Video controlling using hand gestures using automation techniques in ML pipelines: A case study

Seema Sharma [1,2]

[1] Research Scholar, Department Of Information Tecnology, Manipal University Jaipur, Jaipur, India
[2] Assistant Professor, Department of CSE, JECRC University, Jaipur, India

Abstract: In this modern age, the advancement in ubiquitous computing has made the use of a natural user interface very much required. The presence of computers and making use of the facilities of human-computer interaction in our societies will bring and mark a positive impact on our societies. Either it was the day when the technologies had not so advanced or today's when the technologies have advanced so much that we spent most of our time communicating, playing, doing our jobs with the machines and many more, even then human beings had used and are still using a broad range of gestures to communicate or interact with each other. The human gesture is a mode of non-verbal interaction medium and can provide the most intuitive, original, and natural way to interact with computers. The objective of this paper is to make the interaction between humans and computers as natural as the interaction between humans. This paper is to recognize the static hand gesture images (i.e. frames) based on shapes and orientations of the hand which are extracted from an input video stream recorded in stable lighting and simple background conditions. We can use this vision based recognized gestures to control multimedia applications (like Windows Media Player, Windows Picture Manager, VLC Player, etc.) running on a computer using different gestural commands.

Keywords: Automation techniques, machine learning, Video controlling, pipelines

1. Introduction:

In today's world, everyone is surrounded by machines and gadgets at all times. The use of mobile phones, tablets, computers, and smartwatches has increased manifold due to the creation of new technology and Internet use. The system that contains a combination of hardware and software for communication between the user and the computer or machine, makes up the human-computer interface (HCI). Normally, switches touch screens, and control elements are used. An easier way of communication is through the touchless user interface, which could be through voice or gesture input. Since voice input has already been applied extensively in every field; this project focuses on gesture input. Instead of using a mouse or keyboard, users can show hand gestures to perform operations or control certain functions of the computer. This could include playing a video, scrolling through a document, or opening an application [1,2,]. This project aims to focus on the field of gesture recognition. Through this application, users will be able to control multimedia on the system using hand gestures. The input is taken through the webcam. The programming language used is Python. A common scenario where this application will be able to help users is in the kitchen, where users may not be able to touch device screens or press buttons with their hands while cooking. Another scenario is devices used at the beach or poolside, where users will have their hands too wet or sandy to type. Some of the other common applications are in automated homes, driving safety, television control, and much more. With upcoming IoT (Internet of Things) devices, it is possible to switch off lights, control surround sound systems, access devices, and change room temperature through the connected devices. Many of these devices use gesture recognition and voice input[2,3].

Gesture recognition is the process by which systems can see, recognize and respond to gestures shown by the user. This perceptual user interface provides a way for users to communicate with the system without the need for clicking and typing. Gestures can come from any bodily motion or state. Usually, they come from the hand or face. Currently, there is a lot of focus on understanding emotions from facial expressions and gesture recognition through hand movements. To interpret sign language, many approaches have been made by using computer vision techniques[3,4]]. This will help to build a better bridge between humans and machines, compared to GUIs (graphical user interfaces) or text user interfaces, which limit the input to the mouse and keyboard. Using this concept of gesture recognition, it is possible to make conventional input to systems redundant and bring newer technologies into the light to increase user productivity as well[1,10,11].

The problem with traditional systems can arise in many ways. Firstly, for people with disabilities, it is not possible to click buttons or type for every operation. Secondly, if some part of your computer is not working properly, it is not possible to rely on the traditional use of a mouse and keyboard for all inputs. Thirdly, even if people are able and have usable systems, nowadays, every part of life is being automated and simplified. Therefore, people prefer it if there is a way to sit back and relax without having to control the laptop through the usual methods every time. Whether it is while watching a movie, listening to songs, or reading a PDF, users prefer to touch less engagement with the screen and control through intuitive, new, and creative ways.[12] Many such techniques require complex algorithms and a lot of packages. But in this project, an attempt to create a simple, yet powerful tool is made. Traditionally, computer vision is used to handle gesture recognition. However, there are easier ways to do the same operations. Since Python is taking over the programming world from all aspects, it is the best way to start. Using the automation techniques available in Python, it is possible to perform system operations. The objective and importance of hand gestures employing automation techniques are discussed in Section 2, related work is discussed in Section 3, and an innovative method is discussed in Section 4. Outcomes and performance metrics are shown in Section 6. And Section 7 will discuss the conclusion.

