JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Leveraging Hand Gesture Recognition and LLM for Developing a Non-Contact, Real-time Virtual, Immersive Teaching Aid

[1]Dr. Seema Kolkur, [2] Yashas Vaddi, [3]Nimish Tilwani, [4]Madhava Ved

[1]Associate Prof, [2]Student, [3]Student, [4]Student
[1] Department of Computer Engineering,
[1]Thadomal Shahani Engineering College, Mumbai, India

Abstract : In today's digital age, where technology continues to blur the lines between the physical and virtual worlds, the way humans interact with computers is undergoing a significant transformation. Hand-tracking technology, which has seen substantial advancements in recent years, offers a more natural and engaging way for users to interact with digital environments. By leveraging this technology, the proposed system aims to transcend traditional input methods, such as mouse and touch, which can often feel restrictive and less engaging. In addition, adding Large Language Models (LLM) to effectively process the input marks it's possible use in various applications. This paper addresses the challenge of developing a robust and accessible hand-tracking based drawing application that works efficiently with widely available hardware, such as webcams**.**

***IndexTerms* - Computer Vision, Hand Gesture Recognition, LLM.**

I. Introduction

With the advent of technology, people are spending a lot of time online and perform multiple activities in the virtual world. With this, the interaction of human with machines has undergone a dramatic change. Traditionally, humans used input devices like keypad, mouse, touch screen etc. for effective communication with machines. Over the time, non-contact inputs like as voice, facial expressions, physical motions, and gestures are used as inputs for many applications [1]. With vision-based techniques, a camera is used to provide contactless communication between humans and machines. Computer vision cameras are widely available, inexpensive and easy to use. Hand gestures is an effective communication method which is why many researchers worked on hand gestures recognition for various applications [2]. Hand-tracking technology, which allows users to use their natural gestures to interact with digital devices, holds the promise of creating a more engaging and accessible user experience. Such interfaces feel more intuitive and engaging, making it easier for users to interact with digital content.

This paper explores a way of developing a robust and accessible hand-tracking based teaching aid application. The application is designed to capture real-time video from the webcam, process the video feed to detect and track hand movements, and translate those movements into drawing actions on a virtual canvas. This approach not only enhances the user experience but also opens up new possibilities for interactive applications across various domains, including education, entertainment, and creative arts.

II. Literature Review

Hand gesture recognition is a widely used technology in various applications in the past. In [3], authors used two colored markers on the tips of the fingers to generate eight hand gestures. These movements were detected and fed as instructions to computer for specific tasks. [4] showed use of a data glove with ten sensors to collect gesture data. It measured the joint angles of five fingers as main parameter. In a system presented in [5], a red-colored LED light source was attached to the finger of the user. Hand movements were tracked by tracking light movements. But system had basic assumption that user writes only English alphabets which will be recognized, send to text editor and also that there is no other red color object in the focus of webcam. [6] used Ultrasonic Distance Sensors to identify hand gestures. The distance between the hand and the distance sensor was used to record the gestures. [7] used a skeleton based model which used the geometric shape of the hand along with hand skeleton linked joints. [8] used Neural Network where contour of the hand is used as the feature to recognize static hand gestures in an image or video for Nepali Sign Language (NSL) recognition. [9] used Leap Motion Controller, which can track hands for creating a painting application. [10] demonstrated a glove with an attached ping pong ball to track the hand gestures and write on the screen. Most of these systems heavily relied on some sensor or physical device as an additional hardware. [11] created a virtual paint application by using computer vision and hand tracking. [12] used Convolution Neural Networks is for detecting and recognizing drawn gesture in front of camera specifically for English text.

III. METHODOLOGY

The proposed application is developed as an educational aid where a user can make a drawing or write some text as he/she would write on a black board in a class room setting. The advantage of the system is; it does not need any specialized hardware or sensors for the same. Also it is contact less, virtual and can be used from anywhere. A user standing in front of a webcam can just start writing in air i.e. a virtual canvas provided by the system. Overall architecture of the application is both robust and adaptable, ensuring that the application performs well under various conditions. The working flow of the proposed virtual board application is shown in Fig 1. Based on index finger movements of either hand, a user can draw some shapes or text on virtual canvas. The final image is saved in jpeg format. If the user wants to process the text from hand-written image, the saved image is passed on to an OCR module. This module extracts text from the input image and saves it in a text editor for future usage. If an input image contains a simple mathematical equation, user may wish to execute it and find result. In such a case the extracted text is sent to chatGPT, which evaluates it and gives result. ChatGPT is an AI-powered chatbot that uses an LLM to understand and generate human-like text.

