[1] Ammar Aldallal
[2] Qassim Abdali

Smart Presentation Management Using Gesture Recognition and Machine Learning

Journal of Electrical Systems

Abstract: - With the growing reliance on visual information, computer vision offers solutions to complex problems and enhances user experiences across multiple domains. This paper focuses on integrating hand gestures into software applications to improve human-computer interaction. The advancement of gesture recognition systems is vital for various applications, including gaming, virtual and augmented reality, assisted living, cognitive development assessment, and industrial uses like human-robot interaction and autonomous vehicle control. This research presents a system for controlling PowerPoint presentations using hand gestures detected through computer vision. Utilizing OpenCV and MediaPipe, the system analyzes video input from a webcam to track hand movements in real-time, converting specific gestures into commands. The proposed system offers an intuitive and engaging alternative to traditional input devices, allowing presenters to navigate slides, annotate content, control a pointer, and apply zoom using natural hand gestures. The implementation demonstrates the system's effectiveness in accurately recognizing hand gestures and triggering corresponding actions, validated through practical examples and high accuracy in various functions. This approach addresses the high cost of smart boards and the need for more interactive presentation tools, particularly in educational settings, enhancing engagement and accessibility.

Keywords: About smart hand gesture, PowerPoint presentations, computer vision, OpenCV, MediaPipe, machine learning.

I. INTRODUCTION

The Computer Vision has gained significant importance as a technology that enables machines to extract meaningful information from visual data. It plays a crucial role in various fields, including autonomous vehicles, robotics, healthcare, and human-computer interaction. As the reliance on visual information continues to grow, computer vision has become a key solution for addressing complex problems and improving user experiences in multiple fields.

Recently, computer scientists have employed various computational algorithms and methods to address problems and simplify our lives. Integrating hand gestures into software applications has significantly enhanced human-computer interaction. The advancement of gesture recognition systems is crucial for the development of these interactions, and the use of hand gestures is becoming increasingly common across different fields. Hand gesture applications are now evident in gaming [1], virtual reality [2, 3], augmented reality [4], assisted living [5], and cognitive development assessment [6]. Additionally, the latest developments in hand gesture recognition have attracted industrial interest, particularly for human-robot interaction in manufacturing [7, 8] and autonomous vehicle control [9].

One area where computer vision has made notable strides is interactive interfaces that utilize natural gestures and movements. Traditional input devices, such as keyboards and mice, can be less intuitive, particularly in scenarios such as presentations. Recognizing this challenge, this work aims at developing a system the controls Power Point presentation using hand gesture using computer vision. Where it capitalizes on the capabilities of computer vision libraries, such as OpenCV and MediaPipe, to analyze video input from a webcam and track hand movements in real-time by detecting specific gestures and translating them into commands.

The motivation behind this work is the need for a more intuitive and engaging means of interacting with presentation content. This proposed system eliminates the reliance on traditional input devices, allowing presenters to navigate slides, annotate content, control a pointer, and apply zoom using natural hand gestures.

This work addresses specific challenges in the field of education. Firstly, the high cost of smart boards is a barrier to their widespread use in schools. The project aims to develop a cost-effective alternative that provides interactive presentation tools, making them more accessible to schools with limited budgets. Secondly, traditional presenting methods often lack interaction between the teacher and the content, resulting in less engaging lessons. By

[1] Department of Telecommunication Engineering, Ahlia University. aaldallal@ahlia.edu.bh

[2] Department of Computer Engineering, Ahlia University. 201710207@ahlia.edu.bh

Copyright © JES 2024 on-line : journal.esrgroups.org

introducing features such as controlling the content and drawing on them, this work aims to enhance the level of engagement between the teacher and the content and creates more exciting learning and teaching experiences. Lastly, the project also supports teachers by providing additional tools that can assist in delivering the idea such as pointing on the important content and zooming on them. The proposed system works on wide range of devices.

The remaining of this paper is organized as follows: Section 2 reviews similar work done on hand gesture and its application and technologies applied so far. Section 3 presents the methodology and system components. Section 4 presents the system implementation: How the mentioned components in section 3 has been used in association with the software to build the system. Finally, is the conclusion which summarizes the findings and results, highlighting the future work that can further improve the system.

