A Survey on Gesture Controlled Virtual Mouse and Voice Assistant in the field of AI and ML

Dr. Surekha KB , Prerna Shankar, Priya Pawan Chaudhary, Piyush Kumar, Ritesh Mehta

BMSIT&M, *Bangalore*

ABSTRACT

The Gesture and Voice Controlled Virtual Mouse revolutionizes the way users interact with computers by seamlessly integrating voice commands and hand gestures, reducing the reliance on direct physical contact. Through the application of machine learning techniques, this system digitally manages input and output processes, enabling users to navigate and execute actions like left- and right-clicking using dynamic hand motions in lieu of a traditional physical mouse. This innovative implementation leverages machine learning for precise hand identification and integrates Python modules for effective voice assistance. By adopting this approach, the project not only simplifies fundamental mouse operations but also extends its capabilities to include tasks such as adjusting brightness, controlling volume, and managing applications, all while accommodating variations in ambient noise. Notably, the project boasts a voice assistant feature, providing an additional layer of convenience. Users can communicate with the virtual assistant through verbal or text instructions. In summary, the Gesture and Voice Controlled Virtual Mouse represents a sophisticated fusion of Computer Vision, Human-Computer Interaction, Speech Recognition, and Hand Gesture Recognition.

Keywords: *Computer Vision, Human-Computer Interaction, Speech Recognition, Hand Gesture Recognition*

I. INTRODUCTION

In the ever-evolving landscape of technology, the significance of Human-Computer Interaction (HCI) is reaching new heights. With the widespread adoption of touch screen technology in mobile devices, human-computer interaction has become more intuitive and natural. However, this seamless interaction paradigm has been somewhat limited in desktop systems, where the cost and complexity of implementing touch screen interfaces have posed significant challenges.

This is precisely where Computer Vision steps in as a game-changing innovation. By harnessing the power of Computer Vision, it can revolutionize the way humans interact with computers.

Instead of relying on conventional input devices like mouse and keyboards, Computer Vision offers an exciting alternative: a virtual engagement device that operates solely with a camera. This innovative approach eliminates the need for physical touch-based interfaces, opening up a world of possibilities for more affordable and accessible HCI solutions.

The core concept behind this technology lies in hand gesture recognition algorithms and methods. By leveraging advanced algorithms, computers can interpret and respond to various hand gestures made in front of a camera. This gesture recognition technology empowers individuals to interact with machines in a manner that feels incredibly natural, bridging the gap between human actions and digital responses.

The applications of gesture recognition technology are vast and diverse. From augmented reality experiences that blend the virtual and real worlds seamlessly, to enhancing computer graphics for immersive visualizations, the potential applications span a wide spectrum. In the realm of computer games, this technology opens doors to interactive gaming experiences that respond dynamically to players' gestures, creating a new level of engagement and excitement.

Moreover, gesture recognition technology holds promise in the field of biomedical equipment, offering innovative solutions for individuals who face challenges in controlling their limbs. By translating subtle hand movements into digital commands, individuals with limited mobility can gain newfound independence and access to digital interfaces, improving their overall quality of life.

This endeavor represents a pioneering step toward a future where technology adapts to human behavior, making interactions with computers more natural, accessible, and inclusive. Such innovative projects harness technologies such as computer vision, speech recognition, and machine learning to create a more intuitive and accessible human-computer interaction experience, offering increased convenience and productivity for users and enables users to control their computers and interact with digital devices through gestures and voice commands.

II. LITERATURE REVIEW

The paper named "Human Computer Interaction using manual Hand gestures in real time" by Mohammad Alsaffar, Abdullah Alshammari et al. suggests that the system's development primarily relied on key components such as the ADV7183 video encoder, the parallel peripheral interface and the DMA Controller, along with asynchronous memory SDRAM which work in tandem. The PPI, when combined with the DMA Controller, facilitates hardware-based image processing. Subsampling thus enhances processing speed by reducing memory access without imposing a substantial performance impact on the application.

