HandiVision AI: A Confidence-Aware System for Touchless Computer Interaction Using Hand Gestures

Prof. Swati Lokhande[1] Samiksha Dinesh Raut[2] Tanishq Balasaheb Mane[3] Pragati Shivdas Ingale[4] Mansi Satyavan Giranje[5]

[1,2,3,4,5]Department of Artificial Intelligence and Machine Learning

[1,2,3,4,5]Zeal Polytechnic, Pune, Maharashtra, India

Abstract — HandiVision AI is a computer vision system that will allow one to interact with computers without touching using real-time gesture recognition. It will decrease the need of traditional devices like a mouse and keyboard by detecting hand gestures made by capturing the hand movements using a webcam. By using Python, OpenCV and MediaPipe the hand landmarks are detected and gestures can be recognised. These gestures can perform tasks such as moving the cursor, clicking and scrolling. The system has a handwriting/drawing feature allowing users to draw shapes, write letters and make strokes using their fingers. The proposed system also has an active feature allowing users to select the hands they want to interact with whilst using the system. A safety feature has been put into the system to pause any actions being performed if there is uncertainty in tracking or multiple hands are present. This will decrease the amount of accidental commands and drawings. This system demonstrates how artificial intelligence and computer vision can be used in human-computer interactions. The performance of the system can be measured using gesture detection accuracy, speed of interaction, accidental actions and drawing consistency.

Keywords: Hand Gesture Recognition; Human-Computer Interaction (HCI); Computer Vision; MediaPipe Hand Tracking; OpenCV-Based Gesture Detection; AI-Powered Touchless Interface

I. INTRODUCTION

Human-Computer Interaction (HCI) is an important part of how users can communicate with computer systems. Normally, computers are controlled using a mouse, keyboard and touchpad. However, due to the popularity of these controllers, they can become inconvenient due to requiring physical contact. With developments in Artificial Intelligence (AI) and Computer Vision, hand gesture recognition can provide a great way to control computers without physical touch.

Hand gesture recognition is used for allowing computers to understand and interpret hand gestures into commands. Using a webcam the system can interpret hand gestures to perform actions in real-time. Computer vision helps identify key parts of the hand such as fingertips and finger joints. These features can then be used to recognise certain gestures.

HandiVision AI is a proposed AI system which can allow you to interact with computers with hand gestures. It uses Python, OpenCV and MediaPipe to provide real-time gestures using video frames. Once recognised, gestures can be translated into actions for controlling the computer such as moving the cursor, clicking and scrolling. The system also has an air drawing and writing feature. This allows users to draw shapes, write letters and make strokes using hand gestures.

An important feature of HandiVision AI is that it can actively select a hand to track. If there are multiple hands visible on the camera it will try to track the selected active hand. The system also has a safety feature to pause computer actions if it is unsure about tracking your hands. This will help avoid unwanted clicks, commands and mistakes when drawing.

This research aims to produce an easy-to-use touchless system. The system can be used for controlling a computer using hand gestures. The system can also be used for air drawing and writing. The system can also be used to analyse the reliability of gestures and how effective the confidence-based safety features are. This project shows how Artificial Intelligence and Computer Vision can be practically applied to improve Human-Computer Interaction.

II. BACKGROUND AND RELATED WORK

A. Hand Gesture Recognition

Hand gesture recognition can be used to interpret hand motions and translate them into commands. Using computer vision techniques gestures can be identified from webcam footage, without any specialist equipment. This technology can be applied to many different applications such as touchless interface, assistance systems and gesture-controlled computers.

B. Real-Time Hand Tracking Using Mediapipe

MediaPipe Hands is a real time solution for tracking hands. It detects 21 landmarks on their hands such as tips of fingers and joints. This information can be used to identify gestures and track finger movement. MediaPipe will be used to detect and track hand movements using a webcam in HandiVision AI.

C. Touchless Computer Control

Gesture based touch free control of computer systems using pre-defined gestures. Using hand gestures, you will be able to control features on a computer by placing your fingers in certain positions and movements.

