

AI-Based Hand Gesture Controlled Media Player

Ritik Sharma¹, Yash Tripathi², Rinku Agrawal³, Yash Pratap Singh⁴, Shagun Sharma⁵

^{1,2,3,4,5}Department of Computer Science and Engineering, VCTM Aligarh

Abstract— Artificial Intelligence and Computer Vision are really changing the way we interact with computers. We usually control media players with things like keyboards, mice, remote controls or touch screens. Sometimes these ways of controlling things are not very convenient especially when we need to control things without touching anything.

In the few years people have been doing a lot of research on recognizing gestures. This is a part of how we interact with computers. Gesture-based systems let us talk to machines using movements of our hands instead of using physical devices. This project is about making a media player that we can control with hand gestures that are captured by a webcam. The system uses Computer Vision to find the parts of our hands and figure out where our fingers are, in real time.

The system we made uses Python, OpenCV and MediaPipe to track our hands and recognize gestures accurately. We assigned hand gestures to different things we can do with media players, like play pause make the volume louder or softer go to the next song and go to the previous song. Our system is designed to be easy to use, efficient and touchless so it is easy for people to control media players. Artificial Intelligence and Computer Vision are what make this possible.

Index Terms— Artificial Intelligence, Computer Vision, Hand Gesture Recognition, OpenCV, MediaPipe, Human-Computer Interaction, Media Control System

I. INTRODUCTION

Artificial Intelligence is a technological development in modern computing. It has changed how humans interact with machines by enabling systems to understand actions, speech and behavior [1]. Artificial Intelligence plays a role in Human-Computer Interaction, where machines respond to natural human movements and commands [14].

With the growth of multimedia applications media players are now a part of life. People use media players to watch videos listen to music learn online and for entertainment. Most media players are controlled using keyboards, mice or touch screens. However,

these methods are not always convenient especially when users need hands- interaction.

Gesture recognition technology offers a way to interact with computer systems [7][8]. In gesture-based systems the user makes hand movements in front of a camera. The system recognizes those movements as commands. This approach makes interaction more natural and efficient because humans can communicate with machines using physical gestures with their hands.

Recent advancements in computer vision and machine learning have made it possible to detect hand movements accurately in time [2][5]. Tools like OpenCV and MediaPipe help, with image processing and hand landmark detection [3][6].

These technologies enable developers to create systems that track finger positions and identify gestures with accuracy. The proposed research focuses on developing an AI-Based Hand Gesture Controlled Media Player. This media player allows users to control media functions without physical input devices. The system captures video from a webcam detects hand landmarks recognizes gestures and converts them into media control commands. Artificial Intelligence powers this media player to understand and respond to hand gestures.

II. RESEARCH OBJECTIVES

The main thing this research is trying to do is to make a system that can control media using gestures. This system will use Artificial Intelligence and Computer Vision to work [1][2]. The goal is to see if we can use hand gestures to control things like videos and music without using the controls.

To get this done we have a thing we want to accomplish. First, we need to look at the technology that already exists for recognizing gestures and figure out how it works. This means we have to understand how Computer Vision and Artificial Intelligence can

detect and understand what our hands are doing.

Another thing we want to do is make a system that can see what our hands are doing in time using just a regular webcam. This system should be able to tell where our fingers are and match them to things like play, pause and turning the volume up or down [6].

We also want to see how well our system works and if people can really use it. We will look at things like how accurate it's how fast it responds and if it is easy to use [5]. We also want to explore how this kind of system can be used in life like, in smart homes or to help people or in places where we interact with things. We will be looking at gesture-based media control systems. How they can be used in different places. Gesture-based media control systems are what we are trying to make. We want to see how they can be used.

III. RESEARCH GAP

We have made a lot of progress with Artificial Intelligence and gesture recognition. There are still some problems with the systems we use to control media. Most media players need us to use keyboards or remote controls or touchscreens [8]. This is not very flexible. It is not very convenient in some situations. When we use gestures to control things it does not always work well. Many systems have trouble understanding what we are doing with our hands [7]. They get confused if the light is not right or if the background is messy. They also have trouble if we are holding our hands in ways. Another big problem is that many gesture recognition systems are slow. They need a lot of power to work [5]. They need special equipment. This makes them hard for regular people to use. Current systems also only let us use a gesture. This makes them not very useful. A lot of the systems that researchers have made are also not very easy to use. So, we need to make a system that can understand gestures and control media in time. It should work with the hardware we already have. It should be easy to use. We need a system, like this because the ones we have now are not very good. Artificial Intelligence and gesture recognition can be used to make a system.

IV. LITERATURE REVIEW

Gesture recognition is something that a lot of people have been looking into in the field of computer vision and human-computer interaction [7][8]. People are

trying to figure out ways to get machines to understand the movements of hands and turn them into commands that make sense.