II. Related Work

Raindrop the main computer vision libraries adopted in this study are OpenCV and MediaPipe. which are adopted by several studies for various computer vision tasks as follows:

OpenCv and MediaPipe

OpenCV (Open Source Computer Vision Library) is an open-source computer vision and machine learning software library. It provides a comprehensive set of tools and functionalities to perform various computer vision tasks such as image processing, object detection, face recognition, and more. OpenCV is designed to be highly efficient and usable in real-time applications. It is widely used in academic research, commercial applications, and industries like robotics, surveillance, and autonomous vehicles.

Rathod, et al. [10] developed a real-time emotion detection device using OpenCV and bio-signals, that can detect human emotion based on the way the face looks and it showed very high accuracy.

Authors of [11] found that an OpenCV-based implementation outperformed a Cognitive Services API in emotion recognition after testing both for thousands of times using two Android applications that were developed just to test both.

Murali et al [12] presents a Sign Language Recognition System using Convolutional Neural Network (CNN) and computer vision techniques, specifically utilizing OpenCV for image processing tasks. OpenCV is employed for key steps such as image resizing, applying masks, dilation, and erosion operations.

Authors of [13] outlines a methodology for sign language recognition using Convolutional Neural Networks (CNN). The initial step involves data collection using a webcam to capture hand gestures. The collected images undergo preprocessing, Later, morphological operations, such as dilation and erosion using elliptical kernels, are applied to enhance image quality. OpenCV was integral to the image preprocessing stage, performing tasks such as resizing images to a uniform size and applying morphological operations to enhance image quality. These preprocessing steps were vital for ensuring that the CNN could accurately learn and classify the gestures.

Mishra et al, [14] presents a comprehensive methodology for developing a motion detection system using OpenCV. The system aims to enhance security by detecting and notifying users of any detected motion. The implementation demonstrates high accuracy and reliability in various environments, making it a viable solution for enhancing security through real-time motion detection.

Hasan and Sallow [15] emphasize the critical role of OpenCV in developing face detection and recognition systems. OpenCV provides essential tools for image processing, such as Haar Cascade classifiers for object detection, Local Binary Patterns (LBP) for texture analysis, and advanced algorithms like Eigenfaces and Fisherfaces for face recognition. The library's capabilities in image transformation, filtering, and real-time video analysis facilitate efficient implementation of complex computer vision tasks, making it a pivotal component in the system described.

Giri [16] used a combination of Convolutional Neural Network (CNN) and OpenCV for live emotion detection. Where OpenCV plays a pivotal role in this study by providing the necessary tools for image processing and facial feature extraction. The library is used to preprocess images, convert them to grayscale, and detect facial landmarks, which are crucial for emotion recognition.

MediaPipe is a cross-platform framework developed by Google for building multimodal (e.g., video, audio, and sensor) applied machine learning pipelines. It provides a set of reusable and customizable components for building efficient and scalable pipelines for processing and analyzing multimodal data. MediaPipe is designed to enable the rapid prototyping and deployment of machine learning models in various applications, including computer vision, augmented reality, and gesture recognition.

Lugaresi et al, [17] demonstrates the efficacy of MediaPipe through various use cases, including real-time object detection and face landmark detection with segmentation. The system achieved efficient processing with minimal latency, ensuring smooth operation and accurate synchronization of annotations with video frames. For instance, the object detection pipeline achieved real-time performance by combining detection and tracking tasks, running them in parallel to optimize resource usage. MediaPipe's flexibility and efficiency in managing computational loads across different platforms were highlighted, showcasing its capability to handle complex perception tasks in real-world applications with high accuracy and reliability.

Kumar [18] discusses the implementation of a hand gesture recognition system using MediaPipe, a framework developed by Google for building multimodal machine learning pipelines. MediaPipe plays a crucial role in this system by providing the tools for real-time hand tracking and gesture recognition. The framework captures hand gestures via a webcam, processes the images to identify hand landmarks, and uses these landmarks to classify gestures. MediaPipe's robust and efficient processing capabilities, including background subtraction and noise removal, are essential for accurately detecting hand positions and movements, thereby enabling the system to recognize various hand gestures.

