GESTURE PLUS:
A Novel Approach to Enhance Interactive Media

GROUP 01

GUIDE

PROF. SAVITA LADE

B820 KASTURI PEDNEKAR

B831 ARYA SAHANE

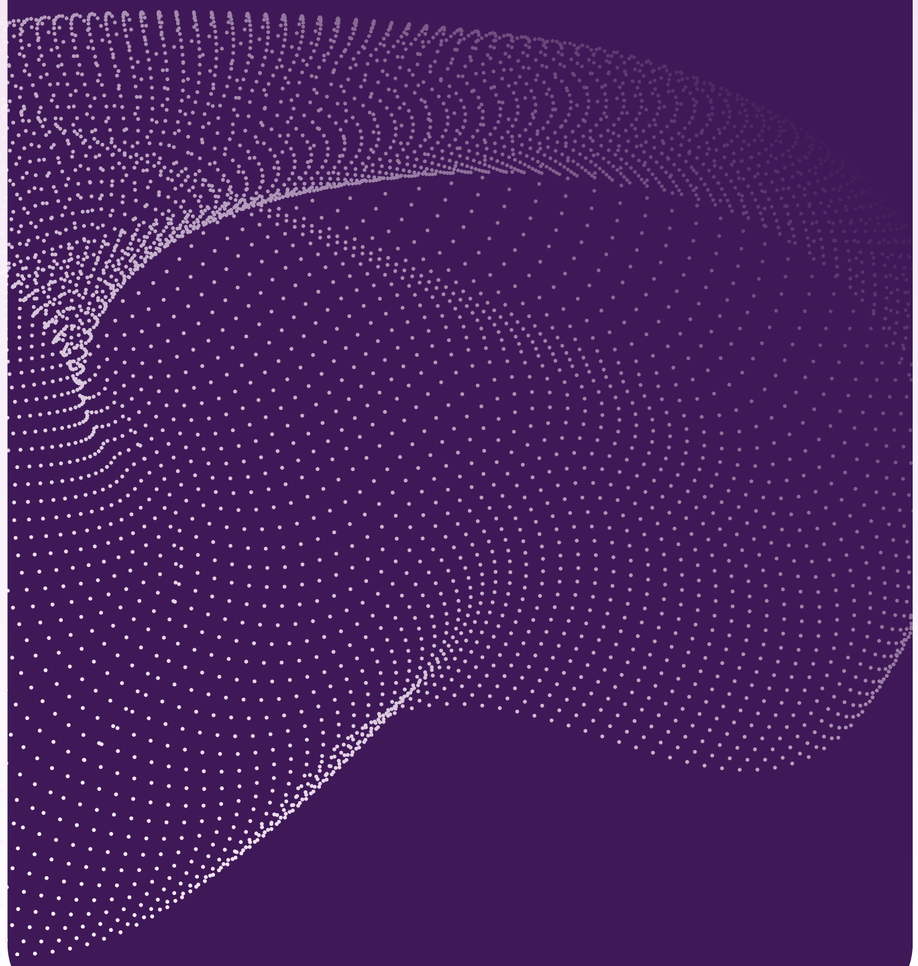
B836 SHAIKH UZAIR AFNAN

B858 AKANSHA WANKHADE

Affiliation

MCT's Rajiv Gandhi Institute of Technology, Mumbai University

Abstract



GesturePlus is an advanced technology designed to enhance human-computer interaction through a combination of voice commands, hand-gesture recognition, and a chatbot integrated with a voice module. It aims to bridge the gap between users and digital interfaces by providing a seamless, intuitive, and touch-free experience.

Core Technologies:

- Machine Learning and Computer Vision for gesture recognition
- A chatbot for voice and text interaction and command generation
- Multi-platform compatibility for enhanced accessibility

Objectives & Research work

1. Develop an Intuitive Human-Computer Interaction System

2. Achieve High Accuracy in Gesture -Recognition with minimum Latency

3. To Integrate Voice Command Functionality for hands-free system control

4. Ensure Real-Time Responsiveness

5. Create a Versatile Multi-Modal Interface

6. Promote Scalability and Future Upgradability

TITLE	AUTHORS	SOURCE	YEAR	DESCRIPTION
Real-time Virtual Mouse System using RGB-D Images and Fingertip Detection	Dinh-Son Tran, Ngoc-Huynh Ho, Hyung-Jeong Yang, Soo-Hyung Kim, Guee Sang Lee	Springer	2020	In the realm of human-computer interaction (HCI), creating a real-time fingertip-gesture-based interface remains challenging due to sensor noise, varying light conditions, and the complexity of tracking fingertips across diverse subjects. One popular method for interacting with computers without a physical mouse device is using fingertip tracking as a virtual mouse.
Gesture Recognition Based Virtual Mouse and Keyboard	B. R. Sandhya, C. Amrutha, S. Ashika	Springer	2023	The system employs hand gestures for virtual mouse and keyboard control, highlighting challenges like gesture recognition accuracy and environmental dependencies.
Human-Machine Interaction Sensing Technology Based on Hand Gesture Recognition: A Review	Lin Guo; Zongxing Lu; Ligang Yao	IEEE	2021	This paper provides a comprehensive overview of hand gesture recognition (HGR) technology, focusing on the various sensing methods employed. It explores the challenges and opportunities in HGR research, aiming to guide future advancements in human-machine interaction.