The implementation of the system is done using a dedicated processor with encoders and controllers. Space limitation is a drawback of the processor because of saving the gesture in each iteration and also requires hardware devices to perform the mouse operations.[1]

The paper "Virtual Mouse and Assistant: A Technological Revolution Of AI" by Jagbeer Singh ,Yash Goel et al. suggests that the proposed system operates seamlessly without the need for external hardware, relying on Computer Vision and AI ML algorithms for automatic hand gesture recognition and assistance. When performing mouse operations via hand gestures, the initial step involves utilizing OpenCV to activate a webcam and capture real-time images of the user's hands. These images are processed to track landmarks and then various mouse operations are then executed. The drawback of the system was that the system used complex hand gestures to perform mouse operations, i.e., the system employed intricate hand gestures for executing mouse functions which could be difficult for the users.[2]

The paper titled "Virtual Mouse Using Hand Gesture Recognition- A Systematic Literature Review" by Prathamesh Shinde, Prof. Pradnya V. Kulkarni and others provides a thorough examination of existing methods and techniques implemented or proposed, along with an exploration of challenges in the field of virtual mouse using hand gestures. The suggested approach outlines a systematic methodology beginning with image capture through a webcam or built-in camera. The system is then configured to identify hands, allowing users to specify parameters such as the number of hands, minimum detection confidence, and minimum tracking confidence. Upon image acquisition, the conversion to RGB and application of image processing techniques, such as background subtraction and color segmentation, facilitate the extraction of hand and finger landmarks. The system subsequently tracks the hand, determining its location and unique identification.

The final step involves simulating various mouse actions like left, right, single as well as double clicks. This resulted in a comprehensive process for computer interaction through hand gestures.

One limitation of this system arises when the background color coincides with the designated hand color, potentially causing malfunctions. This issue is exacerbated in environments with varying ambient background lighting. [3].

"Cursor Control System Using Hand Gesture Recognition" by Ashwini M. Patil, Sneha U. Dudhane et al. is a paper of related work that suggests that for implementation, a convex hull algorithm is employed for hand detection and gesture recognition, enabling control of computers and other applications through detected gestures. Skin color characteristics within the YCrCb color space are employed for effective discrimination and background subtraction to distinguish between human skin and background objects. This process involves capturing an initial frame with just the background, and subsequent frames are compared to it. Pixels that pass a specific threshold are considered part of the human body and are retained in a new frame, while others are treated as background. This results in a frame with only the human figure against a zero-color background, facilitating more effective detection. The feature of the algorithm helps in detecting finger tips of the hand and recognizes if the finger is folded.

The drawback is that the system detects skin pixels and segments the hand, both of which need previously captured images.[4]

The paper "Hand Gesture Recognition System Using Camera" by Viraj Shinde, Tushar Bacchav, Jitendra Pawar et al. suggests a system that consisted of three primary stages. The first stage involved object detection, where the goal was to identify hand objects within digital images or videos. While improved environmental conditions and camera equipment can help mitigate these problems, achieving control in real-world settings or product applications can be challenging. Therefore, employing image processing methods served as a more effective solution for handling these image issues and constructing gesture recognition system that is adaptable and resilient.

In the second stage, hand objects were analyzed for identifying specific gestures, addressing key research issues such as extracting distinctive features and selecting effective classifiers. Finally, the third stage entailed the analysis of sequential gestures to recognize user instructions or behaviors.

A limitation of the system was its reliance on the user to wear a black belt on the wrist of the gesturing hand to achieve accurate segmentation of hand shape which will be eliminated in our proposed methodology.[5]

"Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques" by Nasser H. Dardas and Nicolas D. Georganas et al. introduces a real-time system with three models including hand identifying and tracing through background subtraction, second being skin detection, followed by contour comparison algorithms. It uses posture recognition using bag-of-features and multiclass SVM. The algorithm extracts key points from training images, mapping them into a unified bag-of-words vector with k-means clustering which is input to the multiclass SVM classifier. In testing, the classifier categorizes detected hand postures from a webcam, constructing visual word vectors for key points in a small image. A drawback is the

utilization of cluster and multiclass SVM models, which can be challenging and time-consuming for complex tasks like hand gesture recognition. System accuracy is affected by webcam quality, training image quantity, and the choice of cluster count for model construction.[6]

