Virtual AI Mouse by detecting Hand Landmarks

Team : Sharath Kumar Karnati
Wsu id: 011852253

October 27, 2024

Abstract

The AI Virtual Mouse using Hand Tracking leverages computer vision and machine learning to create an intuitive and contactless human-computer interaction system. By employing advanced hand detection algorithms and real-time tracking techniques, this project replaces traditional input devices such as the physical mouse with dynamic hand gestures. Utilizing libraries like OpenCV and MediaPipe, the system detects hand landmarks and interprets various gestures to perform standard mouse functions such as cursor movement, clicking, scrolling, and drag-and-drop. This technology enhances accessibility, particularly for users with physical limitations, and offers a more natural, hygienic, and immersive way to interact with computers. Through efficient gesture recognition and real-time processing, the virtual mouse aims to improve user experience in both personal computing and emerging fields such as virtual reality and augmented reality.

1 Introduction

The rapid advancements in artificial intelligence (AI) and computer vision have transformed how humans interact with technology. One of the most promising areas of innovation is the development of touchless, gesture-based interfaces that provide more intuitive and accessible ways to control digital devices. Traditional input methods like keyboards and mice, while functional, are limited in flexibility and can present challenges for users with physical disabilities. To address these limitations, this project focuses on creating an AI-based virtual mouse using hand tracking technology, allowing users to control computer systems with hand gestures in real-time.

The core objective of this project is to replace the conventional mouse with a virtual alternative that tracks hand movements and interprets gestures for actions such as moving the cursor, clicking, scrolling, and dragging. By utilizing hand landmarks and tracking algorithms, this system aims to improve user experience, enhance accessibility, and offer a hygienic alternative to physical input devices. The applications for such a system extend beyond personal computing, with potential use cases in virtual reality (VR), augmented reality (AR), and smart environments. This project explores the development of a real-time hand gesture recognition system, which serves as the foundation for creating a responsive and accurate virtual mouse.

2 Literature Review

The concept of gesture-based interaction has gained traction in the past decade, particularly with advancements in computer vision and AI. One of the earliest systems, Microsoft Kinect, allowed users to control applications through body gestures but was limited by hardware requirements.

Recent developments have focused on software solutions using standard cameras for gesture recognition, with hand detection and tracking being pivotal. Traditional methods relied on color segmentation and shape recognition, but these were often hindered by lighting and background noise. The shift to deep learning models, particularly convolutional neural networks (CNNs), has significantly improved hand tracking accuracy.

Google's MediaPipe framework has simplified the creation of hand tracking systems, providing efficient real-time landmark detection that models hand key points in 3D space. This has facilitated the use of gestures for controlling virtual objects and user interfaces, paving the way for applications like virtual mouse functionality.

Gesture recognition has also seen significant interest, with gestures such as pinch and swipe being extensively studied. Research by Li et al. (2020) and Zhou et al. (2019) demonstrated the effectiveness of deep neural networks and CNNs in achieving high accuracy in gesture classification for human-computer interaction.

Despite these advancements, challenges like system responsiveness and accuracy under varying lighting conditions remain. This project aims to build on existing systems by optimizing hand tracking performance to create a seamless and efficient gesture-based input method.

3 Technical Plan

Overview

This project aims to develop an AI-based virtual mouse utilizing hand tracking technology to enhance human-computer interaction. The methodology leverages deep learning algorithms and real-time hand tracking frameworks to recognize hand gestures and translate them into mouse movements.

Methodology

1. **Literature Review**: Analyze existing research on hand gesture recognition and tracking techniques, focusing on the use of deep learning methods and frameworks like MediaPipe.

2. **Framework Selection**: Choose the appropriate tools and frameworks, such as MediaPipe for hand tracking and TensorFlow or PyTorch for implementing machine learning models.

3. **Model Development**: Implement deep learning models (e.g., CNN, RNN) for gesture recognition. Fine-tune hyperparameters to enhance model accuracy.