D. Air Drawing and Writing

Users can draw and write by moving their finger in front of the camera. By detecting the position of the tip of your finger you can create lines, shapes and/or letters using a virtual canvas. This gives more than just gesture control for computers; it allows users to interactively draw.

E. Active-Hand Selection and Tracking Reliability

When more than one hand is visible in the scene it can be important to select a specific hand to track consistently. There *can* often be a problem with tracking if the hands move, are occluded or there is a change in lighting. Selecting an active

All rights reserved by *www.ijsrd.com*

hand can help to provide consistent tracking of the desired hand.

F. Safety Mechanisms in Gestures-Based Systems

When tracking is not reliable users may accidentally click or move the cursor and draw. This can be avoided by implementing a safety mechanism that pauses actions on the computer when the selected hand is not being tracked reliably. HandiVision AI will implement this safety mechanism however its efficacy must be measured using experiments. Comparing the performance of the system with and without the safety mechanism will provide a measure of how effective it is.

G. Research Gap

Current hand tracking and gesture recognition allows us to interact with computers without using our hands. This study looks into creating an integrated system which allows users to control their computer, draw/write in the air, select using their hands and pause based on uncertainty. The aim of this research is to create an integrated system called HandiVision AI and determine whether this system will have a safe fail mechanism to reduce accidental actions during operation.

H. Gap and Scope

1) Problem

Current hand gesture recognition systems offer a way of controlling a computer without touching the screen. However, there can be issues with tracking the hand reliably, multi hand interference and accidental commands. Many basic versions just allow control of a cursor and do not offer air drawing/writing capabilities or deal with tracking uncertainty.

This project offers an opportunity to create a comprehensive system that encompasses touchless computer control, air drawing, active hand selection and a safety feature to pause the computer if the hand cannot be reliably tracked. This has been proposed within HandiVision AI by creating a system that will pause the computer if the selected hand cannot be reliably tracked. This should be tested to find out how effective it is at reducing accidental commands, increasing responsiveness and improving usability.

2) Project Scope

This project will create a hand gesture recognition system using a webcam. It will be done using Python, OpenCV and MediaPipe. The system will include:

- Real Time Hand Detection – Detecting hand landmarks (such as fingertips and finger joints) using live webcam video.
- Touchless Computer Control – Allowing control of cursor movement, clicks and scrolling using pre-defined gestures.
- Air Drawing/Writing – Detecting finger movements to draw, write letters and make strokes on a virtual canvas.
- Active Hand Selection – Selects and tracks an active hand to reduce any multi hand interference.
- Tracking Uncertainty – Pauses the computer if the active hand cannot be reliably identified.
- Evaluating System Performance – Evaluates accuracy, responsiveness, accidental command rate and continuous drawing of gestures.

The system will be used on a desktop/laptop with a webcam. Depending on the quality of the webcam, lighting, background complexity and visibility of hands the system may have varying levels of performance. The system will only use pre-defined gestures and will not interpret sign language freely. The effectiveness of the system will be determined by practical testing.

III. Threat Model and Design Principles

A. Threat Model

HandiVision AI will be using a webcam to take input and recognise hand gestures to carry out functions on the computer. Potential errors can occur due to environmental factors, inaccurate hand detection and inaccurate hand movements. Below are the possible threats which can occur:

1) Misclassification of Gestures – A hand gesture is recognised as a different gesture causing unintended actions on the computer.
2) Multi Hand Detection – Multiple hands can be picked up by the webcam causing incorrect tracking of the hand leading to incorrect cursor movements or drawing.
3) Lost Tracking – Hands may become partially hidden, move quickly or be blocked by objects causing lost tracking.
4) Poor Lighting and Background – Poor lighting and complex background can cause issues with detecting the hand landmarks.
5) Accidental Actions – Accidental finger movements can cause unintended clicking, scrolling or drawing due to an incorrectly classified gesture.
6) Unauthorised Inputs – The webcam could detect gestures from another person if they are visible on screen. These gestures could cause unintended actions on the computer without the correct protections in place.