Sánchez-Brizuela et al [19] presents an innovative algorithm for real-time hand segmentation, utilizing MediaPipe's highly optimized hand landmark detection capabilities. MediaPipe provides precise hand landmarks, which are then processed using morphological and logical operators to generate dynamic skin color models for accurate hand segmentation. The algorithm's simplicity and efficiency are bolstered by MediaPipe's robust landmark detection, allowing the segmentation process to be performed on non-accelerated hardware without sacrificing speed or accuracy.

Applications of hand gesture recognition

Hand gestures can be categorized into static and dynamic types. Static gestures involve a stable hand shape, while dynamic gestures include a series of hand movements, such as waving and hand shaking. The primary distinction between posture and gesture lies in their focus: posture emphasizes the shape of the hand, whereas gesture emphasizes hand movement. Research on hand gestures generally follows two main approaches: the wearable glove-based sensor approach and the camera vision-based sensor approach [20]. Wearable gloves employ various sensors to capture hand motion and position by detecting the coordinates of the palm and fingers [21]. Different sensors used in this technique include angular displacement sensors [22], curvature sensors [23], flex sensors [24], accelerometer sensors [25], and optical fiber transducers [26], each utilizing distinct physical principles based on their type.

Range of studies have explored the use of hand gesture recognition for various applications. Triyono [27] developed a sign language translator that will detect the hand gesture and converted it to words using OpenCV to collect the data from the webcam, achieving a 95% success rate in recognizing hand gestures.

Authors of [28] created an AI virtual painter using MediaPipe and OpenCV, allowing users to write in the air and see their words displayed on a screen by using the MediaPipe hand detection model to detect the hand gesture and then it will give a command based the gesture. Zhu et al. [29] focused on recognizing English letter gestures that will, achieving an accuracy of over 81%. Lastly, the study in [30] applied gesture recognition to virtual scene interaction, using MediaPipe and OpenCV to identify hand key points and improve accuracy and stability in virtual environments. These studies collectively demonstrate the potential of hand gesture recognition for a range of practical applications.

All these projects show that using both MediaPipe and OpenCv together will give high accuracy and great results, the project uses MediaPipe module to detect the hand gesture, OpenCv allows the user to use the webcam and then it will collect the data from the webcam, each hand gesture will result in a command such as moving the slides or drawing or using the pointer and the zoom.

III. Methodology

The methodology used in this project consists of several steps that are performed in certain order to achieve accurate hand gesture recognition. These steps are as follows:

The system starts by image processing: where raw video frames are captured from a high-resolution Webcam using the OpenCV library. From these frames hand detection is performed. It involves identifying the hands within the captured frames using the MediaPipe. Once the hands are detected, the next step is to extract landmarks from the hand regions. Landmark extraction is determining specific locations of the hand, such as fingertips, knuckles, and the center of the hand. These landmarks provide essential information for the next step, which is Gesture Recognition. This step involves understanding how hand is positioned and how it moves based on the landmarks on the hand and what gesture it represents. The final step is user interaction. Once a gesture is recognized, actions are triggered based on the recognized gesture. These steps are illustrated as a flowchart in Figure 1:

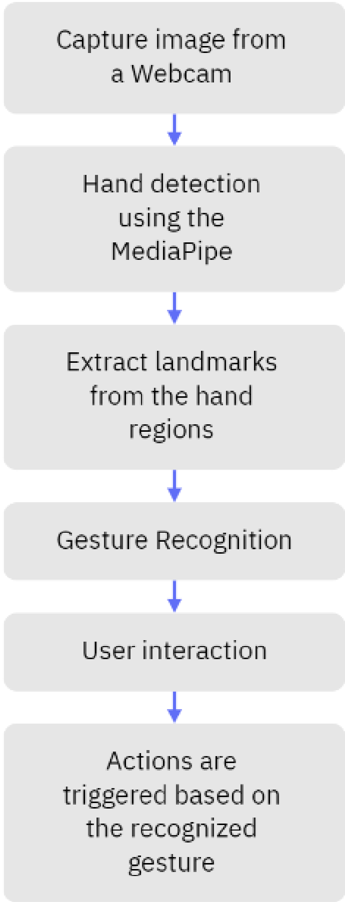

Figure 1: flowchart of the proposed system

A. Hand Tracking Module:

Several libraries are involved in this phase to recognize and track hand movement. these libraries are described as follows:

- CvZone is a computer vision library that offers pre-built modules such as Hand-Detection module. It is accessible for developers to use without wasting time on building the detection module.
- The Hand-Detection module within the CvZone library is used for hand tracking in this project. This module uses computer vision for detecting and tracking hands within a video stream. It locates hands in each frame of the video stream. It uses features such as landmarks and hand orientation to accurately identify hand positions.