4. **Integration**: Combine the hand tracking and gesture recognition modules to translate detected gestures into mouse commands.

5. **Testing and Validation**: Test the integrated system in various environments and refine based on results that i will get.

6. **Documentation**: Document findings, methodologies, and results throughout the project lifecycle.

3.1 Work-flow chart:

[Start] → [Capture Video] → [Detect Hands]
→ [Find Landmarks] → [Recognize Gestures]
→ [Control Mouse] → [Display Feedback] → [End]

4 Intermediate Results

So far, I had successfully implemented the hand tracking functionality using MediaPipe. The following observations and knowledge have been acquired:

- **Real-time Hand Detection**: The system accurately detects hands and identifies key landmarks, providing a solid foundation for gesture recognition.

- **Gesture Interpretation**: Basic gestures, such as moving the index finger for cursor movement and clicking by pinching the index and middle fingers, have been successfully implemented.(but still not fully functional there are some issues in the code)
- **Performance Insights**: The system exhibits responsiveness under varied lighting conditions, though further optimization is needed for consistent accuracy for hand tracking.

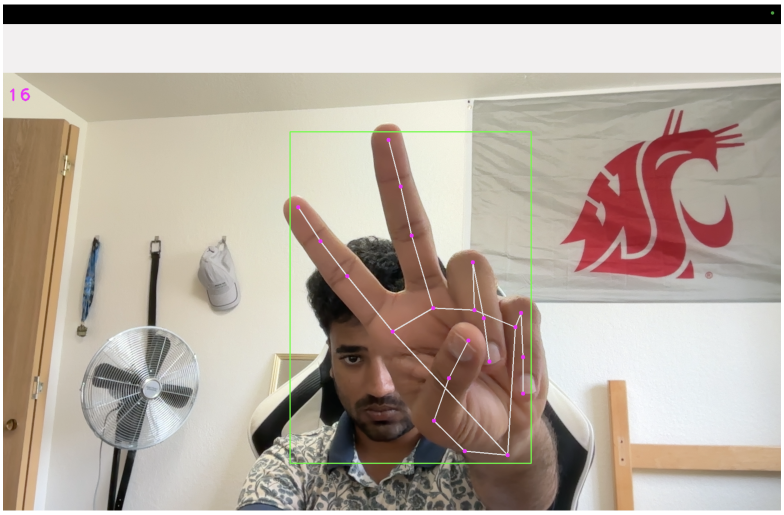

Figure 1: Hand Tracking results.

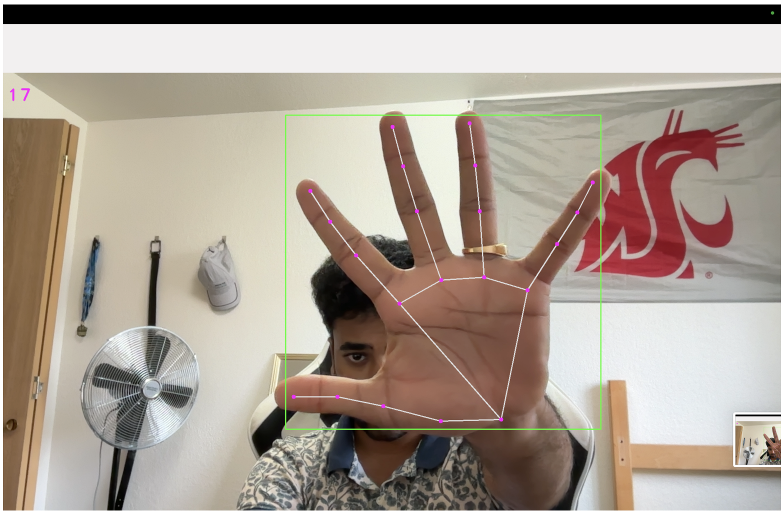

Figure 2: Hand Tracking results.

