IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, November 2024

GestureLink: Immersive Hand Guided Virtual Interface

Ashutosh Said[1], Shubhangi Said[2], Anand Khatri[3], Sachin Bhosale[4]
Department of AI & DS Engineering[1-4]
Jaihind College of Engineering, Pune, India

Abstract: *This report introduces an innovative Hand Gesture Mouse Controlling system that utilizes hand gestures captured via a webcam through a colour detection technique. This system enables users to navigate the computer cursor and perform clicking and dragging actions using different hand gestures.*
The proposed system relies solely on a low-resolution webcam functioning as a sensor, effectively tracking the user's hand gestures in multiple dimensions. The implementation will utilize Python and OpenCV for seamless integration. Hand gestures offer a natural and effortless means of communication. The camera's output will be displayed on the monitor, and shape and position information of the gestures will be acquired through colour detection. Additionally, the report outlines the implementation of a file transferring scheme using Python server programming.

Keywords: Virtual Mouse, Hand Gestures, Image Capture, Image Processing, Colour Detection

I. INTRODUCTION

Advancements in augmented reality and compact wireless technologies have revolutionized our daily interactions with devices. This research presents a novel AI-driven virtual mouse platform that utilizes hand gestures and hand tip recognition via computer vision to execute mouse tasks on a computer. Its core objective is to supplant conventional mouse peripherals by harnessing the computer's integrated camera or an external webcam. Through the application of computer vision methodologies for hand gesture and hand tip recognition, the suggested system establishes a framework for human-computer interaction (HCI). The AI virtual mouse platform facilitates accurate tracking of fingertip motions captured by the integrated camera or webcam. facilitating mouse cursor operations, scrolling, and cursor navigation. Unlike wireless or Bluetooth mouse devices that rely on additional components such as a dongle and battery, this system empowers users to control computer mouse operations solely through hand gestures using their existing camera hardware.

Within the proposed framework, frames are captured by the webcam and then undergo processing to identify different hand gestures and movements of the hand tip, which in turn activate particular mouse operations. The AI virtual mouse system is developed utilizing the Python programming language and integrates the OpenCV library for computer vision operations.

The MediaPipe package is employed for hands tracking, while the Pynput, Autopy, and PyAutoGUI packages enable window screen navigation and essential functions like left-click, right-click, and scrolling. Evaluation of the proposed model reveals exceptional accuracy levels, indicating its efficacy for real-world applications even without a dedicated GPU, relying solely on CPU capabilities.

II. LITERATURE SURVEY

A Low-Power mmW RadCom System for Short-Range Hand Gesture Sensing and Data Synchronization

This article investigates the amalgamation of multifunctional radar sensing and wireless communication (RadCom) to facilitate hand gesture recognition and data synchronization in smartphones. A novel low-power RadCom design centred on a 1-bit comparator is introduced, allowing for high-resolution sensing and rapid data transmission. The transmitter produces IR-UWB pulses and ASK symbols, achieving data transfer rates of up to 5 Gbps over short ranges. Furthermore, the 1-bit comparator architecture is proven effective in precisely discerning hand gestures and converting propagation scenarios into range Doppler maps (RDM).

Deep Facial Expression Recognition: A Comprehensive Survey

This survey delves into the realm of deep facial expression recognition (FER) amidst challenging real-world conditions. Leveraging the power of deep neural networks, this paper explores solutions to issues such as overfitting due to limited training data and expression-unrelated variations. It covers datasets used in the literature, data selection criteria, and evaluation principles. Additionally, it presents the standard pipeline of deep FER systems, innovative neural network architectures, and training strategies. The survey concludes with insights into performance benchmarks, related issues, and future directions for robust deep FER systems.

Single-hand Gesture Recognition in Manipuri Classical Dance using Skeletonization Technique

This study centres on identifying single-hand gestures in Manipuri classical dance, a vital aspect of Indian heritage. With 25 Asamyukta Mudras and 12 Samyukta Mudras, this dance style suffers from inadequate datasets for scholarly exploration. The research introduces the development of a Manipuri single-hand gestures dataset, gathered from 6 participants. It also deliberates on diverse gesture recognition approaches and emphasizes a feature extraction method employing skeletonization. This method facilitates the identification of 25 Asamyukta Mudras and differentiation between left- and right-hand utilization.