The proposed system performs several functions to convert hand movement into action where two of them could be considered as setup and others are controlling functions

B. Setup functions:

The setup functions are performed as a preparatory for the hand gesture to function properly. These functions are threshold and hand classifiers, and explained as follows:

a. Threshold

At the beginning, the system functions based on hand movements, it needs to make sure that the hand is above a threshold. This threshold represents the horizon line that activates the system. This is set to avoid the reaction of the system for any unintended hand movement of the presenter. Also, the presenter must be aware that his hand must be raised to a certain level to activate the gesture. The virtual threshold is illustrated in Fig. 2 as a gray horizontal line

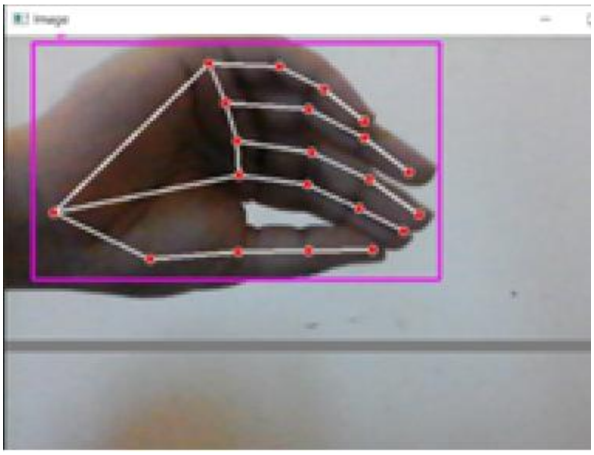

Figure 2: The hand must be above Threshold (The gray horizontal line)

b. *Hand Classifier*

The system classifies the detected hands as left or right based on their labels, and stores them in left_hand and right_hand variables, so the user can use the right hand for certain gestures and the left for the others. For this project only left hand will be used to control the slides.

C. *Presentation functions*

The control functions of the proposed system are used to control the power point in terms of zooming, move one slide to the left, one slide to the right, show a pointer on the screen, and write on the screen or on the slide itself. The explanation is as follows:

a. *Zooming*

By calculating the distance between the index finger and thumb landmarks with the hand positions as in Figure 3 it measures the distance between them. Enlarging the distance causing slide to zoom in and reducing the distance zooms out of the slide.

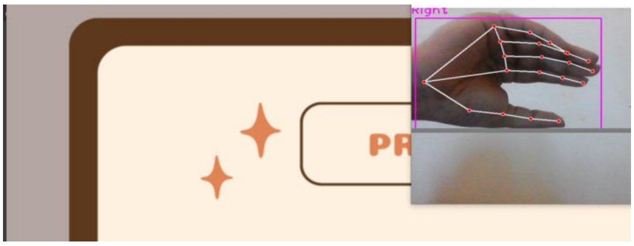

Figure 3: Screenshot of Zoom Function

b. *Moving slide left*

When detecting the left hand with only the thump is up the slides will move to the left one slide at a time as shown in Figure 4

Figure 4: Moving slide left

c. *Moving slide right*

When detecting the left hand with only the pinky is up the slides will move to the right one slide at a time as shown in Figure 5.