"A Real-Time System for Hand Gesture Recognition Using Motion History Image," authored by Chen-Chiung Hsieh and others, introduces a face-centric, adaptive skin color model alongside a motion history image technique that relies on the direction of hand movements. The paper defines six hand gestures, comprising of four dynamic gestures and two static gestures. Specifically designed Harr-like features are employed to detect the four directional dynamic hand gestures, while the skin color models that adjust or modify themselves based on facial characteristics identify the static hand gestures within an area of focus (AOF) around the face. The system underwent testing with five participants, revealing experimental results that indicated an accuracy of 94.1% on an average, validating the proposed system's feasibility. A constraint of the structure was the utilization of more intricate hand movements for mouse operations [7].

The paper "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques" by Munir Oudah et al. examines various techniques for gesture recognition. Explored methods include identification of hand gestures relying on the
Color-based recognition, Appearance-based recognition, Computer Vision Approach, Instrumental Glove Approach among others. The paper highlights a flaw in interaction systems, underscoring the necessity for reliable and robust algorithms. Our suggested methodology will enhance this approach using computer vision for hand gesture detection. Utilizing a camera vision-based sensor proves advantageous, as it enables contactless communication between humans and computers.[8]

The paper "Static Hand Gesture Recognition with Parallel CNN" by Qing Gao, Zhaojie Ju, Jinguo Liu and others introduces a technique involving two channels of parallel Convolutional Neural Networks (CNNs). One channel analyzes images recording all the gestures of hand in the RGB color space based on color information, while the other processes hand gesture images using depth information. In the experimental phase, the system's performance was compared to single-channel RGB-CNN, single-channel Depth-CNN, and prior research by other scholars. The outcomes revealed that employing the parallel CNNs approach improves the precision of interpreting hand gestures. This ensures smoother spatial Human-Robot Interaction (HRI) tasks, and the parallel CNN proves effective in varying illumination and complex backgrounds.
The drawback of this system is that it uses parallel CNN: one for color of hand and the other for depth hand gesture images making it more complex to fuse the two.[9]

The paper "The Virtual Assistant- A wearable device for Independent Living of the Visually Impaired" by Balamurugan M et al. suggests the utilization of the device is primarily made up of two parts: the first is a headset that includes an IR sensor, microphone, and earbuds, and the second is a processing unit. The RF module and Arduino UNO provides wireless communication, while IR sensors detects objects in front of the user.

The drawback is that hardware is required for building RF Receiver, decoder, encoder, transmitter, etc. for developing smart device which is costly and difficult to maintain. The proposed technology will not make use of any hardware devices in order to execute tasks according to user instructions.[10]

The paper "A Study on Speech Recognition Technology" authored by Dr. Ramalatha Marimuthu et al. underscores the intricacies of speech technology, encompassing processes like sampling, bit resolution, and the recognition of vocal signals. An inherent strength of this technology lies in its natural and unobtrusive signal production, rendering it well-suited for applications involving remote users and tasks reliant on speech interfaces, showcasing the technology's versatility. Notably, speech verification achieves impressive accuracy while placing minimal computational demands on systems.

Nevertheless, a notable drawback of this system is the inherent challenge of maintaining robustness amidst the variability of the communication channel. This limitation poses a considerable hurdle to achieving consistent performance in diverse settings, pointing to an area for potential improvement in future developments of speech recognition technology.[11]

