ISBN: 978-93-340-4751-6

Touchless Tech and Voice Commands

Shreesha B [1], Yashavanth S [1], Spandhan Prasad S N [1], Rohith H P [2]
[1] Department of Artificial Intelligence and Machine Learning, VCET, Puttur, Karnataka, India
[2] Assistant Professor, Department of Computer Science and Engineering (Data Science), VCET, Puttur, Karnataka, India
Email ID: shreeshapilinja@gmail.com

***Abstract* — An innovative approach to human-computer interaction, leveraging machine learning and computer vision to interpret human hand gestures into input operations for various applications and games. The system captures these gestures using a standard webcam, eliminating the need for specialized hardware. It offers a range of input types, enhancing its functionality with speech commands and live transcription capabilities. The system simulates native touchpoints in the air, offering a range of gestures for both single and bi-modal hand gesturing. Our exploration has led to the development of a system with the potential to revolutionize digital device interaction, achieving low-latency input delivery through specific API calls within enabled applications. By providing an affordable solution for touchless interaction, our research could potentially transform the way we interact with digital devices, making it more intuitive, enjoyable, user friendly, and immersive. In technical terms, our system enables effective control of various applications such as maps, 3D models manipulation, gaming, and mouse functions through intuitive hand gestures, offering users an immersive and touchless interaction experience.**

Keywords— Machine Learning, Computer Vision, Human-Computer Interaction, Speech Commands, Touchless Interaction.

I. INTRODUCTION

In the field of Human-Computer Interaction (HCI), building a smooth and efficient communication bridge between people and technology is the ultimate goal. This is accomplished by giving technology human traits, which facilitates natural communication between people and machines. User-centered design, which puts users' requirements, skills, and preferences first during system development, is a pillar of HCI. The goal is to create user interfaces that are not only intuitive and exciting to use, but also functional. They seek to produce a seamless and intuitive user experience by giving the demands of the user first priority and iteratively improving design elements.

Direct physical contact with input devices, such as a mouse or touchscreen, has facilitated human-computer interaction. This method's drawbacks, meanwhile, are its intricacy and requirement for direct personal contact. Many efforts have been undertaken in the last ten years to create computer vision-based methods for computer interaction. By removing interaction from the confines of the two-dimensional plane, computer vision is a field that can significantly improve the naturalness of these interaction metaphors.

Using gesture-based controls to zoom, pan, and rotate views, for example, can improve navigation apps like maps and provide users a more natural and engaging experience. Users of the system can engage virtually with virtual items as if they were real by manipulating and navigating 3D models. More immersive experiences that mimic holding and interacting with real-world physical objects are made possible by this capability. The system has the potential to enhance gameplay by enabling users to move, select, and manipulate items in the virtual world with hand movements. Applications can benefit from touchless capabilities when the system is used for mouse activities. This allows users to click, drag, and drop things without making physical contact, much as touch interactions on mobile devices.

II. RELATED WORK

Shreesha B et al. [1] investigated how to combine computer vision and gesture recognition methods to improve Human Computer Interaction by utilizing cutting-edge libraries and algorithms. A range of applications are explored, such as gesture-based keyboards and calculators, emotion-driven music selection systems, and virtual mouse control utilizing hand and eye movements. With the goal to facilitate smooth engagement with digital interfaces, real-time camera feeds allow the recognition and tracking of facial landmarks, hand gestures, and body motions.

Dr Ranjeet Kumar et al. [2] delve into the development of voice assistant systems using Python, highlighting their role in modern technology. They discuss the integration of AI, machine learning, and neural networks, alongside reviewing related works in speech recognition. Their research methodology focuses on speech recognition, Python backend development, and Google Text to Speech integration.

Prof. P Ajitha et al. [3] presented a gesture-based volume control system that utilizes OpenCV, Mediapipe, PyCaw, and NumPy modules. With this method, users can change the volume on their computer by making hand motions that are photographed by a camera. This research demonstrates how computer vision and machine learning may be used to create interactive and user-friendly user interfaces that control audio output without the need for physical input devices.

ISBN: 978-93-340-4751-6

Kavitha R et al. [4] conducted a comprehensive literature review of hand gesture-controlled virtual mouse systems that leverage Artificial Intelligence (AI) technology. They highlight how the development of AI has led to an increase in the use of hand gesture recognition for operating virtual gadgets. The technology converts hand gestures into mouse movements on a virtual screen by using a camera to record hand movements and AI algorithms for detection. Hand gestures and voice commands are recognized by machine learning and computer vision algorithms without the need for extra hardware.