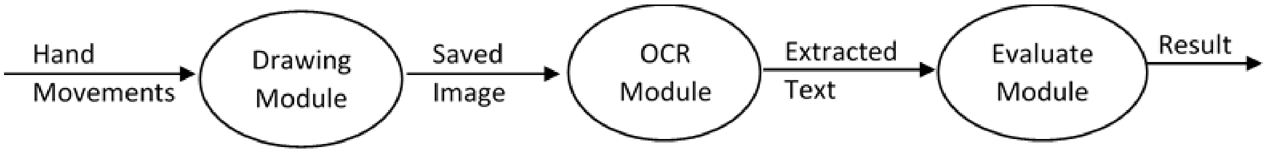

Fig. 1 Workflow of Proposed Application

Drawing module is the heart of this application. Figure 2 gives additional details to the working of drawing module. The key components are:

1. Video Capture: The system captures live video feed from a standard webcam. This video feed serves as the primary input for the application. The system captures frames from the video, flips them for a mirror effect, and converts them to RGB format for hand tracking. This preprocessing step ensures that the video feed is suitable for hand landmark detection.
2. Hand Tracking: The Mediapipe library is used to detect and track hand landmarks in the video feed. Mediapipe is a powerful tool that provides high-fidelity tracking of hand movements in real-time [13].
3. Gesture Recognition: The system identifies specific hand gestures, such as drawing, colour selection, and canvas clearing. This is achieved by analyzing the hand landmarks detected by Mediapipe.
4. Drawing Module: Based on the detected gestures, the system renders lines and shapes on the virtual canvas. The drawing module ensures that the hand movements are accurately translated into drawing actions. Drawing points are stored in deques for different colours, allowing efficient addition and removal of points. This ensures smooth and continuous drawing. Colour selection and canvas clearing are implemented based on the position of specific fingers in the hand landmarks. This allows users to control the application using simple and intuitive gestures. To complete the drawing a user can touch index finger and thumb.
5. Frame Rendering: The completed Canvas image is saved to the storage device. The images are saved or screenshot is taken when the program detects both the hands on the screen.

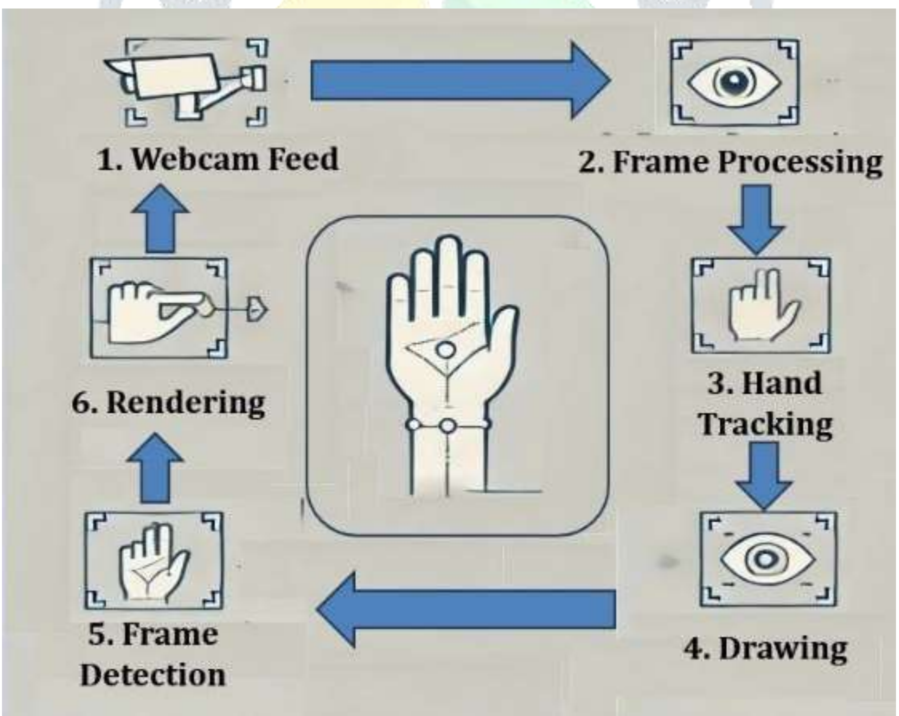

Fig. 2 Detailed working of Drawing Module

IV. RESULTS

Figure 3 shows operational mode of the drawing module. It clearly indicates hand landmark points as detected by mediapipe. It shows that a user has selected blue color and has started drawing on the virtual canvas. Figure 4 depicts a sample input a user can provide on the virtual board. Figure 5 shows sample image captured by drawing module. As the text contains a mathematical equation, this string “WHAT IS?” is added to the image. This image is passed to OCR module, the extracted text is displayed as shown in

Fig.56. This image is then passed on to Evaluate module i.e. ChatGPT. ChatGPT is turn calculates and returns result as shown in Fig.7.