Gesture Controlled Mobile Robot: A Remote Gesture-Controlled Robot System

This research presents a novel system for remote-gesture-controlled robots utilizing hand-worn glove controllers fabricated via 3D printing. Utilizing the Arduino platform, the system facilitates concurrent manipulation of an industrial-style robotic manipulator arm and a robotic vehicle via hand gestures, thereby broadening the horizon of human-robot interaction. The investigation verifies the viability of glove controllers within this framework, proposing potential industrial applications for gesture-controlled robotics with future enhancements.

III. PROBLEM STATEMENT

The objective at hand involves establishing a means of human-computer interaction devoid of physical connections to the computer. Various approaches were proposed, but they all necessitated physical hardware manipulation. Another suggestion entailed utilizing the photoelectric effect principle; however, it requires specialized and economically impractical hardware. Consequently, the chosen course of action entails creating a virtual mouse through cost-effective image processing techniques. This solution effectively eliminates the challenges. Our workflow involves processing through open resources, without the use of sensors, and maintaining a low-cost approach.

IV. GOALS AND OBJECTIVES

The objective at hand involves Develop an AI-powered application that can be seamlessly integrated into a user's computer system. The application should be designed to operate using a webcam, ensuring easy installation and setup for users. It should enable intuitive functionality, such as drag and drop capabilities, as well as a convenient scrolling feature. The primary goal is to create a virtual input system that functions effectively on all surfaces. The application should efficiently translate hand gestures or motions into mouse inputs, accurately reflecting the intended screen position. The user interface (UI) of the application should be user-friendly, promoting ease of use and intuitive navigation. It is crucial for the program to prioritize speed and responsiveness, ensuring optimal performance without any noticeable delays or lag. Heavy processing tasks that may disrupt the user experience should be avoided to maintain a smooth interaction.

V. MOTIVATION

The motivation behind this project lies in the quest to harness the potential of AI within a streamlined framework. The primary objective is to delve into the realm of vision-based interfaces as a response to the inherent constraints and unnaturalness associated with conventional input devices like mice and joysticks. This endeavour gains particular significance in the context of intelligent environments, where there is a pressing need for interactions that feel intuitive and enable effective teleoperations.

It becomes increasingly evident that the era of the Virtual Mouse is on the horizon, poised to supplant physical mice in the not-so-distant future. This transition is driven by a growing desire for remote control and interaction with technological devices, all without the need for additional peripheral tools such as remotes or keyboards. This shift not only promises heightened convenience but also holds the potential for significant cost savings. Historically, users have been reliant on physical mice, necessitating a dedicated surface area and being bound by the limitations of cable lengths. In stark contrast, the Virtual Mouse obviates these requirements by leveraging a mere webcam to capture images of the user's hand position, thereby accurately ascertaining the desired pointer location. In essence, the driving force behind this project is the aspiration to craft an object tracking application that paves the way for seamless human-computer interaction, heralding a transformative paradigm shift in the manner in which we engage with computer systems.

VI. SYSTEM ARCHITECTURE

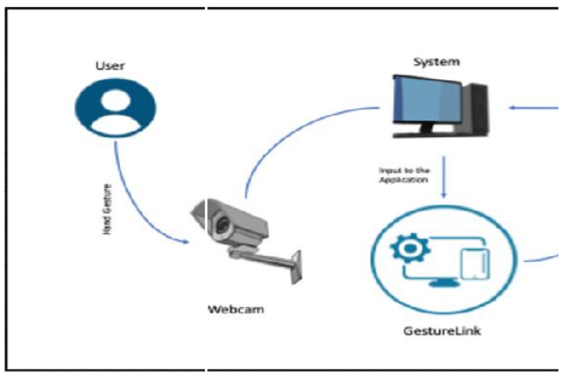

Fig. 1 System Architecture

Fig. 1, we present an overview of the image analysis techniques utilized for converting surface-touches into keystrokes in camera-based virtual keyboard systems. These schemes typically follow a sequence of image analysis steps. The initial step involves capturing an image I am using the system's camera, which is then subjected to skin segmentation to obtain a binary image. This binary image, represented by region H, corresponds to the hand(s) present in the scene. This research presents a novel system for remote-gesture-controlled robots utilizing hand-worn glove controllers fabricated via 3D printing. Utilizing the Arduino platform, the system facilitates concurrent manipulation of an industrial-style robotic manipulator arm and a robotic vehicle via hand gestures, thereby broadening the horizon of human-robot interaction. The investigation verifies the viability of glove controllers within this framework, proposing potential industrial applications for gesture-controlled robotics with future enhancements. Once the fingertip positions are estimated, the next challenge is to determine which fingertips are in contact with the virtual keyboard's keys. To address this problem, we propose a technique called shadow analysis. This method allows for the efficient mapping of touch points to key presses. To simplify the mapping procedure, we posit that the virtual keyboard keys can be represented as rectangles within a two-dimensional space (R2). Additionally, we presume that the keyboard layout is predefined during compilation and that the keyboard mat comprises control points enabling the derivation of a perspective correction transformation. Leveraging this information, we apply a straightforward formula to convert the touch points' coordinates in the keyboard-space to corresponding key presses.