2. Motivation:

The companies are certainly at the forefront of the touchless gesture recognition technology field at the moment, but other companies also contain some sort of value in this field. These are the Swedish eye-tracking device (Tobii Rex), technology for message writing in the air (Air writing), and motion sensor devices (Leap Motion). The companies and products released in this field are currently very expensive and cannot be used by the general public, as their access is usually limited to experimental purposes. Due to huge initial and maintenance costs as well as complex hardware and software devices, these technologies have not reached the overall audience of computer and other electronic device users. Several existing projects cater to the needs of multimedia control using hand gesture recognition. However, it has been observed that none of these are renowned.

Overall, everyone who uses computers daily, even for work, can benefit from this project. It can be applied even in business scenarios where presentations are given daily and it is not always possible to use the keyboard or mouse to change slides. This application will provide an easier, faster, and more natural way of multimedia control. The programming is done using Python, with preinstalled webcam (basic). The Media Pipe library is a framework for building multimodal(e.g video, audio, or any time series data), cross-platform (i. android, IOS, web, edge devices) applied ML pipelines. Media pipe also facilitates the deployment of machine learning technology into demos and an application on a wide variety of different hardware platforms is used. The PySeriallibrary is also used to make sure the input data from the sensors come into the system bit by bit. The gesture recognition is done by the sensors. The hand is detected and distance is calculated, before determining the gesture and sending data to the Python file, which performs appropriate action by identifying which application is currently in focus.

The motivation for the studies is that there are a lot of critical situations in the day-to-day lives of

disabled people. We have come up with a small part of such life to make it easier using computer vision technology. In this paper, we have discussed a low system that uses hand gesture recognition technology to control the VLC media player. Among many computer visions based interactive systems, designing hand gestures and facial expression-based HCI system retains to be a highly challenging task. Our main purpose is to find a no tangible way to interact with the computer.

This finding about important contributions can be brief as follow:

Efficiency: The proposed system will be able to successfully control the multimedia applications on the computer by recognizing hand gestures and acting appropriately. It provides an efficient touchless interface for the users.

Improved accuracy: By focusing on only hand gestures, it will provide a high accuracy rate in recognizing the gestures, in varying scenarios like light and color.

Usability: The system will be easy to use even for someone new to the technology of gesture recognition systems. Itwillbeeasytolearnandadaptto

Extensibility: The system can be added on to by anyone with minimal programming knowledge to extend its services across all system applications and many types of devices. It contains Python modules to provide automation which can work on many operating systems.

Simple algorithms: The algorithms used do not massively increase the computational burden on the system. They are modest and easily extensible.

This paper also discusses the existing technology with their limitation .this paper also depicts that the proposed fame work is more accurate than the other.

3. Related work:

There are many companies all around the world producing gesture recognition technology with their systems. Some of these are described in detail below. Intel Corporation's touchless multifactor authentication (MFA) helps improve efficiency and convenience in healthcare organizations while avoiding security risks. This solution uses a combination of device recognition and facial recognition techniques. One of Microsoft Corporation's projects aims for camera-based gesture recognition to view and control surgeries without contact. It also analyses the challenges in deploying such systems. The Elliptic Labs software suite provides gesture recognition functions through the existing microphone and earphones. This technology can be used to recognize your hand gestures and move the objects on the screen[13,14].

In 2017, five engineering students from a college in Mangaluru created a hand gesture recognition module that used a computer webcam. They used the methods of convexity, canter of gravity, and identification to determine the various gestures. After recognition, the cursor of the system is moved and performs operations on the system, according to the gesture. The drawback of this system is that the hand is detected using the background subtraction method, in which the background is kept as a white screen. Other objects in the camera view or varying hand colors result in failed results[5].