Table 2.1: REVIEW OF RECENT RESEARCH ARTICLES ON GESTURE CONTROLLED VIRTUAL MOUSE

Article	Title Of Paper	Description	Advantages	Disadvantages
[1]	Human Computer Interaction using manual hand gestures in real time	The implementation of the system is done using a dedicated processor with few encoders and controllers.	The system identifies 79% of the frames successfully, resulting in a notable reduction in processing time.	Space limitation is a drawback of the processor while saving the gesture in each iteration.
[2]	Virtual Mouse and Assistant: A Technological Revolution Of AI	The hand gestures and assistant both work automatically without the help of any external hardware by using Computer Vision and AI ML algorithms.	The system achieves real-time capture of hands with substantial accuracy.	The system uses complex hand gestures to perform mouse operations.
[3]	Virtual Mouse Using Hand Gesture Recognition- A Systematic Literature Review	The paper addresses all implemented or proposed methods and techniques comprehensively and their diverse challenges in the hand gesture recognition for a virtual mouse.	The system captures hand images, converts them to RGB, and conducts image processing, including background subtraction and color segmentation.	The background color coincides with the designated hand color, causing few malfunctions. This issue is exacerbated in environments with varying ambient background lighting.
[4]	Cursor Control System Using Hand Gesture Recognition	The convex hull algorithm is employed to determine whether a finger is located within the palm area by addressing the challenge of identifying the largest polygon encompassing all vertices.	The characteristics of the algorithm help in detecting the tips of the fingers of the hand and recognizes if the finger is folded or not.	The system detects skin pixels and segments the hand, both of which need previously captured images.
[5]	Hand Gesture Recognition System Using Camera	The stages are object detection, object recognition and analyzing sequential gestures.	The system's advantage lies in minimizing external interfaces such as a mouse and keyboard.	For the precise segmentation of hand shape, users need to wear a black belt on the wrist of the gesturing hand.
[6]	Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques	During the training stage, the cluster and multiclass SVM classifier models were constructed, and these models were subsequently employed in the testing phase to identify hand gestures captured by a webcam.	The system has effective real-time performance, maintaining satisfactory results across varying frame resolutions. It also exhibits high classification accuracy even under conditions of variable scale, different orientation, illumination, and also in cluttered background.	The algorithms for multiclass SVM classifier models that were used can be challenging and time-consuming for complex tasks.

[7]	A Real Time Hand Gesture Recognition System Using Motion History Image	The paper suggests an adaptive skin color model based on facial features and a method utilizing motion history images derived from hand movement direction.	Recognition rates are not only good for static gestures but also for dynamic gestures.	More intricate hand movements are the principal drawbacks of its methodology.
[8]	Hand Gesture Recognition Based on Computer Vision: A Review of Techniques	The paper reviews all the different techniques including glove based attached sensor and color based hand gesture recognition.	The various techniques for hand gesture recognition identify a variety of gestures to perform mouse operations.	Hand gesture recognition seeks to rectify a flaw in interaction systems, emphasizing the necessity for the development of reliable and robust algorithms.
[9]	Static Hand Gesture Recognition with Parallel CNN	Parallel CNNs is designed where one channel analyzes images capturing hand gestures in the RGB color space based on color information, while the other processes hand gesture images using depth information.	It works well in changing illumination and complicated backgrounds.	It considers parallel CNN: one for the color of the hand and the other for the depth of the hand gesture images making it more complex to fuse the two.

Table 2.2: REVIEW OF RECENT RESEARCH ARTICLES ON VOICE ASSISTANTS

Article	Title Of Paper	Description	Advantages	Disadvantages
[10]	The Virtual Assistant -A wearable device for Independent Living	The device is primarily made up of two parts: the first is a headset that includes an IR sensor, microphone, and earbuds, and the second is a processing unit.	The RF module and Arduino provide efficient wireless communication, while IR sensors detects objects in front of the user.	Hardware is required for building RF Receiver, decoder, encoder, transmitter, etc. for developing smart device.
[2]	Virtual Mouse and Assistant: A Technological Revolution Of AI	The system identifies user commands using a speech recognition module.	The suggested system functions as a proficient voice assistant, accomplishing tasks with optimal efficiency.	Only limited functionalities are added.
[11]	A Study on Speech Recognition Technology	Speech technology is described by using processes like sampling, bit resolution and identification of speech signals etc.	Application of speech recognition technology is increasing day by day for remote users and speech interface tasks.	Robustness to channel variability is the biggest challenge to the current systems.

