Intuitive Presentation Using Hand Gestures: A Vision Based Approach

O Vasantha Kumari
Dept. of Computer Science and Engineering (Asst. Prof.)
Dr. M.G.R. Educational and Research Institute
Chennai, India
vasanthakumaricse@gmail.com

Venkata Sai Teja K.
Dept. of Computer Science and Engineering (Student)
Dr. M.G.R. Educational and Research Institute
Chennai, India
saitejak1312@gmail.com

B Matthew Mineeth
Dept. of Computer Science and Engineering (Student)
Dr. M.G.R. Educational and Research Institute
Chennai, India
minithbmatthew@gmail.com

Kishan Kumar Parida
Dept. of Computer Science and Engineering (Student)
Dr. M.G.R. Educational and Research Institute
Chennai, India
kishankumar060802@gmail.com

***Abstract*—Human-machine interaction for current presentation techniques heavily relies on external controllers like clickers and remotes. While these traditional methods eliminate the need for manual device control, they still require physical interaction to manage presentation slides, disrupting the presenter's flow. This paper introduces a novel hand gesture recognition system aimed at reducing the reliance on external controllers and enhancing overall smoothness and accuracy, thereby improving the audience and user experience during presentations. Our approach utilizes real-time image processing through machine learning, integrated with deep learning modules, to identify and recognize hand gestures for performing various tasks and controlling the presentation.**

***Index Terms*—hand gesture recognition, Machine Learning, Deep learning, Enhancing performance**

I. Introduction

Imagine yourself attending a captivating and engaging presentation. The speaker is charismatic, the content is engaging, but then... they fumble with a bulky clicker, the lights flicker as they switch slides, and precious momentum is lost. Frustration ripples through the audience as outdated technology causes a loss of immersion. Unfortunately, this scenario is all too common in the present world.

"Slides are crutches, but crutches can become prisons if you rely on them too much." Garr Reynolds, author of "Presentation Zen,"

Presentations are used almost everywhere, like in education, business, public speaking, etc. Therefore, it's very important that we use the right methods to present our case. Using intuitive methods can help maintain engagement between the audience and the presenter without killing momentum through clicker-flicker and keyboard shortcut interruptions.

In today's world, where AI is taking over almost everything, it makes things easier and more engaging. We decided to make our approach based on machine learning. Our system uses machine learning to control the presentation flows without the need for any physical controlling devices like clickers or keyboards, using only our hand gestures to control the navigation of the presentation slides. This approach helps us avoid all the disadvantages of traditional presentation techniques.

Our system uses image recognition to use real-time image processing to interpret the hand gestures performed by the presenter. In our current system, we use left and right swipes to navigate between slides and specific finger configurations to trigger additional functionalities, like using a two-finger gesture to trigger an independent pointer and a single-finger gesture, which works as a free-style marker.

There are projects that use hand gesture recognition to do simple commands like swiping your hand over your phone to wake it up, etc., but all of these projects use hand gestures to do simple tasks, which does not fully explore the potential of using hand gestures. Through this project, we also hope to explore the boundaries of what we can achieve with simple hand gestures, hopefully expanding it beyond presentation in the future.

In conclusion, using this method, we can smoothly transition between controlling our presentation slides and giving our speech. In the future, we hope to integrate AI so we can effectively tell when the presenter is using hang gesturing to emphasize his speech and when he is using hand gestures to control the presentation. We also hope to use the same hand gesture to go beyond the presentation software and control some simple computer commands without the use of a mouse or keyboard.

So effectively, after the full completion of this project, we can give a seamless presentation where we give our speech while the slides change with just a small flicker of our hand movement without interrupting our lecture or killing the momentum and engagement of the audience.

II. Existing work

In the ever-changing IT industry, a slew of methods are being discussed to enhance the functionalities of presentation software and presentation techniques. All these studies, however, don't deal with using hand gestures for presentation; they are mainly focused on improving functionalities in presentation software using AI and the latest feature, UI improvements.

A thorough analysis of the present studies and methods gives you a plethora of presentation methods. Some are even using voice commands to control the presentation flow. Several significant methods are being discussed that involve varying technologies to make the presentation techniques easier. Some presentation controls used in the present world are:

A. Keyboard Shortcuts

- You can move through slides in most presentation software using keyboard shortcuts.