The proposed system will attempt to limit the risks by using active hand selection, ensuring accurate tracking, using pre-defined gestures and pausing actions if the tracking becomes unsure. This will improve accuracy but there may still be some errors.

B. Design Principles

The design principles below will be used to ensure the system can provide accurate, easy and convenient interaction.

1) Fast Response – The webcam and hand movement must be processed quickly.
2) Accurate – The hand landmarks and pre-defined gestures must be recognised accurately.
3) Active Hand Tracking – The system must attempt to keep tracking on the active hand. If multiple hands are detected this should have less of an effect.
4) Safe – If the system can't identify the active hand then computer actions should be paused. Only when the system is confident with tracking will actions resume.
5) Non Contact – The user should be able to perform computer actions and draw on the virtual canvas without having to touch any traditional input devices.
6) Modular – The webcam input, hand detection, gesture recognition, computer control, air drawing and safety should all be divided into modules.

All rights reserved by www.ijsrd.com

7) User Friendly – The interface should give visual feedback on what the webcam can see, recognised gestures, what mode the system is in and what the tracking status is.
8) Secure – The webcam should only process what is needed for the application. The video shouldn't be saved or sent and the computer control should only do what is supported.

This will provide a solid base for creating a non contact computer interaction system. The safety features in the proposed system should be tested in different lighting, hand positions and multi-hand situations to check how effective they are.

IV. PROPOSED ARCHITECTURE AND WORKFLOW

A. System Design

HandiVision AI has been developed in a modular fashion. The system uses the OpenCV library to take live video input from a webcam. The MediaPipe library is used to detect and track key points of the hand. Using Python the gestures detected will be processed and the computer will perform an operation via the PyAutoGUI library. Air drawing and writing can also be done using the system. The system includes a function to detect which hand is being used. This reduces the issue of multiple hands by pausing when an unsure tracking is detected.

B. System Flow

Live video is captured using the webcam and hand key points are detected using MediaPipe. The active hand is selected and tracked. The tracking reliability is checked and gestures are recognised based on a set of gestures. Depending on the selected mode the computer will perform actions such as moving the mouse, clicking and scrolling. Alternatively the user can air draw/write on a virtual canvas. The results are displayed on the UI and the process is repeated until the application is stopped.

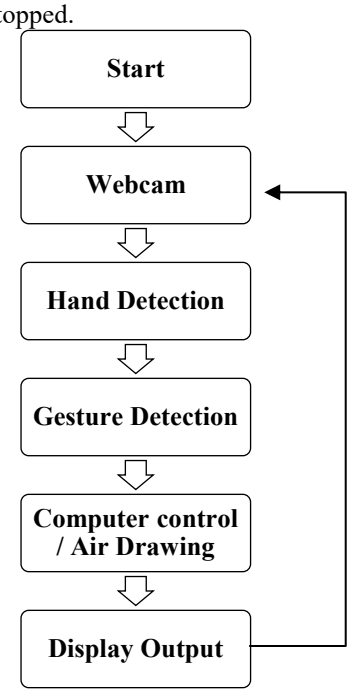


Fig. 1: Architecture Diagram of HandiVision AI

Figure 1. This project aims to provide a user with a contactless way of interacting with a computer through hand gestures. A webcam will be used to monitor and capture the hands movements. This video will then be processed using OpenCV, which is useful for processing video images. MediaPipe will be used to detect a users hand and locate key points of interest on the hand such as fingertips and joints.

After detecting the hand, the hand will be tracked. This is done by analysing the position of the fingers to recognise what gesture they are making. We can then turn these gestures into specific commands using Python, and PyAutoGUI. This allows the user to control the mouse by moving it, clicking and scrolling.

The system has an air drawing feature. By detecting the movement of the fingertips, this will allow the user to draw/write on a virtual canvas. This can be useful for simple drawing.

The system also has checks to ensure it is working correctly. If more than one hand is detected or the active hand cannot be tracked, the system can pause for a short period of time to prevent accidental commands. The recognised gestures and drawing will be shown on the screen.