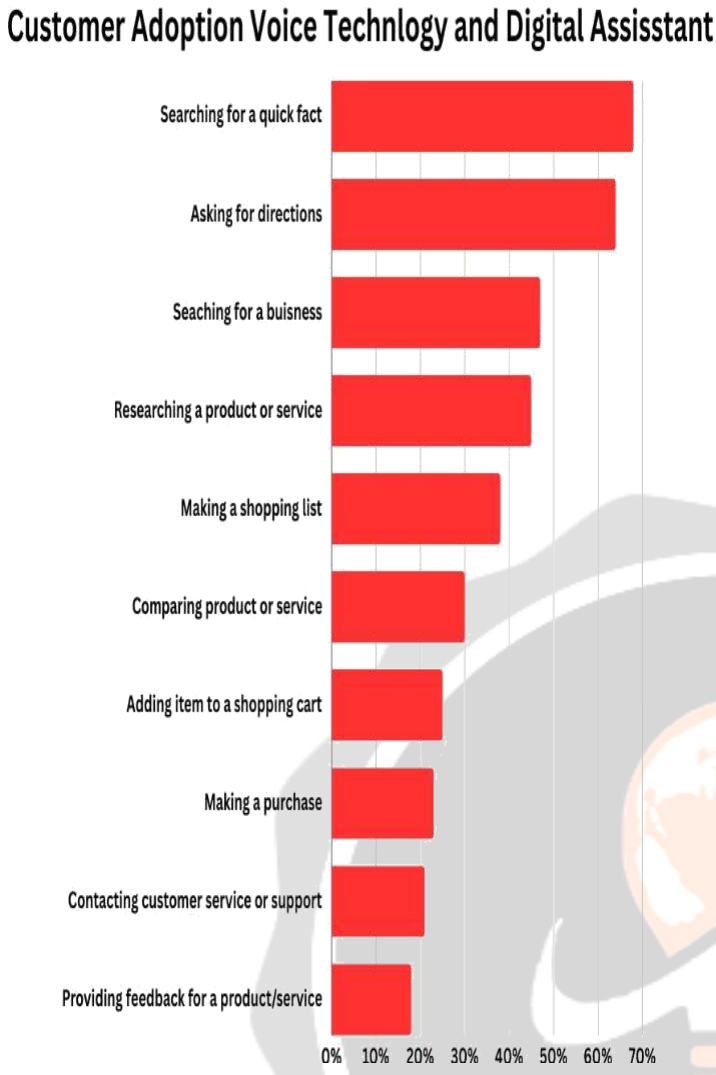

Fig 2.1: Statistics Of Consumer Adoption to Voice Technology and Digital Assistants

As shown in Fig. 2.1, Voice technology and digital assistants have revolutionized the way users interact with technology, offering a hands-free and natural user experience. Their convenience, accessibility, and personalization have fueled their rapid adoption, particularly among younger generations.

III. CONCLUSION

In conclusion, the intersection of Computer Vision and Human-Computer Interaction (HCI) heralds a groundbreaking era in technology, offering innovative solutions to longstanding challenges. Utilizing gesture recognition algorithms and leveraging cameras as virtual engagement tools has the potential to revolutionize our computer interaction methods. The introduction of this technology opens doors to a myriad of possibilities, from creating immersive augmented reality experiences to revolutionizing the gameplay.

Moreover, the impact extends far beyond entertainment, reaching into the realm of biomedical applications, where individuals with limited mobility can regain independence through intuitive digital interfaces.

At the heart of our initiative lies the goal to democratize HCI, making it more affordable, accessible, and inclusive for everyone. By eliminating the barriers posed by expensive touch screen technology and intricate desktop setups, it is paving the way for a future where individuals can effortlessly communicate with computers using natural hand gestures.

As researchers delve deeper into the realms of computer vision and gesture recognition, it is not merely redefining the way user interact with machines; it is also fostering a sense of empowerment and independence among users, particularly those with physical limitations. Our vision is to create a world where technology adapts seamlessly to human behavior, enhancing lives, fostering creativity, and bridging the gap between the digital and physical worlds.