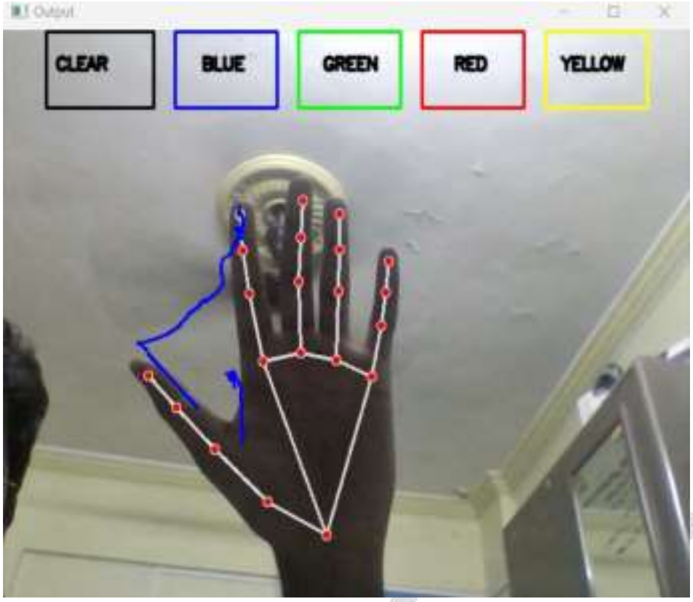

Fig. 3 Operational Drawing Module

Fig. 4 User input on Virtual Board

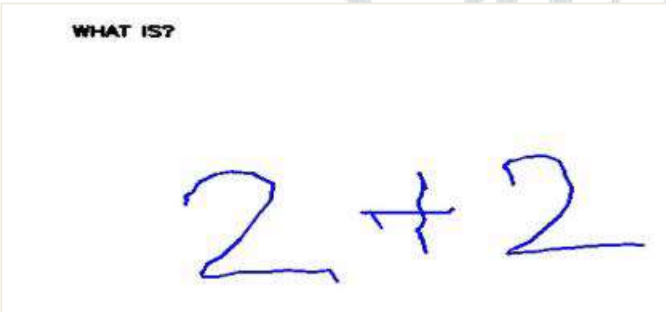

Fig. 5 Sample Image Captured

Fig. 6 Text Extracted by OCR Module

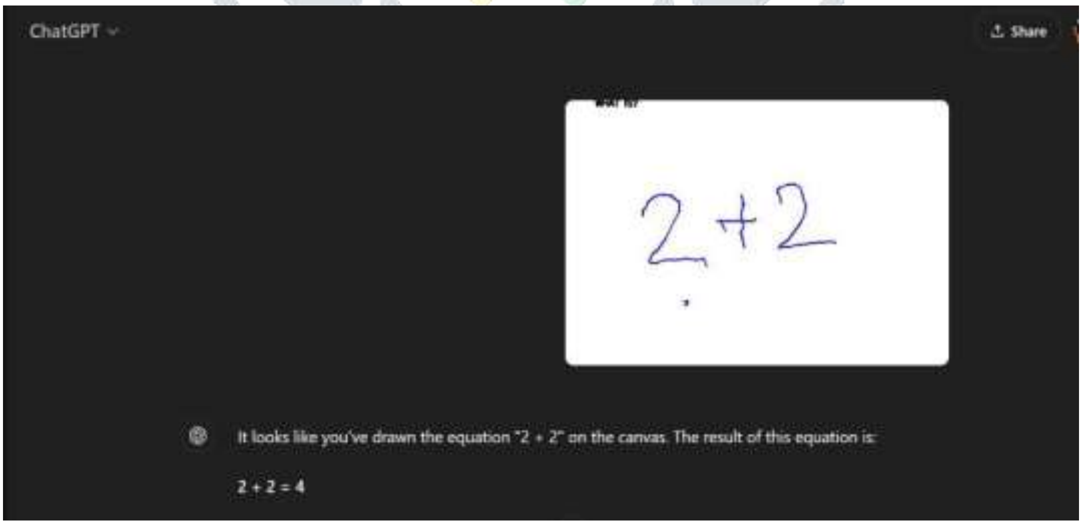

Fig. 7 Result from ChatGPT

IV. Conclusion And Future Work

This paper demonstrates a hand-tracking based drawing application using computer vision and a LLM based ChatGpt model. It provides an intuitive and interactive user experience, allowing users to draw on a virtual canvas using natural hand gestures. It can be used as a teaching tool with basic shapes, colors and also as a simple calculator. The input drawing is saved as a Jpeg image. If the input has text rather than drawing it is saved with a text editor. Results for simple calculations ae computed quickly. By leveraging widely available hardware and open-source software, this application becomes more accessible and practical for everyday use. While it is very feasible and effective, there are several areas for future improvement and development:

- Enhancing Gesture Recognition Accuracy: Future work could focus on improving the accuracy of gesture recognition by incorporating more advanced machine learning techniques and training the system on a larger dataset of hand gestures.
- Adding More Gestures: Additional gestures could be implemented to expand the functionality of the application. For example, gestures for drawing different shapes, erasing, or adjusting brush sizes could be added.
- Multi-User Support: Future versions of the application could support multiple users, allowing collaborative drawing sessions. This would involve detecting and tracking multiple sets of hand landmarks simultaneously.
- Extending to Other Platforms: The application could be extended to support other platforms, such as mobile devices and tablets, providing greater flexibility and accessibility.
- Integrating with Other Applications: Integrating the hand-tracking drawing application with other creative and educational tools could enhance its utility and appeal. For example, it could be used in conjunction with digital art programs, virtual whiteboards, or interactive learning environments.

REFERENCES