In 2015, four students of a Maharashtrian engineering college created a hand gesture automation software for controlling files like PDFs, videos, and audio. It was done using HSV scaling, three holdings, filtering and then calculating the center of gravity. It identified the fingertips but could not provide a good accuracy rate. There was another notable system created. In 2015, used super-pixels, for hand gesture recognition, which used a special Kinect depth camera. The cost of this camera was high, hence this project, though effective, was not feasible.[6]

In 2014, five Chinese engineering students created a hand gesture recognition system that identified the fingers of the hand. Through background subtraction, binary transformation, segmentation, and finger recognition, the hand gesture was identified. This system, however, only accurately recognized static gestures. If other objects match the color of the hand or are moving, the performance was accurate. [6]

In 2012, four men created a system that controlled PowerPoint presentations using hand gestures. Using segmentation algorithms, they created software that could browse through the slides of a PPT. This system refused to provide

results when the hand gesture was unclear or fingers were not distinctly stretched. [7]

In 2010, two students of IIT created an application for the control of VLC Media Player using hand gestures. The gesture recognition was done using the K-nearest neighborhood algorithm, Lucas Kanade Pyramidical Optical Flow algorithm, and integrated XML files for feature storage. The methods and algorithms used in this project were difficult to understand and implement. Due to this, the runtime and recognition accuracy of the resultant application was relatively low[8].

Some of the disadvantages of existing systems are as follows:-

- Expensive: The present systems are all part of big corporations that release the hardware and software as a huge package which is very expensive for the general crowd to buy and maintain.
- Difficult algorithms: The methods and functions used to implement hand gesture recognition systems are difficult and long. The software involves the integration of many languages and platforms to present the finished application.
- Limited applications: The systems created so far have been limited to a few gestures and one or two applications in control, hence reducing the possibilities for users.
- Low accuracy: With infinite variations possible in terms of hand color, size, shape, and movement, the training data for gesture recognition software should be large to see accurate results. Most existing systems do not have good accuracy in terms of hand gestures.

4. Proposed system:

The project that has been developed with the title "Multimedia Control Using Gesture Recognition" aims to provide a solution to the drawbacks of the existing systems in this field, mentioned above. This project will be able to provide a touchless user interface for computer systems. The hardware devices used are also cheap and can be implemented in future system designs as part of the computer itself. The proposed system can be used to control and do actions on applications like VLC media player, and Windows media player It can provide functions that will enable the user to play, pause, toggle mute, forward, rewind, increase volume, and decrease the volume in VLC media player and similar actions in Windows media player for songs. Through this system, users will be able to interact with the screen without having to type. It can be used by people who are disabled. It can also be used by older generations who do not know how to operate computers. It can also be used by anybody who is unable to type or has a computer with a malfunctioning keyboard. It can be extended and maintained with ease by anybody with minimal programming knowledge. The programming is done using Python, with preinstalled webcam (basic). The Media Pipe library is a framework for building multimodal(e.g video, audio, or any time series data), cross-platform (i. android, IOS, web, edge devices) applied ML pipelines. Media pipe also facilitates the deployment of machine learning technology into demos and applications on a wide variety of different hardware platforms is used. The PySerial library is also used to make sure the input data from the sensors come into the system bit by bit. The gesture recognition is done by the sensors. The hand is detected and distance is calculated, before determining the gesture and sending data to the Python file, which performs appropriate action by identifying which application is currently in focus.

Hardware Requirements

Processor: Any Processor above 1GHz
RAM: minimum 1 GB
Hard Disk: minimum 5 GB
Camera: Any webcam
Output device: Monitor

Software Requirements

Operating system: Windows 10
Editor: Python IDLE
Software: Python 3.6+
Command line: Command prompt