In essence, the fusion of Computer Vision and HCI represents a significant leap forward, marking a paradigm shift in the way user perceives and engage with technology. Developers continue to innovate and explore the vast potential of these technologies, and are shaping a future where human-computer interactions are not just interfaces but meaningful and

natural dialogues, enriching the human experience in profound ways. Such a system has the potential to propel the technological world to new heights, enabling a multitude of operations to be executed without the need for physical touch.

IV. REFERENCES

[1] Mohammad Alsaffar, Abdullah Alshammari , Gharbi Alshammari , Tariq S Almurayziq , Saud Aljaloud , Dhahi Alshammari, and Assaye Belay ," Human Computer Interaction using manual Hand gestures in real time"- 2023

[2] Jagbeer Singh ,Yash Goel, Shubhi Jain, Shiva Yadav, "Virtual Mouse and Assistant: A Technological Revolution Of AI" - 2022

[3] Renuka Annachhatre, Miti Tamakuwala,Prathamesh Shinde, Prof. Pradnya V. Kulkarni, " Virtual Mouse Using Hand Gesture Recognition- A Systematic Literature Review"- 2022

[4] Ashwini M. Patil, Sneha U. Dudhane , Monika B. Gandhi, Nilesh J. Uke, "Cursor Control System Using Hand Gesture Recognition" - 2021

[5] Ashwini M. Patil, Sneha U. Dudhane , Monika B. Gandhi, Nilesh J. Uke, "Cursor Control System Using Hand Gesture Recognition"-2021

[6] Nasser H. Dardas and Nicolas D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques" – 2021

[7] Chen-Chiung Hsieh, Dung-Hua Lio and David Lee, "A Real Time Hand Gesture Recognition System Using Motion History Image" - 2020

[8] Munir Oudah , Ali Al-Naji and Javaan Chahl, "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques"-2020

[9] Qing Gao, Jinguo Liu, Zhaojie Ju , Yangmin Li, Tian Zhang , and Lu Zhang, "Static Hand Gesture Recognition with Parallel CNN" – 2020

[10] Balamurugan M, Logesh , Nirmal Kumar , Prasath S, Saranraj N, "The Virtual Assistant-A wearable device for Independent Living"-2023

[11] Arul.V.H , Dr. Ramalatha Marimuthu, "A Study on Speech Recognition Technology" – 2021

[12] Md. Zahirul Islam Department of Computer Science and Engineering University of Chittagong Chittagong, Bangladesh; "Static Hand Gesture Recognition using Convolutional Neural Network" Article. April 2021

[13] Mayur V. Gore Department of electronics Government college of engineering Aurangabad; "Human Computer Interaction using Hand Gesture Recognition" Article. April – 2021

[14] Lokesh Gagnani, Hetvi Patel, Sakshi Chaturvedis, Rajeshwari Jaiswal Computer Engineering Department, LDRP Institute of Technology and Research Gandhinagar Gujarat India; "Gesture Controlled Mouse and Voice Assistant" Article. 10 Oct 2022

[15] Swapnil D. Badgujar, Gourab Talukdar, , Omkar Gondhalekar,Mrs. S.Y. Kulkarni, "Hand Gesture Recognition System"- 2020

[16] Hsiang-Yueh. Lai , Han-Jheng. Lai , "Real-Time Dynamic Hand Gesture Recognition"-2020

[17] Ahmad Puad Ismail , Farah Athirah Abd Aziz, Nazirah Mohamat Kasim, Kamarulazhar Daud, "Hand gesture recognition on python and opencv" – 2020

[18] Ruchi Manish Gurav, Premanand K. Kadbe, " Real time Finger Tracking and Contour Detection for Gesture Recognition using OpenCV " – 2020

[19] Ehsan Niloy, Jannatun Meghna, Mohammad Shahriar, "Hand Gesture-Based Character Recognition Using OpenCV and Deep Learning"-2021

[20] Abhay Dekate, Chaitanya Kulkarni, Rohan Killedar Department of Computer Engineering, AISSMS College of Engineering, Pune, Maharashtra, India; "Study of Voice Controlled Personal Assistant Device" Article. December 2021