VII. PROPOSED SYSTEM

To detect hand gestures and track hand movements, the study employs the MediaPipe framework alongside the OpenCV library for computer vision tasks. The algorithm integrates machine learning principles to facilitate the tracking and recognition of hand gestures and fingertip positions.

A. MediaPipe

MediaPipe, an open-source framework developed by Google, is designed for integration into machine learning pipelines. This versatile framework is particularly advantageous for cross-platform development, leveraging time series data as its foundation. With its multimodal capabilities, MediaPipe offers versatility across various audio and video sources, making it a valuable tool for developers seeking to build and analyze systems using graph-based approaches. By leveraging the MediaPipe framework, developers can effectively construct and evaluate systems, streamlining the development process. At the heart of MediaPipe are its pipeline configurations, which dictate the sequence of operations within the system. These pipelines are flexible and can operate seamlessly across different platforms, ensuring scalability for both mobile and desktop applications. The MediaPipe framework comprises three essential components: performance assessment, a framework for acquiring sensor data, and a library of reusable components known as calculators. A pipeline, represented graphically, consists of interconnected calculators through which data flows in the form of packets. This design allows developers to customize or replace calculators within the graph, enabling the creation of tailored applications. Through the interaction of calculators and streams, a comprehensive data-flow diagram emerges, illustrating the interconnectedness of calculators via streams.

B. Hand landmark detection

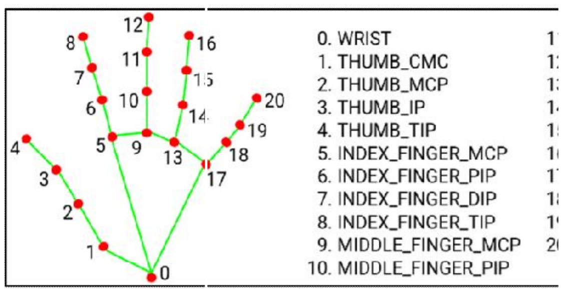

Fig. 2 Hand Landmark Key Points

MediaPipe employs a single-shot detector model to achieve real-time detection and recognition of hands or palms. Specifically, the model undergoes training tailored for palm detection within the hand detection module, as palms prove easier to train for this purpose. This decision stems from the enhanced performance of non-maximum suppression, especially when dealing with smaller objects like palms or fists. In Fig. 2, the hand landmark detection model assumes a pivotal role by pinpointing 21 joint or knuckle coordinates within the hand area. OpenCV, a widely used computer vision library, furnishes an array of image-processing algorithms for object detection. Written in Python, OpenCV empowers developers to construct real-time computer vision applications. It functions as a versatile tool for processing images and videos, facilitating seamless integration of tasks such as face detection and object recognition.

C. Mathematical Model

We need to define the relationship between the image captured by the webcam and the position of the virtual pointer. We'll use some basic geometric concepts and transformations to represent this relationship.

Let's assume that the webcam captures a 2D image of the user's hand position, denoted as (x_cam, y_cam). This image will be processed to extract relevant features representing the position of the hand. Let's call these extracted features (x_hand, y_hand).

We can define a transformation function that maps the coordinates of the hand in the image (x_hand, y_hand) to the coordinates of the virtual pointer (x_virtual, y_virtual) on the computer screen. Let's call this transformation function T.

So, the mathematical model can be represented as follows:

Image Processing:

The webcam captures the image of the user's hand (x_cam, y_cam). Image processing algorithms extract the relevant features (x_hand, y_hand) from the captured image.

Transformation Function:

$$T: (x\_hand, y\_hand) \rightarrow (x\_virtual, y\_virtual) \quad (1)$$

Equation (1) is the transformation function T takes the coordinates of the hand in the image and maps them to the coordinates of the virtual pointer on the computer screen.

The specific details of the transformation function T will depend on the algorithms and techniques used for object tracking and mapping the hand position to the screen position. Common approaches may include calibration of the camera, perspective transformation, and homography to establish the mapping between the hand position and the screen position.