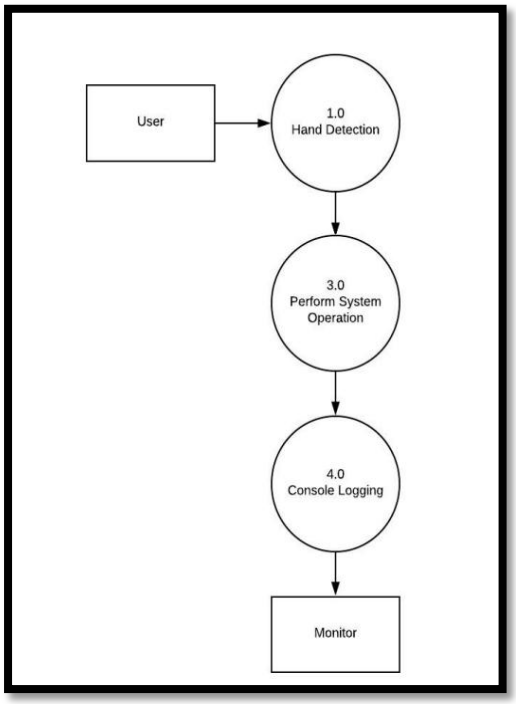

Fig 1 DFD

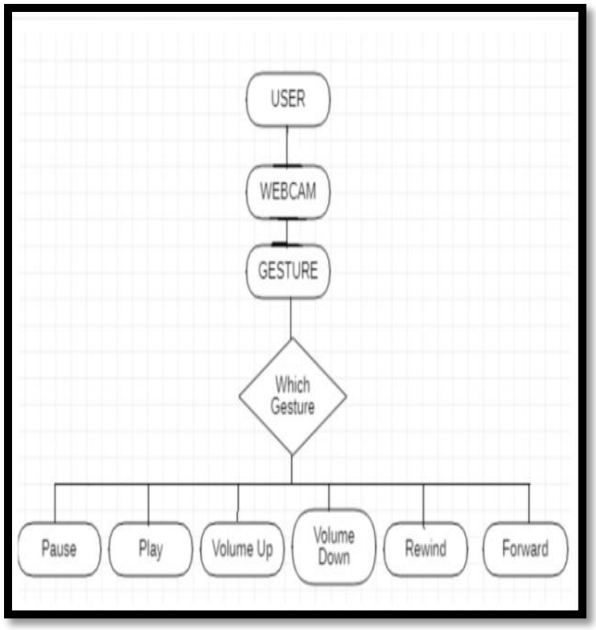

Fig 2 control flow

5. Experimental Result and Testing:-

To evaluate the functionality of any projected release, testing is done on various aspects. The main purpose of testing is to find out whether the application meets the requirements of the user and to identify if there are any defects to be rectified. The various features of the system are evaluated against the ideal or requested specifications to determine quality. Various types of testing are done at various levels of the software development process.

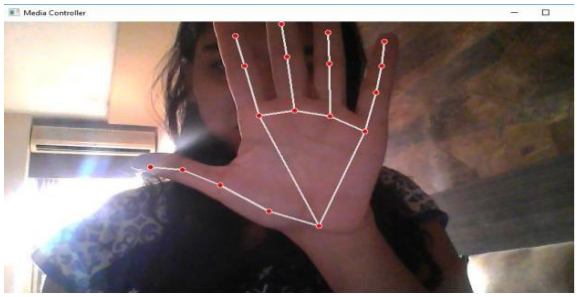

Fig 3 :- Hand Detection

Table 1:- Test Case for Initial Recognition

Serial Number of Test Case	TC 01
Module Under Test	Python code
Description	When the program is run, the first output should be the recognition of the Command Prompt application in focus.
Output	Console logging in Command Prompt with the application name.
Remarks	Test Successful.