The old ways of doing gesture recognition used devices like data gloves and sensors [8]. These methods worked well but they were expensive and not very convenient to use every day. Now that computer vision has gotten better people like using vision-based gesture recognition systems because they do not need any hardware.

Some studies have shown that image processing is really important for detecting and analyzing hand gestures. OpenCV is a tool that people use a lot to capture and process video frames [3], which lets them analyze hand movements in time. Googles Media Pipe has made hand tracking systems a lot better [6][13] by giving us a model that has already been trained to find 21 landmarks on the hand.

People have also been using machine learning to make gesture recognition more accurate [5][15]. They have trained models using sets of data to teach them to recognize different hand gestures. Even with all this research gesture recognition systems still have some problems, like being sensitive to the environment having trouble, with gestures that are not clear and being complicated to compute. Gesture recognition systems are still trying to overcome these challenges and researchers are working hard to improve gesture recognition systems.

V. RESEARCH METHODOLOGY

The AI-Based Hand Gesture Controlled Media Player is made using a method that focuses on getting things done. This method is about designing, making and testing a system that can recognize gestures in real time using computer vision techniques.

The first thing to do is figure out what technologies are needed to recognize gestures. We use computer vision libraries like OpenCV and Media Pipe because they have tools for processing images and finding the landmarks on a hand [3][6]. We use Python to make the system because it is easy to use and has a lot of support for intelligence and image processing libraries.

Next the system is set up to capture video from a webcam. The video is broken down into frames that are looked at one by one to see if a hand is there. Media Pipe is used to find the fingers by looking at the

landmarks on the hand [6]. After the fingers are found the system looks at how they're arranged to figure out what gesture is being made. The system checks if the fingers are open or closed and compares them to patterns, we already know to figure out what the gesture means.

The system is tested in situations to see how well it works. We look at things like how bright the light's how far the hand is, from the camera and which way the hand is facing [5]. This way of doing things makes sure the system is simple works well and is practical while still being able to recognize gestures. The AI-Based Hand Gesture Controlled Media Player is tested to make sure it can recognize gestures in conditions. The AI-Based Hand Gesture Controlled Media Player is made to be easy to use.

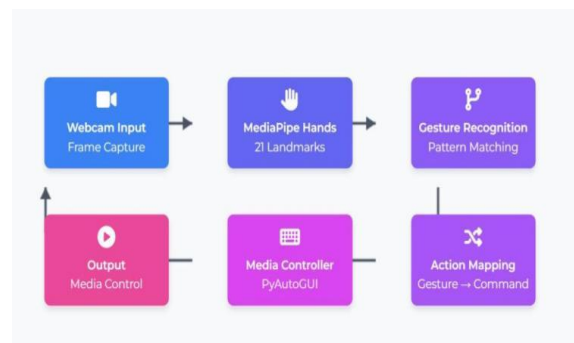
VI. SYSTEM ARCHITECTURE

The AI-Based Hand Gesture Controlled Media Player has a lot of parts that work together. These parts help detect gestures and control the media player.

The first part is the Input Module. It uses a webcam to capture video in time. The camera keeps sending video frames to the computer for it to look at.

The next part is the Frame Processing Module. This part uses OpenCV to change the video frames into a format that the computer can use to detect gestures [12].

Then there is the Hand Detection Module. This is where Media Pipe comes in. It finds the hand. Figures out where all the different parts of the hand are. It can find twenty-one points on each hand.



After that the system moves on to the Gesture Recognition Module. This part looks at the fingers to see which ones are open and which ones are closed. The AI-Based Hand Gesture Controlled Media Player

then sends the gesture it found to the Control Module. This part turns the gesture into a command. It uses PyAutoGUI to send keyboard commands to the media player [11].

Finally, the Output Module shows the gesture it found on the screen. It also does what the gesture is telling it to do with the media. The AI-Based Hand Gesture Controlled Media Player is really good, at controlling the media player with gestures.

VII. WORKING MECHANISM

The gesture-controlled media player works by looking at video in time and recognizing gestures. When you start the program, the camera turns on. The system starts taking pictures. It takes each picture. Uses OpenCV to get it ready to find hands [12]. The picture is changed into RGB format because the MediaPipe model is better at understanding RGB pictures.

The MediaPipe model looks at the picture. Finds a hand if there is one [6][13]. It figures out where all the joints are on the hand. The system checks if each finger is open or closed by looking at the points on the hand. It counts how many fingers are, up and different fingers mean things.



When the system sees a gesture, it matches it with something the media player can do. For example, if your hand is open that might mean play. If you make a fist that might mean pause. The media player gets the command. Pyautogui helps it work like you typed it on the keyboard [11]. There is a wait before the next thing happens so you do not get the same command twice. This keeps happening until you stop the program. The gesture-controlled media player uses the MediaPipe model and the gesture-controlled media player uses PyAutoGUI to control the media player. The gesture-controlled media player is always looking at the video to find gestures.

