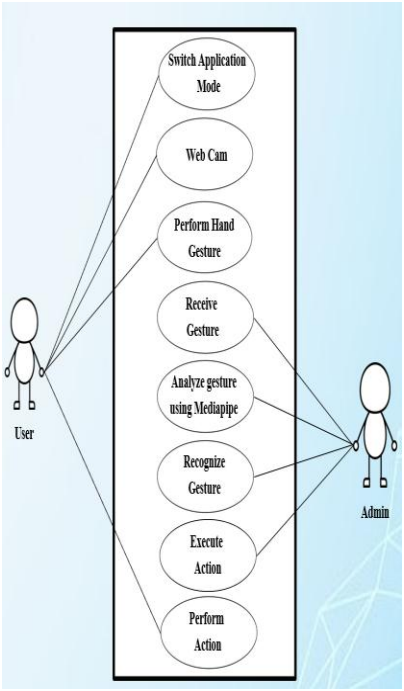

Figure 2 – Use Case

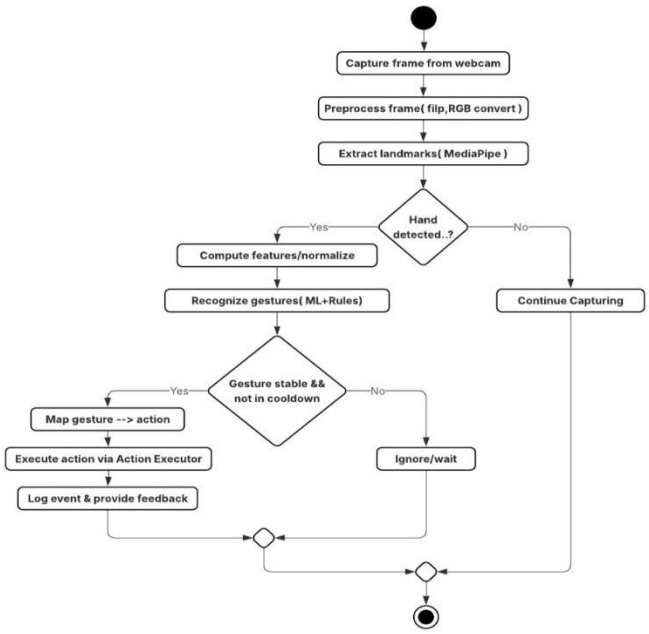

Figure 3 – Activity Diagram

Altogether, these figures give a full picture of how the user hand signs are captured, examined, and converted into corresponding actions.

The workflow of the system is as outlined in Figure 1 where it starts with a webcam providing the live video feed that is continuously being monitored. The frames grabbed from the footage are detached by the frame grabber unit and passed on to the pre-processing module where besides other things, resizing, normalization, and color space conversion take place. From

there, the preprocessed frames are fed into the Media Pipe system that provides a hand model with numerous landmarks, finger joints, key points of the hand, etc. The detected points are the key inputs of the gesture analysis module.

The gesture pre-processing unit on the backend uses these points or landmarks to derive geometrical characteristics of the hand such as the distances and angles between various landmarks, indicating the performed gestures. These features could be kept on the user's device for reference as well as for personalization. The gesture recognizer compares the extracted features with predefined or user-calibrated gesture patterns to identify the most suitable gesture class. The action executor gets a signal from the recognition module and triggers the matched device-action that ca

n be mouse cursor movement, clicking, scrolling, volume control, or application navigation, etc. The modular design of the system makes it very fast, extendable, and capable of supporting real-time interaction.

Figure 2 depicts the system from a use case perspective, which emphasizes the connection between the user, the system, and control administration. The user interacts with the system by performing hand gestures in front of a webcam, the system receives these gestures, and the gesture recognition pipeline analyzes them. The system supports application mode switching, allowing operation in different contexts such as desktop interaction, presentation control, scrolling, and system control. The administrative control monitors gesture logic and configuration, thus ensuring that the system behaves correctly.

The detailed workflow highlighted in Figure 3 represents the system's functional progression. Initially the system records the sequences of the frames and then the preprocessing of the captured images takes place which is followed by the extraction of the hand landmarks. At this point, the system needs to make a decision whether hand detection has been successful or not; if not, frame capturing continues without interruption. When the hand is detected, the system calculates the normalized representations, and then, gesture recognition takes place by applying the mixture of the rule-based decisions alongside the pre-trained models. After recognition, the potential gesture is subjected to stability and cooldown verification prior to being translated into an action. Actions will be executed upon confirmation of valid gestures accompanied by event journaling and user notification. At the same time, unstable gestures will be disregarded to preclude any unintentional actions.

The whole proposed architecture binds together the user's side interaction, backend processing, gesture recognition, and action realization in one solid and trustworthy framework. It is a multi-layered and modular composition that not only increases reliability but also decreases the likelihood of inadvertent triggers and at the same time, serves the user with a natural and intuitive gesture-based interaction that is fast and convenient for human–computer interaction (HCI) in the 21st century.

V. OUTPUT

OUTPUT SCREEN

VI. CONCLUSION and FUTURE WORK

This work presented a vision-based Gesture Control System aimed at enabling natural and touchless human–computer interaction. By interpreting real-time hand gestures captured through a camera, the system allows users to perform various computer operations without relying on conventional input devices such as a keyboard or mouse. The proposed system demonstrates how gesture recognition can be effectively used to improve interaction flexibility, accessibility, and user experience. Through structured gesture recognition, mode-based operation, and reliable action execution, the system provides an efficient solution for intuitive computer control. The overall design highlights the potential of gesture-based interfaces as a practical and user-friendly approach in modern human–computer interaction environments.

As future work, the system will be extended by developing a dedicated web-based platform to make the gesture control software easily accessible to a wider set of users. The future enhancement aims to provide the system as a software-style installation, allowing users to install the application on their systems in a simple and user-friendly manner. During installation, the application will be configured to automatically register itself within the system environment, including adding the required execution path to an autorun batch file. This approach will enable the gesture control system to start automatically when the system boots, ensuring seamless availability without requiring manual execution. By integrating a web-based distribution mechanism and automated startup configuration, the future version of the system will improve usability, accessibility, and ease of adoption while maintaining the core gesture-based interaction capabilities.

VII. REFERENCES

[1] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, et al., “MediaPipe: A framework for building perception pipelines,” arXiv preprint arXiv:1906.08172, 2019.

[2] I. S. MacKenzie, Human–Computer Interaction: An Empirical Research Perspective. Morgan Kaufmann Publishers, 2013.

[3] Z. Zhang, “Microsoft Kinect sensor and its effect,” IEEE Multimedia, vol. 19, no. 2, pp. 4–10, 2012.

[4] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, et al., “Real-time human pose recognition in parts from single depth images,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[5] A. Mittal, A. Zisserman, and P. H. S. Torr, “Hand detection using multiple proposals,” Proceedings of the British Machine Vision Conference (BMVC), 2011

[6] R. Poppe, “A survey on vision-based human action recognition,” Image and Vision Computing, vol. 28, no. 6, pp. 976–990, 2010.

[7] S. Greenberg and C. Buxton, “Usability evaluation considered harmful (some of the time),” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2008.

[8] S. Young, M. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, et al., The HTK Book. Cambridge University Engineering Department, 2006.

[9] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4–20, 2004.

[10] P. Taylor, A. Black, and R. Caley, “The Festival speech synthesis system,” Proceedings of the ESCA Workshop on Speech Synthesis, 1998.