B. Mouse/Clicker

- Using a mouse or a presentation clicker or remote control is a common method. Clicking the left mouse button or using the clicker's buttons typically advances to the next slide, and right-clicking or using other buttons might move to the previous slide.

C. Navigation Panel

- Many presentation tools have a navigation panel or slide sorter view that allows you to see all the slides at once. You can click on a specific slide in the panel to jump directly to it.

D. Presenter View

- Advanced presentation software often includes a "Presenter View" that shows different information on the presenter's screen compared to what the audience sees. It typically includes slide thumbnails, speaker notes, and a timer. The presenter can navigate through slides independently of what the audience sees.

E. Voice Commands

- Some presentation software supports voice commands for navigation. This allows presenters to verbally instruct the software to move to the next or previous slide.

F. Interactive Features

- In some modern presentation tools, especially for online or interactive presentations, you might have interactive features like clickable links or buttons within the slides that control navigation.

G. Remote Control Apps

- For presentations conducted on tablets or smartphones, remote control apps can be used to control slide navigation. These apps turn your device into a remote control, allowing you to move through slides.

G. Custom Animation and Transitions

- You can set up custom animations and transitions between slides to control the flow of information. This is often used for adding visual effects or revealing content gradually.

Always familiarize yourself with the specific features and functionalities of the presentation software you are using, as the methods for controlling slide navigation can vary between tools. Popular presentation software includes Microsoft PowerPoint, Google Slides, Apple Keynote, and others.

Despite the advances made in the field, problems remain. Using all the existing methods gives us limited interactivity, lack of engagement, dependency on equipment, limited accessibility, difficulty in collaboration, lack of presenter feedback, and compatibility issues. By expanding on previous studies, we hope to improve our understanding of these problems. We hope to solve these problems by using our software, which uses physical hand gestures to control the presentation.

We will elaborate on our methodology in the following sections of this paper, detailing the application of the algorithms used. We hope to provide a thorough and forward-thinking contribution to the subject of presentation control methods.

III. Algorithms

Using cutting-edge machine learning algorithms is essential for precise control, enhanced interaction, and efficient development. The cvzone. HandTracking module in the system uses various algorithms for hand detection, including skin color segmentation as a preliminary step and also blob analysis with landmark detection. Let's delve into some potential methods used.

A. Skin Color Segmentation

Skin Color Segmentation is threshold based on a predefined skin color range in HSV or YCbCr color space. It also sets an adaptive threshold based on local statistics to handle lighting variations and skin chrominance models using CrCb space and skin region thresholds.

- Skin color segmentation is the process of identifying and separating skin pixels from other pixels in an image.
- HSV stands for hue, saturation, and value. Hue represents the color itself, saturation represents the intensity of the color, and value represents the brightness of the color.
- Skin colors typically have a hue in the range of 0–30 degrees, a saturation in the range of 0.2–0.8, and a value in the range of 0.3–1.0.

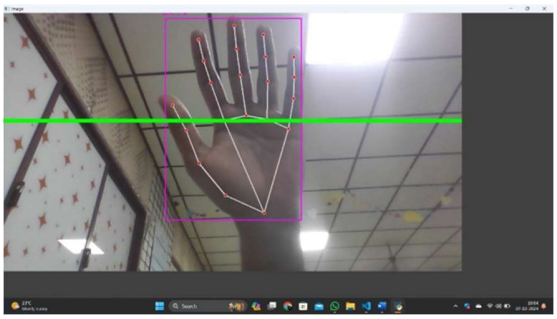

Fig. 1. Hand Gesture Recognition

B. blob Analysis

Blob analysis is a computer vision technique used to identify and extract features from connected regions of pixels in an image. And these blobs are typically defined as regions of pixels that are brighter or darker than their surroundings.

Blob analysis can be used to identify and track objects in a video sequence or to count the number of objects in an image. The blob analysis can be applied directly or indirectly, using the following concepts:

1) Keypoint Detection: Keypoint Detection uses feature detectors like Harries Corners or DoG to identify potential landmark locations.

Key to the success of this vision-based approach is the utilization of feature detectors such as Harris corners or Difference of Gaussians (DoG). These detectors play a crucial role in identifying potential landmark locations on the human hand, enabling the system to recognize and interpret various gestures accurately.

- Harris Corners: Harris corner detection is a widely-used technique in computer vision. It identifies points in an image where there are variations in intensity in multiple directions, making it ideal for locating key features in an image. In the context of hand gesture recognition, Harris corners can be employed to pinpoint distinctive landmarks on the hand, facilitating precise tracking and gesture interpretation.