Kavana KM and Suma NR [5] presented a real-time on-device hand tracking solution for predicting a human hand skeleton using a single RGB camera. Touchless interaction has gained significant attention due to its wide-ranging applications, from medical systems to gaming, benefiting individuals with hearing difficulties who rely on sign language for communication.

Quam D L [6] explained about how a hardware-based system is developed. Although this model produces incredibly accurate results, many movements are challenging to execute while wearing a glove that severely limits the user's hand's range of motion, speed, and agility. Also wearing gloves for a long time will result in skin diseases and is not best suited for the users with sensitive skin type.

III. System Architecture

Our research is focused on integrating gesture and voice controls to transform user-computer interaction. Our goal is to improve interactions by addressing the several issues related to conventional input devices and making them more inclusive, smooth, and intuitive. We have determined the drawbacks of conventional input devices, such as a physical mouse, in a variety of contexts, including gaming, presentations, touch pads, and many more. We therefore concluded that hardware requirements, including operating system compatibility, webcam, microphone, and speaker, are based on the issues that have been identified.

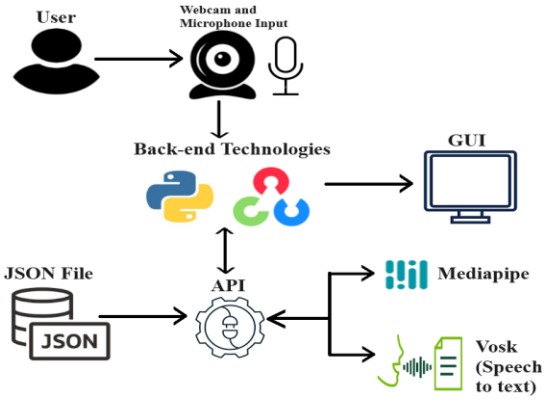

Fig. 1. System Architecture of Touchless and Voice Command Technology

A voice command and touchless technology application is intended for the system architecture shown in the Fig. 1. The user engages through a webcam input at the start of it. This kind of interaction isn't restricted to any one format; it may also be done using voice commands or hand movements. The webcam records these exchanges and feeds the data into the system for analysis. The back-end technologies, namely Python and OpenCV, are responsible for processing these inputs. Interpreting the webcam inputs is a critical function it performs. An Application Programming Interface, is used to make sure that different parts of the system integrate and communicate seamlessly.

Apart from interpreting gestures, the system also improves interaction possibilities by translating spoken words into text. Vosk, a speech-to-text tool, helps with this. Using machine learning and computer vision, this novel method of human-computer interaction converts hand gestures made by users into input for a range of games and apps. The system does not require additional hardware because it uses a normal webcam to capture these motions. With voice commands and real-time transcription capabilities, it improves its functionality with a variety of input methods. The technology offers a variety of motions for both single and bi-modal hand gesturing, simulating natural touchpoints in the air. The potential is to completely change how we interact with digital gadgets by offering a cost-effective alternative for touchless interaction that makes it more immersive, intuitive, pleasurable, and user-friendly.

IV. System Implementation

The implementation of the gesture recognition system involves the integration of two key modules the speech module and the hand gesture module. These modules utilize the Vosk and Mediapipe libraries respectively to interpret user gestures and speech. The system simulates native windows touchpoints in the air and also provides speech commands and live transcription capabilities.

A. Hand Gesture Module

The hand gesture module uses the Mediapipe library to detect 21 landmarks on each hand, such as the index tip, pinky base, and wrist. It calculates primitives like index pinched, thumb stretched, and palm facing camera. This module is responsible for interpreting hand gestures and translating them into specific actions or commands.

B. Speech Module

The speech module employs the Vosk library to recognize the user’s speech. Vosk is an offline open-source speech recognition toolkit that supports many languages. It provides a streaming API for the best user experience and allows quick reconfiguration of vocabulary for optimal accuracy.

ISBN: 978-93-340-4751-6

C. *Integration of Modules*

The integration of the hand gesture module and speech module is achieved through API interactions.

V. RESULTS

The results of our research indicate that the system successfully interprets a variety of hand gestures and voice commands, providing an interactive and immersive user experience. The system's performance was evaluated in various applications, demonstrating its versatility and potential.