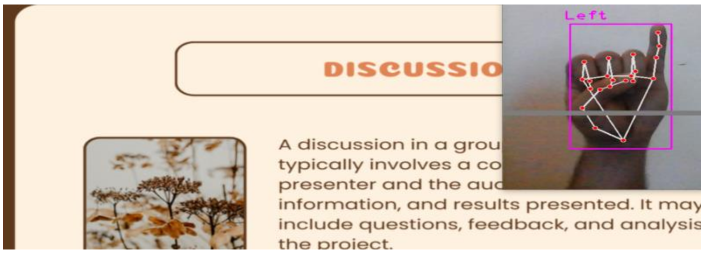

Figure 5: Moving slide right

d. Pointer

When detecting the left hand with the index and middle fingers are up as shown in Figure 6 it will show a pointer on the screen as red dot that will move around based on the hand movement.

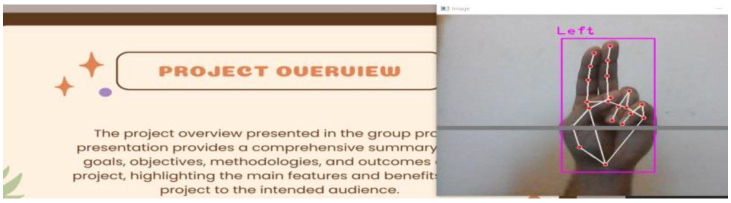

Figure 6: functioning as a pointer (shown as dot in this figure)

e. Writing

When detecting the left hand and with the index is up it will start writing on the screen according to the hand movement as illustrated in Figure 7.

Figure 7: Hand Gesture for writing

As illustrated by the above figures, the results of using MediaPipe and OpenCv were outstanding, all the functions worked at a high level of accuracy.

IV. Conclusion

This project demonstrates the effective application of computer vision and hand gesture recognition in creating an intuitive and interactive presentation control system. By utilizing OpenCV and MediaPipe, the system can accurately detect and interpret hand gestures, enabling users to navigate slides, annotate content, control a pointer, and apply zoom functions without the need for traditional input devices. The developed system addresses specific challenges in the educational field by providing a cost-effective alternative to smart boards, enhancing teacher-student interaction, and making presentations more engaging. The project shows promising results in terms of accuracy and functionality, indicating that the combination of MediaPipe and OpenCV is a robust solution for hand gesture recognition. Future work can focus on further improving the system's accuracy, expanding its range of gestures, and integrating additional features to enhance user experience. Overall, the project contributes to the advancement of interactive technologies and provides a valuable tool for educators and presenters.

References

[1] Desai, S.; Desai, A. Human Computer Interaction through hand gestures for home automation using Microsoft Kinect. In Proceedings of the International Conference on Communication and Networks, Xi'an, China, 10–12 October 2017; pp. 19–29.

[2] Rajesh, R.J.; Nagarjunan, D.; Arunachalam, R.M.; Aarthi, R. Distance Transform Based Hand Gestures Recognition for PowerPoint Presentation Navigation. Adv. Comput. 2012, 3, 41.

[3] Kaur, H.; Rani, J. A review: Study of various techniques of Hand gesture recognition. In Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 4–6 July 2016; pp. 1–5.

[4] Murthy, G.R.S.; Jadon, R.S. A review of vision based hand gestures recognition. Int. J. Inf. Technol. Knowl. Manag. 2009, 2, 405–410.

[5] Khan, R.Z.; Ibraheem, N.A. Hand gesture recognition: A literature review. Int. J. Artif. Intell. Appl. 2012, 3, 161. [CrossRef]

[6] Suriya, R.; Vijayachamundeeswari, V. A survey on hand gesture recognition for simple mouse control. In Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, India, 27–28 February 2014; pp. 1–5.

[7] Garg, P.; Aggarwal, N.; Sofat, S. Vision based hand gesture recognition. World Acad. Sci. Eng. Technol. 2009, 49, 972–977.

[8] Dipietro, L.; Sabatini, A.M.; Member, S.; Dario, P. A survey of glove-based systems and their applications. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 2008, 38, 461–482. [CrossRef]

[9] LaViola, J. A survey of hand posture and gesture recognition techniques and technology. Brown Univ. Provid. RI 1999, 29. Technical Report no. CS-99-11.

[10] Rathod, P., et al. (2016). Bio-signal based emotion detection device. 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 105-108.

[11] Prieto, L. A. B., & Komínkova-Oplatková, Z. (2017). A performance comparison of two emotion-recognition implementations using OpenCV and Cognitive Services API.