- Difference of Gaussians (DoG): DoG is another powerful feature detection technique that involves taking the difference between blurred versions of an image. This method is particularly effective for identifying key points in images with varying levels of intensity. In the context of hand gesture recognition, DoG can be applied to detect subtle variations in hand movements, enhancing the system's ability to interpret intricate gestures accurately.

IV. Tools and Libraries

A. Python

Python is a popular high-level, interpreted, dynamically typed programming language that is easy to understand and use. Applications including web development, data analysis, automation, artificial intelligence, and machine learning make extensive use of its versatility. Python's clear and simple syntax prioritizes the readability of code, which lowers program maintenance costs.

B. CvZone

Pandas is an open-source Python toolkit for data manipulation and analysis. The Data Frame, a two-dimensional table used to store structured data, is essential to Pandas. It offers essential data structures, such as data frames and series, that facilitate effective data analysis and manipulation. Pandas is a flexible tool for data loading because it can read data from a variety of file formats, including Excel and CSV.

C. NumPy

An open-source Python library for numerical computation is called NumPy. It offers mathematical functions to manipulate these arrays as well as the necessary tools for working with sizable, multi-dimensional arrays and matrices. NumPy is a core Python library for scientific computing and data processing, with an emphasis on performance and economy.

D. Streamlit

With the help of the open-source Python library streamlit, online applications for data science and machine learning may be quickly and easily created. It makes the process of creating interactive web apps out of data scripts easier. Streamlit's simple and intuitive API makes it easy for users to create web apps quickly and effectively. Automatic widget creation, real-time updates, and the seamless integration of interactive elements like tables and charts are some of the key features.

E. Mediapipe

MediaPipe is a tool for analyzing sensory data, mostly visual data. It is an open-source, cross-platform framework used for building machine learning pipelines that process various media inputs to accomplish tasks like real-time hand tracking, facial recognition, and pose detection. MediaPipe provides in-built machine learning tools for building custom pipelines. There are three prebuilt libraries for vision, text, and audio. Based on the app that uses it, import a vision, text, or audio library. For hand gesture detection, the MediaPipe Tasks vision module needs to be imported.

F. Aspose.slides module

Aspose.slides is a class library that enables you to read or write any PowerPoint documents. Aspose.slides offers various unique features, such as:

- Loading, opening, and viewing presentations
- Editing the presentations
- Converting presentations to PDF, Word, JPG, PNG, HTML, GIF, and many other formats
- Rendering and Printing presentations
- Encrypting and Manipulating presentation

V. Flow Chart

In many institutions, conferences, and other domains, the usual presentation is done through a remote device to control the slides of the presentation, but this approach allows the user to manipulate the slides of the presentation through hand signs or hand gestures and for drawing annotations and writing on the slides with hand gestures. This approach improves human-machine interaction.

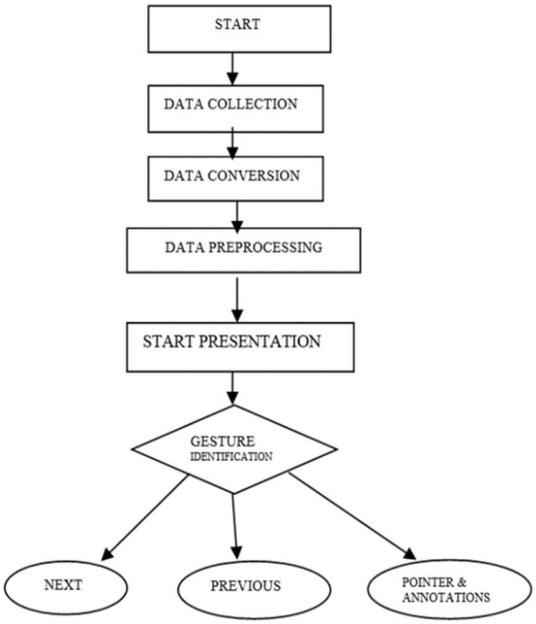

Fig. 2. Flow chart hand gesture controlled presentation

A. Data Collection

The collected data is in the form of visual data, which is inputted through the system's integrated camera and the external webcam or any other source. The cvzone's hand tracking module allows the inputting of the data and keeps capturing the visual data from time to time in the form of video. In this case, the hand's data is tracked in real-time.