Interaction with the Computer:

Number Once the transformation function T maps the hand position to the virtual pointer position, the computer system responds to the new pointer coordinates (x_virtual, y_virtual) as if it were receiving inputs from a physical mouse.

The mathematical model provides a framework to understand how the image captured by the webcam is translated into the position of the virtual pointer on the computer screen. The implementation of the transformation function T may involve various computer vision and AI techniques, and its accuracy and efficiency will be critical to the success of the Virtual Mouse application.

VIII. CONCLUSION

The primary aim of the AI virtual mouse system is to replace the traditional physical mouse with hand gestures for controlling mouse cursor functions. This system employs a webcam or built-in camera to detect and process hand gestures and hand tips, enabling the execution of specific mouse functions.

According to the model's outcomes, it is evident that the proposed AI virtual mouse system performs remarkably well and demonstrates higher accuracy compared to existing models. Furthermore, it effectively addresses most of the limitations observed in current systems. The enhanced accuracy of the proposed model renders it suitable for real-world applications, offering a touchless alternative to mitigate the spread of COVID-19, as users can manipulate the virtual mouse through hand gestures without physical contact.

Nevertheless, the model does have certain limitations, such as a slight decline in accuracy when performing right-click mouse functions and difficulties in accurately clicking and dragging to select text. To tackle these limitations, future endeavors will concentrate on refining the fingertip detection algorithm to produce more precise and dependable results.

In conclusion, the AI virtual mouse system, featuring hand gesture control, presents significant potential as an innovative and hygienic input method. Its superior performance and applicability for real-world scenarios mark it as a compelling advancement in human-computer interaction. As researchers continue to refine and optimize the model, it stands to revolutionize the way users interact with computers and contribute to a safer and more efficient technological experience.

IX. ACKNOWLEDGMENT

The authors extend their appreciation to Shubhangi Said, who served as the research guide and Head of the Department of AIDS at Jai Hind College of Engineering, Pune, affiliated with Savitribai Phule Pune University. Shubhangi's invaluable guidance, support, and insights significantly contributed to the success of this research.

Furthermore, the authors wish to acknowledge Anand Khatri, Head of the Department of Computer Science at Jai Hind College of Engineering, for his continuous encouragement and support throughout the research process. Anand's expertise and mentorship have played a crucial role in shaping this work.

The authors are also thankful to Jai Hind College of Engineering, Pune, and Savitribai Phule Pune University for providing the research environment and resources necessary to conduct this study.

This research was made possible through the collaborative efforts and support of these individuals and institutions, and their contributions are sincerely acknowledged.

REFERENCES

[1]. N. Rzaik, C. Dehos, M. Zarudniev, A. Siligaris, and J. L. J. Gonzalez, "A mmW RadCom System for Short-Range Hand Gesture Sensing and Data Synchronization," 2023 18th Conference on Ph.D Research in Microelectronics and Electronics (PRIME), Valencia, Spain, 2023, pp. 13-16.

[2]. S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, July-Sept. 2022.

[3]. M. Devi, A. Chakraborty, A. Roy, and D. Majumder, "Single-hand Gesture Recognition of Manipuri Classical Dance of India based on Skeletonization Technique," 2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC), Silchar, India, 2023, pp. 1-7.

[4]. E. Solly and A. Aldabbagh, "Gesture Controlled Mobile Robot," 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Istanbul, Turkiye, 2023, pp. 1-6.

[5]. S. V. Sonekar, H. Dhoke, V. Mate, S. Dhewle, and M. Patil, "Real-Time Sign Language Identification using KNN: A Machine Learning Approach," 2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP), Nagpur, India, 2023, pp. 1-4.

[6]. T. M P, M. R. Bharamagoudra, S. Z. Noorain, and S. Maria SP, "Hand Gesture Detection using Transfer Learning with Deep Neural Networks," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-5.

[7]. M. F. Nurnoby and E. -S. M. El-Alfy, "Multi-culture Sign Language Detection and Recognition Using Fine-tuned Convolutional Neural Network," 2023 International Conference on Smart Computing and Application (ICSCA), Hail, Saudi Arabia, 2023, pp. 1-6.

[8]. M. S. Reddy, S. Kukkamudi, R. Kunda, and T. Mamatha, "Virtual Mouse Using Hand Gesture," 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), Chickballapur, India, 2022, pp. 1-5.

[9]. Godbole, V. Gondke, and K. K. Devadkar, "Hand Gesture Identification and Voice Command Based Hardware Reduction," 2022 IEEE Pune Section International Conference (PuneCon), Pune, India, 2022, pp. 1-6.