A Brief Review on Voice Assisted Hands-free Virtual Painting System

A.R. Jayasri¹, Dr. P. Kanimozhi² & Dr. T. Ananth kumar³

¹,²Department of Computer Science and Engineering, IFET College of Engineering, Villupuram, India.
Corresponding Author (A.R. Jayasri) Email: jayasriar2002@gmail.com

DOI: https://doi.org/10.46431/MEJAST.2023.6401

Copyright © 2023 A.R. Jayasri et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Article Received: 03 August 2023 Article Accepted: 20 October 2023 Article Published: 05 November 2023

ABSTRACT

This study employs various Human-Computer Interaction methods to introduce an innovative system that empowers individuals with diverse abilities to use computers in novel ways. It combines gesture-based virtual mouse control and painting, along with eye and voice integration. Advanced eye-tracking technology to accurately detect and interpret ocular movements is incorporated. These movements are then mapped to specific on-screen actions, facilitated by the computer webcam. Simultaneously, hand gesture-based mouse control is also available, where it captures and analyzes intricate hand motions, translating them into precise on-screen commands. The hand tracking painting feature revolutionizes by providing users with a virtual canvas. Through hand movements, users can write and draw with ease with various tools, expanding the potential for interactive creativity. Easy switching between the traditional mouse, virtual mouse, and gesture-based painting modes are controlled by a voice assistant. This integration provides a comprehensive and adaptable digital environment, enhancing user convenience and mobility. This method seamlessly integrates of cutting-edge technologies, culminating in the creation of an inclusive and exceptionally intuitive user experience.

Keywords: Human-computer interaction; Multi-modal interaction system; Assistive technology.

1. Introduction

In the ever-evolving landscape of human computer interaction, innovation continually shapes the way we engage with digital environments. The advancement and widespread adoption of these alternative input devices are crucial for elevating the quality of life for individuals with limb loss. The need for a virtual mouse from the ever-evolving landscape of human-computer interaction, where innovation continually reshapes how we engage with digital environments. Traditional mouse input has served as a fundamental means of navigating computers, but the demand for more intuitive, adaptable, and accessible interfaces has led to the emergence of virtual mouse solutions.

These innovative interfaces offer users a diverse range of interaction methods, including gesture-based control, eye-tracking, and voice assistance, ultimately enhancing the way users interact with their digital devices. A more inclusive digital environment is created by enabling people to communicate with technology using hand gestures, eye motions, and voice instructions. This mouse and paint multi-modality go beyond personal computing. It has potential uses in assistive technology, virtual reality, ATM, design, and other industries, creating new opportunities for creative human-computer interactions.

2. Literature Survey

Human-Computer Interaction (HCI) encompasses a various techniques and approaches that enhance seamless communication and interaction between humans and computers. Empowering users to overcome the limitations of traditional input methods, thereby promoting increased participation and inclusivity in society [1]. With MediaPipe integrated, users are provided with a digital canvas where they can express their creativity through intuitive hand movements. MediaPipe's sophisticated hand tracking capabilities are harnessed to enhance the painting experience significantly. This technology enables users to create with precision and fluidity, expanding the realm of interactive creativity within the digital environment. The authors in [2] used Media Pipe's implementation of Hand detection to
recognize hand gestures. Media Pipe is a cross-platform framework for building pipelines of machine learning algorithms. The Hand detection algorithm uses a convolutional neural network (CNN) to detect hands in images and videos. They used Google's Speech-to-Text API to transcribe voice commands. This limitation may hinder its application in creative tasks or artistic endeavors, where precise drawing control is essential. The system heavily relies on gesture recognition for input commands. While this is intuitive for many tasks, it may pose challenges for users with certain motor impairments who struggle to perform intricate hand movements consistently. In such cases, alternative input methods may be necessary. Mastering the gestures and commands may take time, and some users might find it less intuitive initially.

API uses a variety of machine learning techniques to achieve high accuracy. Air canvas systems typically utilize camera sensors to capture video frames of the air canvas interface. Image processing techniques are then applied to extract features from the video frames, such as object boundaries and hand gestures. Gesture recognition algorithms can identify specific gestures drawn on the air canvas, while machine learning models can learn the patterns of gestures and associate them with corresponding actions. The paper [3] presents a system that combines real-time object tracking and shape recognition to offer remarkable drawing capabilities. It operates by detecting a green object within the camera feed and utilizes OpenCV for tracking the object's movement in the air. Subsequently, it translates these movements into drawing actions on the screen, providing users with a unique and interactive drawing experience. A Convolutional Neural Network (CNN) model is employed to accurately recognize and classify the shapes drawn by the user, enhancing the system's versatility and functionality. Dependence on specific colored object, namely a green one, for tracking and drawing can constrain user choice and creativity, as it mandates the need for the consistent availability and use of a green object as the drawing tool. Users who may not have access to such an object or who prefer alternative colors may find the system less adaptable to their preferences.