B. Data Conversion

In this data conversion, the main functionality is to convert the uploaded presentation file to PNGs or other formats like JPGs or JPEGs.

C. Data Preprocessing

In order to prepare data for analysis and model training, a Python module is imported, which is MediaPipe. MediaPipe is a machine learning integrated library that analyzes sensory data and processes it by creating pipelines for it. real-time hand tracking and pose detection in order to understand the behavior of the hand and its pose.

VI. Implementation

A. Designing User-Interface

It is not practical to execute the code multiple times for each execution, and in order to address such conflicts, a user-friendly interface is employed. To implement the suggested theory, the most straightforward approach for creating the user interface (UI) is by utilizing a Python library known as Streamlit. We can use this library to develop web applications that are driven by data. The following figure illustrates a basic web-based UI designed for intuitive control of presentations.

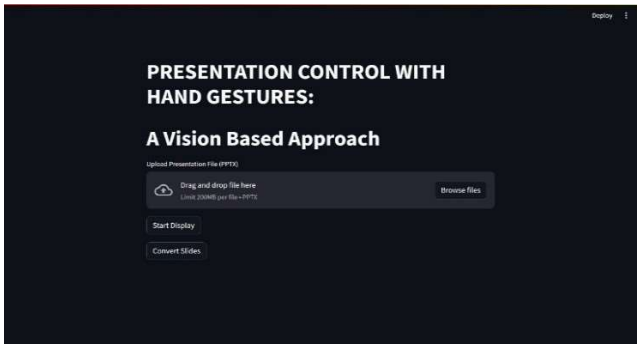

Fig. 3. streamLit user-interface

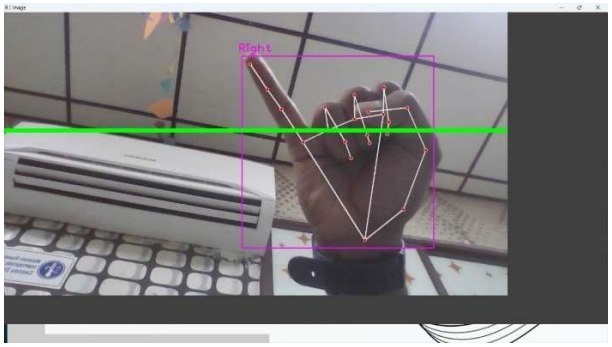

Fig. 4. Hand gesture control presentation to change slide to next

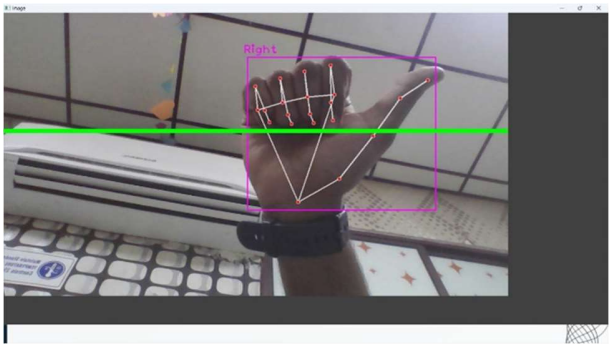

Fig. 5. Hand gesture control for presentation to change slide to previous

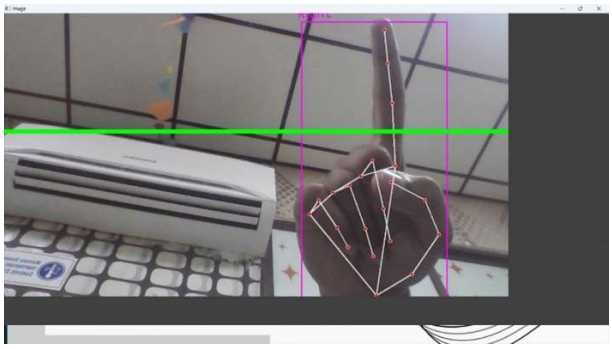

Fig. 6. Hand gesture for controlling pointer during presentation

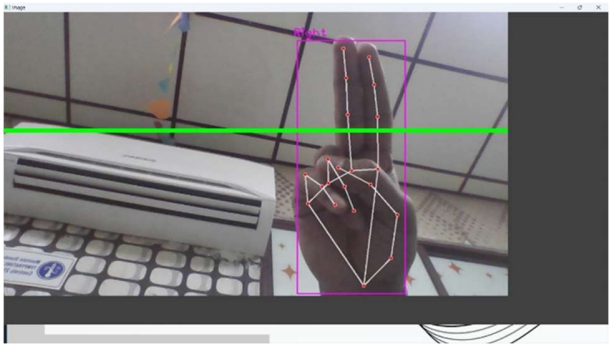

Fig. 7. Hand gestures to draw annotations on the slide

B. Experimental Results:

a. **Gesture Recognition Accuracy:** The project achieves high accuracy in recognizing user gestures, with minimal false positives and false negatives. Evaluation metrics, such as precision, recall, and F1 score, are calculated to assess the gesture recognition performance.