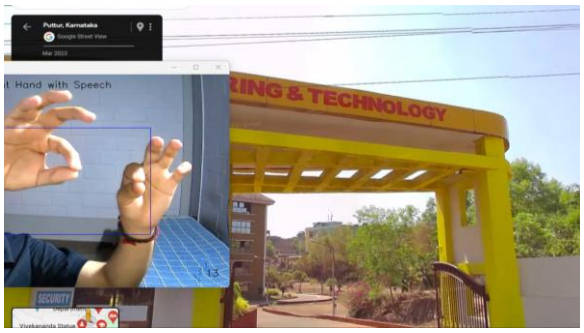

Fig. 2. Gesture-Controlled Zoom Functionality in Google Maps

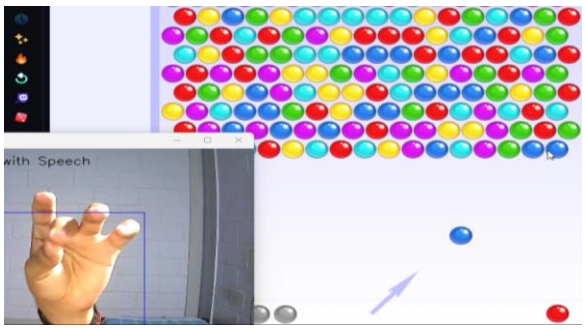

Fig. 3. Playing Bubble Shooting Game Using Hand Gestures

Fig. 4. Drawing in Paint Application Using Hand Gestures

Fig. 5. Rotating Scale Using Hand Gesture for Positioning

Fig. 6. Changing Presentation Slides Through Hand Gestures

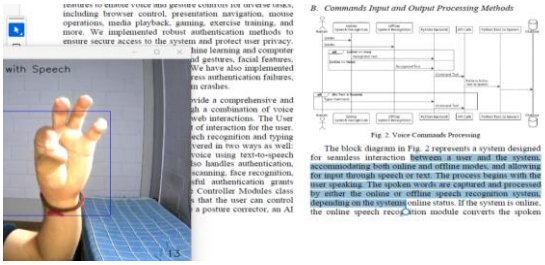

Fig. 7. Selecting Text Through Hand Gestures

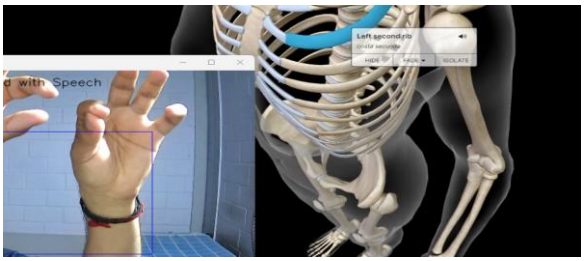

Fig. 8. Controlling a 3D Skeleton Using Hand Gestures in Medical Applications

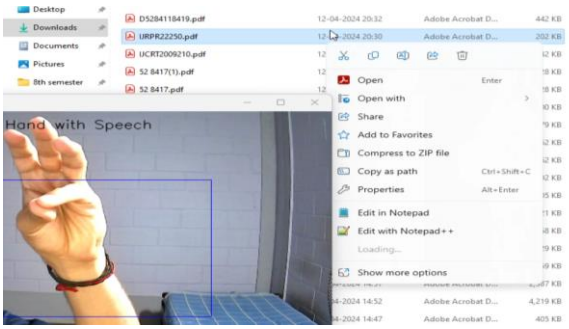

Fig. 9. Performing Left-Click Operation on a File Through Hand Gestures

Fig. 10. Controlling Media Playback with Hand Gestures: YouTube Player Example

ISBN: 978-93-340-4751-6

The system effectively interpreted zoom in and out gestures, enhancing the user's navigation experience in Maps as shown in Fig. 2. It successfully translated hand movements into game controls, providing a more engaging gaming experience, as depicted in Fig. 3. The system allowed users to draw on the screen using hand gestures in Paint Application, demonstrating its potential in creative applications, as illustrated in Fig. 4. It accurately interpreted rotation gestures in positioning a scale, showing its utility in 3D modelling and other similar applications, as demonstrated in Fig. 5. The system effectively recognized slide change gestures, offering a hands-free method for controlling presentations, as shown in Fig. 6. It successfully interpreted text selection gestures, demonstrating its potential in word processing and other text-based applications, as shown in Fig. 7. The system accurately translated hand movements into 3D model manipulations in medical applications and other scientific applications, as shown in Fig. 8. It effectively recognized the left-click gesture, demonstrating its potential as a touchless mouse alternative, as shown in Fig. 9. The system successfully interpreted play, pause, and volume control gestures in YouTube Player, enhancing the user's media consumption experience, as shown in Fig. 10. It can be used like these in various applications. These results demonstrate that our system offers a hands-free and convenient method for users, enhancing accessibility and opening up new possibilities for future exploration and development in the field of human-computer interaction. The system supports English language speech recognition, making it accessible to a wide range of users.