The Air Canvas system employs a sophisticated methodology that encompasses both external hardware components and software components to provide users with a dynamic and interactive drawing experience. The system in [4] integrates an external hardware component, such as a stylus or a tracking device, to enable users to interact with the digital canvas. This component typically incorporates sensors or markers that are essential for tracking the user's movements in the physical space. The system's reliance on an external object, like a stylus or tracking device, to enable drawing subtly introduces an additional layer of complexity for users to manage. This can potentially pose challenges for individuals who may not have access to such external objects or prefer a more straightforward and purely digital drawing experience.

One of the most common approaches for gesture-based mouse control involves the use of inertial measurement units (IMUs), such as the MPU6050, which capture orientation and acceleration data as used in [5]. This data can be processed using various algorithms, including Kalman filtering, to extract meaningful motion patterns. Kalman filtering is a recursive estimation technique that combines sensor measurements with estimated state values to produce more accurate estimates. It is particularly effective in reducing noise and improving the accuracy of motion tracking in gesture-based mouse control systems. However, a drawback of this approach is that individuals with injuries may find it impractical to wear the external glove.
The software component of the AI virtual mouse employs an AI model that interprets input from both the camera and microphone, translating it into meaningful mouse functions based on the user's commands and hand movements. Notably, computer vision techniques are applied to the camera input, enabling the system to recognize and track hand motions and gestures accurately. Machine learning algorithms further enhance the system's capability by learning and adapting to individual user preferences and behavior over time. The microphone input undergoes a transformation through natural language processing algorithms, which convert spoken commands into text and subsequently map them to mouse functions like clicking, scrolling, and dragging. This natural language processing aspect significantly enriches the user experience, providing a versatile means of interaction. The author [6] provides a critical foundation for the integration of AI and multimodal input devices, offering users a more intuitive, adaptable, and accessible means of engaging with digital environments. It signifies a vital progression in the field of HCI and holds significant promise for the continued evolution of technology's role in enhancing user convenience and inclusivity.

The hand recognition for the purpose of digital painting, employing a camera-based system that is robust and versatile in its gesture detection capabilities. This innovative methodology showcases the potential for reimagining human-computer interaction in creative applications. The system described in references [7], [10], and [14] relies on capturing and analyzing user movements through the use of a standard camera. Each frame of this input is meticulously processed using OpenCV and Mediapipe, a computer vision framework. The pivotal role of Mediapipe in tracking the user's hand movements and marking 21 key locations is central to the gesture detection process. The marked landmarks are then matched with a gesture library to discern specific gestures, enabling users to interact with the system or application seamlessly. A distinctive feature of this system is its independence from additional sensors like gyroscopes, accelerometers, or specialized IR/UV cameras, making it a more accessible and cost-effective solution. This characteristic contributes to the system's reliability and consistent performance. While it excels at facilitating digital painting, it does not possess the functionality to control the overall system, akin to traditional mouse or cursor control. It's essential to recognize this constraint when considering its potential applications, as it primarily offers drawing options rather than comprehensive system control. The author [8] represents an innovative step in the field of human-computer interaction, offering an efficient and adaptable solution for gesture recognition, particularly in creative applications. It demonstrates the potential for enhancing user experiences and opens up possibilities for broader applications in the realm of HCI.

The paper [9] employs a whiteboard to delve deeply into the development of an advanced hand landmark model, making a substantial contribution to the fields of computer vision and human-computer interaction. This focuses on the precise prediction of hand landmarks, showcasing a robust methodology that is resilient to self-occlusions and partially visible hands. A distinguishing feature of this model is its utilization of regression techniques, wherein it directly predicts the 3D coordinates of hand landmarks following the detection of the hand's palm within the entire image. This approach is crucial in establishing a highly reliable internal representation of hand posture, even when confronted with challenging scenarios like self-occlusions or hands that are partially obscured from view. To ensure the accuracy and reliability of this landmark model, the study meticulously curated ground truth data. These coordinates were further enriched by extracting Z-values from available image depth maps, whenever applicable.
The combination of real-world and synthetic data allowed for a more comprehensive and robust training dataset, thereby enhancing the accuracy of the hand perception model.