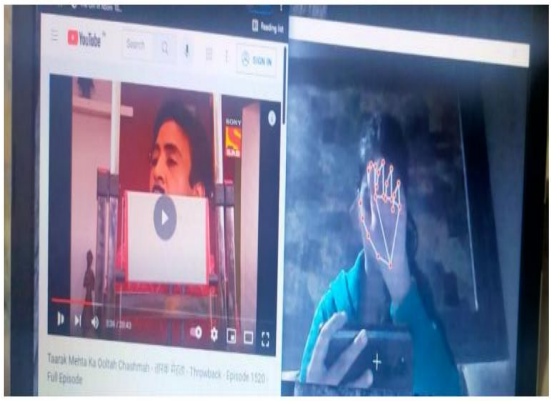

Fig4: - Play video

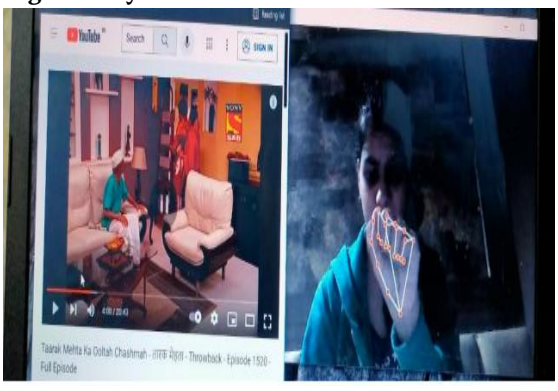

Fig 5:- pause the video

Table 2:- Test Case for VLC play/pause

Serial Number of Test Case	TC 02
Module Under Test	VLC Media Player Play/Pause Action
Description	When the user puts both hands near the sensors, and the VLC media player is active, the video streaming in VLC should toggle between play and pause.
Input	Both hands near sensors (within 5 cm) with VLC media player in focus
Output	If both hands are detected near sensors, the video should play if paused, and vice versa.
Remarks	Test Successful.

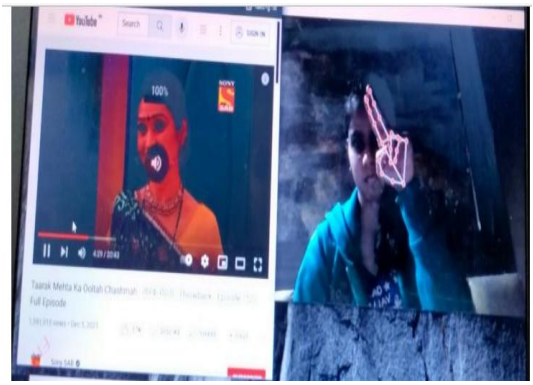

Fig6:- volume increase

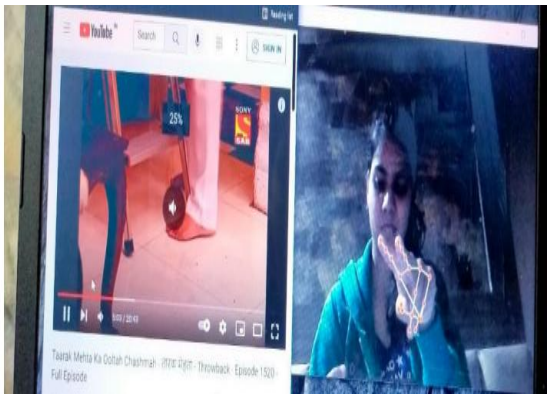

Fig7:-: - volume decrease

Table 3 Test Case for Volume Increase

Serial Number of Test Case	TC 03
Module Under Test	VLC Media Player Volume Increase Action
Description	When the user moves the left hand forward towards the left sensor, with the VLC media player in focus, the volume should increase.
Input	Left-hand stays around15to 25cminfront of the left sensor, then move slowly towards the sensor, with the VLC media player in focus
Output	Volume in VLC media players should increase.
Remarks	Test Successful.

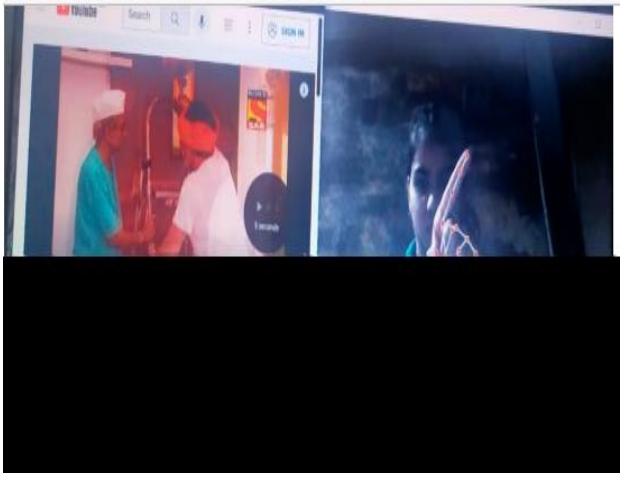

Fig8:- Move forward

Fig 9: Rewind

Table 4:- Test Case for Forward

Serial Number of Test Case	TC 03
Module Under Test	VLC Media Player Volume Increase Action
Description	When the user places their right hand near the right sensor, and Windows Media Player is active, the music player should skip to the next track.
Input	Right hand near the right sensor (within 5cm) and Windows Media Player in focus.
Output	The player skips the remaining length of the current song and moves to the next song in the list.
Remarks	Test Successful.