[12] Murali, R. S. L., Ramayya, L. D., & Santosh, V. A. (2022). Sign language recognition system using convolutional neural network and computer vision. International Journal of Engineering Innovations in Advanced Technology Volume-4 Issue-4, December 2022.

[13] J. Sigut, M. Castro, R. Arnay and M. Sigut, "OpenCV Basics: A Mobile Application to Support the Teaching of Computer Vision Concepts," in IEEE Transactions on Education, vol. 63, no. 4, pp. 328-335, Nov. 2020, doi: 10.1109/TE.2020.
2993013.)

[14] S. Mishra, V. Verma, N. Akhtar, S. Chaturvedi, and Y. Perwej, " An Intelligent Motion Detection Using OpenCV, International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 9, Issue 2, pp.51-63, March-April-2022. Available at doi : https://doi.org/10.32628/ IJSRSET22925

[15] Hasan, R. T., & Sallow, A. B. (2021). Face detection and recognition using opencv. Journal of Soft Computing and Data Mining, 2(2), 86-97.

[16] Giri, S., et al. (2022). Emotion Detection with Facial Feature Recognition Using CNN & OpenCV. In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) pp. 230-232.

[17] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019, June). Mediapipe: A framework for perceiving and processing reality. In Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR) (Vol. 2019).

[18] Kumar, A., Venkatesh, C., DA, L. V., & Harish, S. (2022). Computer vision based Hand gesture recognition system. NeuroQuantology, 20(7), 2859.

[19] Sánchez-Brizuela, G., Cisnal, A., de la Fuente-López, E., Fraile, J. C., & Pérez-Turiel, J. (2023). Lightweight real-time hand segmentation leveraging MediaPipe landmark detection. Virtual Reality, 27(4), 3125-3132.

[20] Oudah, M., Al-Naji, A., & Chahl, J. (2020). Hand gesture recognition based on computer vision: a review of techniques. journal of Imaging, 6(8), 73.

[21] Ahuja, M.K.; Singh, A. Static vision based Hand Gesture recognition using principal component analysis. In Proceedings of the 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE), Amritsar, India, 1–2 October 2015; pp. 402–406.

[22] Jesperson, E.; Neuman, M.R. A thin film strain gauge angular displacement sensor for measuring finger joint angles. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, New Orleans, LA, USA, 4–7 November 1988; pp. 807–vol.

[23] Kramer, R.K.; Majidi, C.; Sahai, R.; Wood, R.J. Soft curvature sensors for joint angle proprioception. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 1919–1926.

[24] Shrote, S.B.; Deshpande, M.; Deshmukh, P.; Mathapati, S. Assistive Translator for Deaf & Dumb People. Int. J. Electron. Commun. Comput. Eng. 2014, 5, 86–89.

[25] Gupta, H.P.; Chudgar, H.S.; Mukherjee, S.; Dutta, T.; Sharma, K. A continuous hand gestures recognition technique for human-machine interaction using accelerometer and gyroscope sensors. IEEE Sens. J. 2016, 16, 6425–6432.

[26] Fujiwara, E.; dos Santos, M.F.M.; Suzuki, C.K. Flexible optical fiber bending transducer for application in glove-based sensors. IEEE Sens. J. 2014, 14, 3631–3636.

[27] Triyono, Liliek & Pratisto, Eko & Bawono, S & Aji, Fendi & Yudhanto, Yudho & Raharjo, Bangun. (2018). Sign Language Translator Application Using OpenCV. IOP Conference Series: Materials Science and Engineering. 333. 012109. 10.1088/1757-899X/333/1/012109.

[28] Nandagopal, S., Soundarya, R., Vaishnavi, S., Vanisri, S., & Hiranya, S. (2022). AI Virtual Painter using OpenCV and Mediapipe. International Journal of Engineering Research in Computer Science and Engineering.

[29] Zhu, H., Deng, C., & Zhu, Y. (2022, December). MediaPipe based gesture recognition system for English letters. In Proceedings of the 2022 11th International Conference on Networks, Communication and Computing (pp. 24-30).

[30] Zhou, Z., & Lv, Y. (2023). The application of gesture recognition based on MediaPipe in virtual scene. Conference on Machine Learning and Computer Application.