The approach outlined in reference [11] stands as a comprehensive methodology for real-time video analysis and the detection of facial features, underlining its notable relevance within the domains of computer vision and human-computer interaction (HCI). The method begins with obtaining video feed from a webcam using Python and OpenCV, thereby forming the foundation for subsequent analyses. In the first segment, the study emphasizes the fundamental step of acquiring a video feed from the webcam. This sets the stage for live stream capture and the subsequent application of object detection techniques. The subsequent segments delve into the detection of human facial features, including the face, eyes, and iris, and the tracking of eye blinks. The employment of dlib library and OpenCV filters showcases the integration of machine learning and computer vision to isolate and analyze specific facial attributes. The approach presented not only contributes to the development of accurate and efficient facial feature detection but also offers valuable insights into the potential applications of such technology in fields such as healthcare, security, and HCI.

The utilization of sign language as the primary mode of communication for the hearing impaired community underscores the significant challenges they face when interacting with those who can hear. For the global population suffering from hearing loss, technology plays a pivotal role in bridging this communication gap and enhancing their quality of life. This research [12] endeavor seeks to provide a comprehensive review and analysis of articles centered around sign language recognition using sensor-based glove systems. The primary objective is to explore the academic motivations, identify prevalent challenges, and offer recommendations within this evolving field. The insights gathered from this study may pave the way for innovative solutions and further advancements, ultimately contributing to the development of technology that facilitates enhanced communication and accessibility for the hearing-impaired community.

The exploration of eye-tracking technology [13], [15], [16] for mouse control has given rise to various methods, each with its distinct advantages and limitations. Two primary approaches is webcam-based eye tracking. The initial method, employing webcams mounted on computer screens or inbuilt webcams, offers a convenient means of tracking eye movements and mapping them to mouse movements. However, it encounters inefficiencies when dealing with vertical eye movements, as detecting the white region of the eye in these scenarios proves challenging, leading to inaccuracies in determining eye movement.

2.1. Problem Statement

The use of traditional computer mouse is associated with several challenges that affect user comfort, accessibility, and overall efficiency. Prolonged use of traditional mouse can result in physical strain, discomfort, and the development of conditions like carpal tunnel syndrome, which can be both painful and debilitating. Additionally, the repetitive and precise movements required by a mouse during extended use can contribute to the development of Repetitive Strain Injuries (RSIs), impacting tendons, muscles, and nerves and leading to long-term pain.

Moreover, traditional mouse pose accessibility barriers for individuals with mobility impairments, including those with limited hand dexterity or hand tremors. This limitation affects a significant portion of the population, making
it challenging for them to interact with computers effectively. Another constraint associated with traditional mouse is the need for a flat and stable surface for accurate tracking. This requirement can be inconvenient when working in various environments or when using laptops on the go, where finding an appropriate surface may be challenging. The physical space occupied by mouse on a desk or workspace is another issue. In smaller work areas, the presence of a mouse can be a constraint, reducing overall workspace efficiency and limiting the available space for other tasks or tools.

3. Conclusion

Research has undertaken a survey of various Human-Computer Interaction (HCI) methods that encompass a wide array of approaches, from gesture-based interactions to eye-tracking and voice-assisted control. The utilization of cutting-edge technologies and methodologies, including the integration of Convolutional Neural Networks (CNN), hand landmark models, whiteboard systems, and speech recognition APIs, highlights the diversity and innovation within the field of HCI. These methodologies cater to different user needs and preferences, and they underscore the continuous evolution of HCI to provide more intuitive and accessible ways for individuals to interact with digital environments. The inclusion of CNNs and hand landmark models demonstrates the application of machine learning and computer vision to empower HCI methods. Whiteboard systems provide visual interfaces for collaborative communication, and speech recognition APIs facilitate voice-driven interactions, offering users greater convenience and accessibility. This survey showcases the dynamic and ever-expanding landscape of HCI methods, emphasizing the importance of accommodating diverse user needs and preferences. The integration of advanced technologies and methodologies provides a glimpse into the promising future of HCI, as it continues to enhance the ways in which individuals interact with and navigate the digital world.

Declarations

Source of Funding
This study did not receive any grant from funding agencies in the public or not-for-profit sectors.

Competing Interests Statement
The authors have declared no competing interests.

Consent for Publication
The authors declare that they consented to the publication of this research work.

Ethical Approval
Not Applicable.

Author’s Contribution
All the authors took part in data collection and manuscript writing equally.

References