6. Conclusion and Further Enhancements

With the advancements in technology to provide novel, convenient and fast methods of human-computer interaction, gesture recognition has received wide appreciation. The various existing systems have good working features but have not been well received by customers. The main problem lies in the fact that these systems have low accuracy rates and complex algorithms. The proposed system will aim to combat these issues and stand out in the crowd of gesture recognition systems. It will provide a touchless user interface for controlling multimedia files and applications such as video players, music players, and documents. It will act as a helping aid for the manipulation of systems for people who have disabilities, who cannot access their input devices, or anyone who prefers this more natural method of communication compared to their methods.

In the future work Hardware that is integrated with the computer and More gesture inputs(not confined to hand gestures)will take. we can facilities for more operations to be performed

Reference:

1. Haria, A., Subramanian, A., Asokkumar, N., Poddar, S. and Nayak, J.S., 2017. Hand gesture recognition for human-computer interaction. *Procedia computer science*, *115*, pp.367-374.
2. Sonkusare, J.S., Chopade, N.B., Sor, R. and Tade, S.L., 2015, February. A review on hand gesture recognition system. In *2015 International Conference on Computing Communication Control and Automation* (pp. 790-794). IEEE.
3. Krishna Reddy, V.V., Bhuvana, K.N.V.S., UmaHarikka, K., Sai Teja, D. and Suguna Kumari, J., 2022. Hand Gesture Recognition Using Convolutional Neural Networks and Computer Vision. In *Cognitive Informatics and Soft Computing* (pp. 583-593). Springer, Singapore.
4. Zhang, Y., Wang, J., Ye, L., Xue, X. and Zhang, Q., 2017. A virtual music control system based on dynamic hand gesture recognition. In *Transactions on Edutainment XIII* (pp. 74-85). Springer, Berlin, Heidelberg.
5. Hsieh, C.C., Liou, D.H. and Lee, D., 2010, July. A real-time hand gesture recognition system using motion history images. In *2010 2nd international conference on signal processing systems* (Vol. 2, pp. V2-394). IEEE.
6. Islam, S., Matin, A. and Kibria, H.B., 2021, December. Hand Gesture Recognition Based Human Computer Interaction to Control Multiple Applications. In *International Conference on Intelligent Computing & Optimization* (pp. 397-406). Springer, Cham.
7. Ansar, H., Jalal, A., Gochoo, M. and Kim, K., 2021. Hand gesture recognition based on auto-landmark localization and reweighted genetic algorithm for healthcare muscle activities. *Sustainability*, *13*(5), p.2961.
8. Rautaray, S.S. and Agrawal, A., 2010. A vision-based hand gesture interface for controlling VLC media player. *International Journal of Computer Applications*, *10*(7), pp.11-16.

9. Jayasekera, M.K.C., 2021. *Real-time New Zealand sign language translator using convolution neural network* (Doctoral dissertation, The University of Waikato).
10. Zhigang, F. Computer gesture input and its application in human-computer interaction. Mini Micro Syst. 1999, 6, 418–421.
11. Tang, M. Recognizing Hand Gestures with Microsoft's Kinect; Department of Electrical Engineering of Stanford University: Palo Alto, CA, USA, 2011.
12. . Bamwenda, J.; Özerdem, M.S. Recognition of Static Hand Gesture with Using ANN and SVM. Dicle Univ. J. Eng. 2019, 10, 561–568.
13. Tekin, B.; Bogo, F.; Pollefeys, M. H+ O: Unified egocentric recognition of 3D hand-object poses and interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4511–4520.
14. Wan, C.; Probst, T.; Van Gool, L.; Yao, A. Self-supervised 3d hand pose estimation through training by fitting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10853–10862