VI. Conclusion

We have proposed a standard set of interaction gestures for Windows PCs. By leveraging machine learning and computer vision, our system interprets human hand gestures into input operations for various applications and games. This system captures gestures using a standard webcam, eliminates the need for specialized hardware. Our system simulates native touchpoints in the air, offering a range of gestures for both single and bi-modal hand gesturing. The integration of gesture and voice controls can significantly improve interactions by addressing the several issues related to conventional input devices and making them more inclusive, smooth, and intuitive. Our research opens up new possibilities for future exploration and development in the field of human-computer interaction.

References

[1] Shreesha B, Yashavanth S, Spandhan Prasad S N, Rohith H P, "Enhanced Input for Human Computer Interaction," International Journal of Creative Research Thoughts (IJCRT), vol. 12, no. 4, pp. c899-c919, Apr. 2024. DOI: http://doi.one/10.1729/Journal.38800.

[2] Dr. Ranjeet Kumar, Muhammad Faisal, M., & Ahmad, O. (2022). "My Voice Assistant Using Python." International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 10, issue 4, pp. 3004. ISSN: 2321-9653.

[3] P. Ajitha, I. Ahmed M. N., M. Seshan K. M., and C. Siva, "Gesture Volume Control," International Journal of Research in Engineering and Science (IJRES), vol. 11, no. 5, pp. 154-160, May 2023. ISSN: 2320-9364.

[4] Kavitha R, Janasruthi S U, Lokitha S, and Tharani G, "Hand Gesture Controlled Virtual Mouse Using Artificial Intelligence," International Journal of Advance Research and Innovative Ideas in Education (IJARIIE), vol. 9, issue 2, 2023, pp. 19380. ISSN(O): 2395-4396.

[5] Kavana, K. M., & Suma, N. R, "Recognition of Hand Gestures Using MediaPipe Hands." International Research Journal of Modernization in Engineering Technology and Science, vol. 04, issue 06, pp. 4149, June-2022. ISSN: 2582-5208.

[6] D. L. Quam, "Gesture recognition with a DataGlove," in IEEE Conference on Aerospace and Electronics, Dayton, OH, USA, 1990, pp. 755-760 vol.2. DOI: 10.1109/NAECON.1990.112862.

[7] Trabelsi, A., Warichet, S., Aajaoun, Y., & Soussilane, S. (2022). Evaluation of the efficiency of state-of-the-art Speech Recognition engines. Procedia Computer Science, 207, 2242–2252. DOI: 10.1016/j.procs.2022.09.534.

[8] Meera Paulson, Nathasha P R, Silpa Davis, Soumya Varma, "Smart Presentation Using Gesture Recognition," IJRTI, vol. 2, issue 3, 2017.

[9] Steven Raj, N., Veeresh Gobbur, S., Praveen., Rahul Patil., & Veerendra Naik. (2020). Implementing Hand Gesture Mouse Using OpenCV. In the International Research Journal of Engineering and Technology (pp. 4257-4261).

[10] Sneha, U., Monika, B., & Ashwini, M. (2013). Cursor Control System Using Hand Gesture Recognition. In the International Journal of Advanced Research in Computer and Communication Engineering (pp. 2278-1021).

[11] Krishnamoorthi, M., Gowtham, S., Sanjeevi, K., & Revanth Vishnu, R. (2022). Virtual mouse using YOLO. In the international conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (pp. 1-7).

[12] Varun, K.S., Puneeth, I., & Jacob, T.p. (2019). Virtual Mouse Implementation using OpenCV. In the International Conference on Trends in Electronics and Informatics (pp. 435-438).

[13] Quek, F., et al. (1994). Towards a vision-based hand gesture interface, in Proceedings of Virtual Reality Software and Technology (pp. 17-31).

[14] Tharsanee, R.M., Soundariya, R.s., Kumar, A.S., Karthiga, M., & Sountharrajan, S. (2021). Deep Convolutional neural network-based image classification for COVID-19 diagnosis. In Data Science for COVID-19 (pp. 117-145). Academic Press.

[15] K. Shima, H. Hakoda, T. Kurihara, B. Sizuki, and J. Tanaka, "Range Selection Using Three Points Touch as Start-up Gesture", IPSJ SIG Technical Report, 2014-HCI-159, pp.1–8, 2014.

[16] T. Kuribara, Y. Mita, K. Onishi, B. Shizuki and J. Tanaka, "HandyScope: A Remote Control Technique Using Circular Widget on Tabletops", in Proceedings of 16th International Conference on Human-Computer Interaction (HCI International 2014), LNCS 8511, pp.69-80, Heraklion, Crete, Greece, June 22–27, 2014.

[17] G. J. Lepinski, T.Grossman, and G. Fitzmaurice, "The design and evaluation of multitouch marking menus", in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, New York, NY, USA, ACM, pp. 2233–2242, 2010.