VIII. ALGORITHM

- 1) Start the program
- 2) Initialize webcam
- 3) Capture video frame
- 4) Convert frame to RGB [12]
- 5) Detect hand using MediaPipe [6]
- 6) Identify hand landmarks
- 7) Check finger positions
- 8) Recognize gesture pattern
- 9) Match gesture with command
- 10) Execute media control action
- 11) Display result on screen
- 12) Repeat until program stops

IX. APPLICATIONS

The Hand Gesture Controlled Media Player that uses Artificial Intelligence can be used in a lot of areas where we do not need to touch anything. One of the uses of this system is in smart homes, where people can control their televisions and music systems and media players without using the remote control or keyboard.

Another big use of this system is to help people. People who have disabilities may have a hard time using the usual devices to control things. The Hand Gesture Controlled Media Player allows them to control media using movements of their hands, which makes it easier for them to do things on their own.

The Hand Gesture Controlled Media Player can also be used in schools where teachers and students use media players to give presentations and learn things online. The Hand Gesture Controlled Media Player is also useful in places like hospitals and laboratories and public places where it's important to keep things clean. The Hand Gesture Controlled Media Player can also be used for things, like games and virtual reality.

X. CHALLENGES

Although gesture recognition systems have benefits there are some challenges to think about when building them. One big challenge is getting the gesture detection right. The system does not always get it right especially if the lighting is bad or the background is busy [5][7].

Another challenge is that people's hands come in shapes and sizes. This makes it hard for the system to

recognize gestures. Real-time processing is also tough because the system has to keep up with video frames without slowing down. The camera quality matters too because low-quality cameras might not be able to detect hand movements

The system also only works with a gesture right now. To add gestures without making mistakes we would need to use more advanced algorithms. This is a challenge because we want to make sure the system is accurate.

We have to consider these challenges when developing gesture recognition systems. Gesture recognition systems can be really useful. We need to make sure they work well.

XI. LIMITATIONS

The system we're proposing has some limitations. One of them is lighting. It works best when the room is well-lit [6]. If it's too dark or too bright the hand detection model won't work right.

- * It also needs users to make gestures within the cameras view

- * The systems performance depends on the computers power. If the computer isn't very powerful it might take a little longer to respond. The system uses -set gesture patterns. This means it can't understand gestures. It also can't learn gestures on its own. With these limitations the system is a simple way to control media with gestures.

XII. FUTURE PROSPECTS

Gesture recognition technology is really going to be big, in the future. We can use machine learning and deep learning to make gesture detection more accurate [15]. The system can also work with voice recognition [9] so users can control media players with their voice and gestures.

Another thing that can be done is to make the system work with TVs, mobile devices and IoT systems. Future systems might use cameras and depth sensors to detect gestures, which will make them work better even when it is dark. Gesture recognition technology will become a way for people to interact with digital things if we keep working on it and making it better.

XIII. CONCLUSION

The Hand Gesture Controlled Media Player that uses Artificial Intelligence shows us how Artificial Intelligence and Computer Vision can make it easier for people to interact with computers [1][2]. This system lets people control the media they are watching using hand gestures so they do not need to use things like a mouse or keyboard. This makes it more natural and easier to use.

The system uses OpenCV and MediaPipe to find the shapes of hands and understand what the gestures mean away [6][12]. It knows what to do when it sees gestures like play or pause a video make the sound louder or softer and move to a different part of the video. Even though it can be hard to get it to work well when the light is not good or the computer is slow the system usually works well when things are normal.

In the future the Hand Gesture Controlled Media Player can be made better by using more advanced Artificial Intelligence techniques, better cameras and working with other smart systems. Controlling things with gestures is going to be a part of how people interact with computers, in the future so the Hand Gesture Controlled Media Player is a good start.

REFERENCES

- [1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson Education, 2021.
- [2] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2022.
- [3] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 2008.
- [4] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012.
- [5] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015.
- [6] Google Research, "MediaPipe: A framework for building perception pipelines," *Google AI Blog / Documentation*, 2020.
- [7] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst., Man, Cybern.*, 2007.
- [8] J. Wachs, M. Kölsch, H. Stern, and Y. Edan, "Vision-based hand gesture applications," *Commun. ACM*, vol. 54, no. 2, pp. 60–71, 2011.
- [9] D. Jurafsky and J. H. Martin, *Speech and Language Processing*. Prentice Hall, 2023.
- [10] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [11] PyAutoGUI Documentation. [Online]. Available: <https://pyautogui.readthedocs.io>
- [12] OpenCV Documentation. [Online]. Available: <https://opencv.org>
- [13] MediaPipe Hands Documentation. [Online]. Available: <https://developers.google.com/mediapipe>
- [14] E. Turban, R. Sharda, and D. Delen, *Decision Support and Business Intelligence Systems*. Pearson, 2019.
- [15] Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.