b. **Slide Control Latency:** The latency between detecting a gesture and performing the corresponding action (e.g., advancing or reversing slides) is measured to evaluate the responsiveness of the system. Low latency indicates smooth and responsive slide control during presentations.

c. **Annotation Precision:** The precision of the annotation feature is evaluated by comparing the drawn annotations with the user's intended gestures. The project accurately captures and represents the user's annotations on the presentation slides, enhancing interactivity and engagement.

VII. Conclusion

A vision-based approach represents a transformable leap in the realm of presentation dynamics. The utilization of computer vision, specifically through keypoint detection using feature detectors like Harris corners and DoG, marks a significant departure from traditional control methods. By incorporating hand gestures as an intuitive interface, the system not only enriches the user experience but also eliminates the need for conventional input devices, fostering a more natural and engaging interaction with presentation content.

The integration of sophisticated algorithms, such as those facilitating gesture mapping, real-time feedback, and user authentication, contributes to the system's adaptability, versatility, and security. Its scalability and compatibility across different platforms ensure widespread applicability, and seamless integration with popular presentation software streamlines the user experience. Beyond the realm of convenience, this project promotes inclusivity by providing a user-friendly and accessible means of presentation control.

The integration of sophisticated algorithms, such as those facilitating gesture mapping, real-time feedback, and user authentication, contributes to the system's adaptability, versatility, and security. Its scalability and compatibility across different platforms ensure widespread applicability, and seamless integration with popular presentation software streamlines the user experience. Beyond the realm of convenience, this project promotes inclusivity by providing a user-friendly and accessible means of presentation control.

The incorporation of user training and on-boarding features ensures a smooth transition for presenters, making the technology approachable and easy to adopt. In essence, "Intuitive Presentation Control using Hand Gestures: A Vision-Based Approach" not only represents the current pinnacle of technology but also sets the stage for future advancements in human-computer interaction, promising more intuitive and immersive presentation experiences.

The implementation of advanced algorithms for gesture mapping, real-time feedback, and user authentication ensures adaptability, versatility, and security. The system's scalability and compatibility across platforms make it universally applicable, while seamless integration with popular presentation software streamlines the user experience. Beyond convenience, the project champions inclusivity by providing a user-friendly and accessible means of presentation control. User training and on-boarding features facilitate a smooth transition for presenters, making the technology approachable and easy to adopt.

In essence, "Intuitive Presentation Control using Hand Gestures: A Vision-Based Approach" not only represents the current pinnacle of technology but also sets the stage for future advancements in human-computer interaction. It promises a more intuitive and immersive era of presentation experiences, where the fusion of vision-based control and user-centric design creates a seamless and captivating presentation environment.

REFERENCES

[1] N. Sharma and S. Mittal “Hand Gesture Recognition for PresentationControl (2021)”

[2] D. Desai, V. Panchal and M. Patel “Smart Presentation System UsingHand Gesture (2022)”.

[3] S. Dey, S. Das and S. Mondal “Automated Digital Presentation Control using Hand Gesture Technique (2022)”

[4] Y. Somani and R. Verma “Smart Presentation Control by Hand Gestures Using Computer Vision and Google’s Mediapipe (2023)”

[5] V. Bhumika and N. Jain “Hand Gesture Controlled Presentation usingOpenCV and Mediapipe (2023)”

[6] M. Kumar and D. Misra “A Robust Real-Time Hand Gesture Recogni- tion System for Presentation Control (2021)”

[7] L. Gupta and R. Aggarwal “Vision-Based Hand Gesture Recognition forPresentation Control Systems: A Survey (2020)”

[8] Y. Zhang, J. Liu and W. Zheng “Towards Natural and Intuitive Presen-tation Control using Hand Gestures and Object Recognition (2022)”