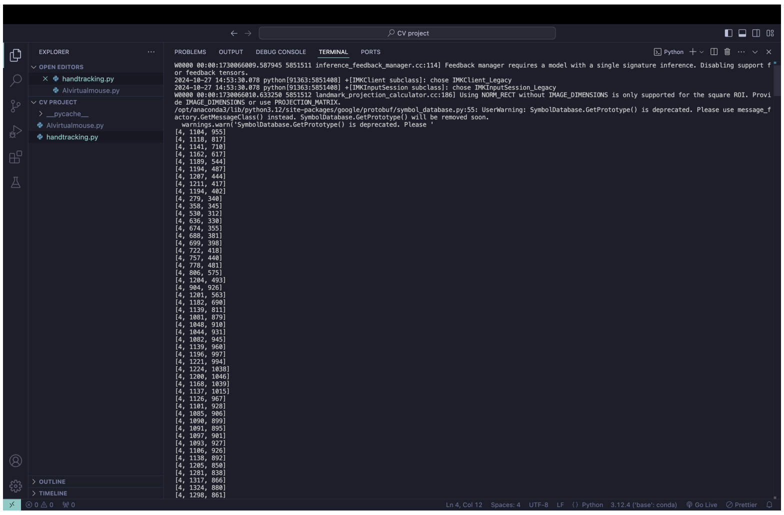

Figure 3: Hand Tracking Results in Terminal.

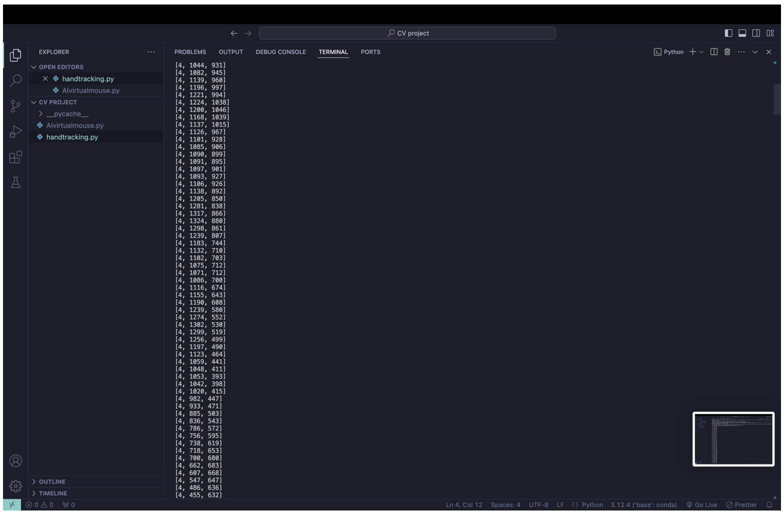

Figure 4: Hand Tracking Results in Terminal.

The more information will be provided in the final report.

5 Future Work

To enhance the project's effectiveness, the following areas will be pursued:

- **Gesture Expansion**: Implement additional gestures to broaden mouse functionalities, such as scrolling and dragging.
- **System Optimization**: Refine the performance of gesture recognition under different environmental conditions, potentially integrating machine learning techniques for improved accuracy without relying on pre-collected training data.
- **User Testing**: Conduct user experience studies to gather feedback and improve interaction intuitiveness.

References

[1] Li, S., Li, W., Liu, Z., Liu, W., and Peng, H. (2020). *Deep learning for hand gesture recognition: A review*. IEEE Access, 8, 77612-77636.

[2] Zhou, X., Zhou, Z., Du, H., and Zhao, X. (2019). *Gesture recognition for human-computer interaction using deep learning*. Journal of Visual Communication and Image Representation, 58, 579-589.

[3] Mehta, D., Liu, P., Bylinskii, Z., and Olwal, A. (2020). *MediaPipe Hands: On-device Real-time Hand Tracking*. Google Research.

[4] Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., and Kautz, J. (2016). *Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4207-4215.

[5] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. (2011). *Real-time human pose recognition in parts from a single depth image*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1297-1304.