Overall, this system provides a simple user friendly system, by incorporating computer vision, hand tracking and gesture recognition.

V. METHODOLOGY AND EVALUATION PLAN

The system will work by taking live video from the webcam. Using OpenCV the video is processed frame by frame. The video will be analysed by MediaPipe to detect the hand and find the landmark points. This will allow the system to track the detected hand and know where each finger is.

From this the system will be able to move the cursor around the screen, click and scroll. Users will also be able to draw and write on a virtual canvas by moving their fingertips. An active hand selection will be used so the system can track consistently. If the system is not detecting the hand it will pause so it does not carry out any unwanted actions.

A. Assessment

The system will be tested to see how well it detects the hand, recognises gestures, controls the cursor and draws in the air. The system will be assessed on how accurate it is at recognising gestures, how fast it responds, how consistent the tracking is and how many unwanted actions occur. This will help identify any issues with the system.

B. Testing

The system will be tested using a webcam, with different lighting conditions, backgrounds and hand positions. The system will be tested with multiple hands within the frame to ensure that the active hand selection works and that it tracks consistently.

VI. DISCUSSION

The HandiVision AI system is an alternative method of interacting with a computer. Computer vision and hand gesture recognition allow basic functions within a computer to be performed using simple hand gestures. OpenCV and

All rights reserved by www.ijsrd.com

MediaPipe are used to detect and track the hands using a webcam.

There is an air drawing and writing functionality which makes the system more versatile and easy to use. Active hand selection helps the computer to consistently track the hands. A pause functionality means that there will be less accidental actions caused by inconsistent hand detection.

It may suffer from issues such as bad lighting, changing backgrounds, hands being obscured and incorrect gesture detection. It will also depend on the quality of the webcam and correct hand placement. Overall it has the potential to make interacting with computers more convenient and accessible using touchless interaction.

VII. Conclusion

The HandiVision AI provides an easy way of interacting with computers without touching the screen using hand gestures. This is done by using the OpenCV, MediaPipe and Python libraries to detect and track the movement of the hands from the webcam. This enables features such as moving the cursor, clicking, scrolling and air drawing/air writing on the virtual canvas.

The active hand selection and tracking features will ensure a reliable experience with less unwanted actions. Although factors such as lighting, changes in backgrounds and how accurate the hands are detected may affect the performance of the system this does provide a good way of interacting with computers without touch.

The system could be improved in the future by adding more gestures, improving the accuracy of detecting the gestures and including more computer features. Overall this aims to make interacting with computers easier, more versatile and friendly.

Acknowledgment

First and foremost, we would like to thank our project guide Prof.Sushmita Choure for guiding us and providing us with the support and motivation throughout the course of our project HandiVision AI.

We thank the Principle and faculty members of Zeal Polytechnic, Pune for their support. We also thank our department for the opportunity of doing this project.

Finally, we would like to thank our team members, friends and families for motivating us and working with us towards successfully completing this project and research paper.

References

[1] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C. L. Chang, M. Grundmann, "MediaPipe Hands: On-device Real-time Hand Tracking". arXiv preprint arXiv:2006.10214, 2020. Available at https://arxiv.org/abs/2006.10214

[2] S. S. Rautaray, A. Agrawal, "Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey", Artificial Intelligence Review, Vol. 43, pp. 1-54, 2015.

[3] V.I. Pavlovic, R. Sharma, T.S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 19, No.7, pp. 677-695, 1997.

[4] S. Mitra, T. Acharya, "Gesture Recognition: A Survey,", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 37, No. 3, pp. 311-324, 2007.

[5] T.S. Huang, V.I. Pavlovic, "Hand Gesture Modeling, Analysis, Synthesis", in Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, 1995.

[6] J.P. Wachs, M. Kölsch, H. Stern, Y. Edan, "Vision-Based Hand-Gesture Applications", Communications of the ACM, Vol. 54, No. 2., pp. 60-71, 2011.

All rights reserved by www.ijsrd.com