[1] Dix, A., Finlay, J., Abowd, G., & Beale, R. (2003). Human-Computer Interaction: 3rd Edition. Prentice-Hall.
[2] Isard, Michael & Maccormick, John. (2000). Hand Tracking for Vision-Based Drawing
[3] Siam, Sayem & Sakel, Jahidul & Kabir, Md., Human Computer Interaction Using Marker Based Hand Gesture Recognition., 2016
[4] M. Lee and J. Bae, "Deep Learning Based Real-Time Recognition of Dynamic Finger Gestures Using a Data Glove," in IEEE Access, vol. 8, pp. 219923-219933, 2020, doi: 10.1109/ACCESS.2020.3039401.
[5] P. Ramasamy, G. Prabhu and R. Srinivasan, "An economical air writing system converting finger movements to text using web camera," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), 2016, pp. 1-6, doi: 10.1109/ICRTIT.2016.7569563.
[6] S. Belgamwar and S. Agrawal, "An Arduino Based Gesture Control System for Human-Computer Interface," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-3, doi: 10.1109/ICCUBEA.2018.8697673.
[7] Q. De Smedt, H. Wannous and J. Vandeborre, "Skeleton Based Dynamic Hand Gesture Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, pp. 1206-1214, doi: 10.1109/CVPRW.2016.153.
[**8]** Acharya, Manoj & Pant, Dibakar. (2016). Computer Vision Based Hand Gesture Recognition For Speech Disabled Persons. Journal of the Institute of Engineering. 11. 30. 10.3126/jie.v11i1.14693.
[9] R. Lyu et al., "A flexible finger-mounted airbrush model for immersive freehand painting," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 2017, pp. 395-400, doi:10.1109/ICIS.2017.7960025.
[10] Prajakta Vidhate, Revati Khadse, Saina Rasal, “Virtual paint application by hand gesture recognition system”, 2019 International Journal of Technical Research and Applications, 2019, pp. 36-39
[11] Niharika M,Neha J, Mamatha Rao, Vidyashree K VIRTUAL PAINT APPLICATION USING HAND GESTURES, International Research Journal of Engineering and Technology (IRJET) Volume: 09 Issue: 04 Apr 2022
[12] V. Joseph, A. Talpade, N. Suvarna and Z. Mendonca, "Visual Gesture Recognition for Text Writing in Air," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 23-26, doi: 10.1109/ICCONS.2018.8663176.
[13] Hand landmarks detection guide https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker last accessed on 18 Nov 24