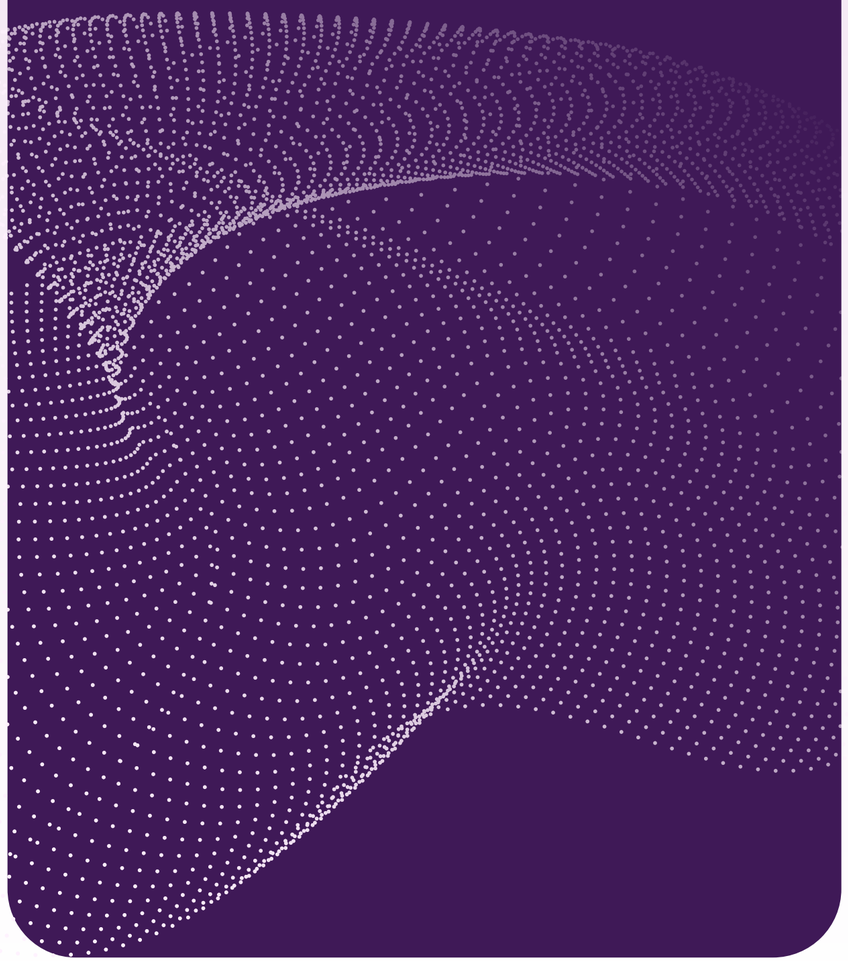
TITLE	AUTHORS	SOURCE	YEAR	DESCRIPTION
Hand Gesture Recognition Based on Computer Vision: A Review of Techniques	Munir Oudah, Ali Al-Naji, Javaan Chahl	MDPI	2020	This paper reviews various computer vision techniqu for hand gesture recognition, discussing their merits and limitations under different circumstances.
Methods, Databases and Recent Advancement of Vision-Based Hand Gesture Recognition for HCI Systems: A Review	Debajit Sarma, M. K. Bhuyan	Springer	2021	This paper provides an overview of vision-based han gesture recognition methods, including recent advancements and applications in human-compute interaction.
Wizard of Wikipedia: Knowledge-powered conversational agents	Emily Dinan , Stephen Roller , Kurt Shuster , Angela Fan, Michael Auli, Jason Weston	Conference paper at ICLR 2019	2019	This research introduces a large dataset of open-doma conversations grounded in Wikipedia knowledge, addressing the lack of knowledgeable dialogue benchmarks. It also presents architectures that can retrieve, process, and use this knowledge to generat informed responses, demonstrating improved performance in knowledge-based open-domain conversations as measured by both automatic metri and human evaluations.

Proposed System

GesturePlus is an innovative human-computer interaction system that leverages advanced machine learning and computer vision technologies to enable users to navigate and interact with computers through hand gestures and voice commands. The system is designed to provide a hands-free, intuitive method of engagement that can be particularly useful in sterile environments or for users with limited physical mobility.

- **Multimodal Interaction:** Combines gesture and voice-based inputs for an alternative, intuitive way to interact with computers.
- **Real-Time Response:** Detects palm movements and voice commands for actions like clicks, volume/brightness control, file navigation, and web search.
- **Cross-Platform Compatibility:** Works on macOS, and Windows using a modular architecture for adaptability.
- **Need for Alternative Input:** Reduces reliance on traditional keyboard/mouse input with a user-friendly, intuitive design.
- **Challenges:** Ensuring accurate hand tracking, robust gesture recognition, voice command processing, and seamless module integration.

Methodology



Gesture Recognition:

- Captures video via OpenCV, preprocesses it, and detects hand landmarks using MediaPipe.
- Calculates distances and ratios to classify gestures using numpy and Python's math library.
- Uses temporal smoothing for stable gesture recognition.
- Executes actions like mouse control, scrolling, or brightness/volume adjustments based on OS-specific libraries.

Voice Command Processing

- Captures audio in real-time using SpeechRecognition, adapting to ambient noise.
- Uses Google's speech-to-text API for transcription.
- Parses commands using a rule-based system to perform actions (e.g., web search, file navigation).
- Implements text-to-speech responses via pyttsx3 for user feedback.

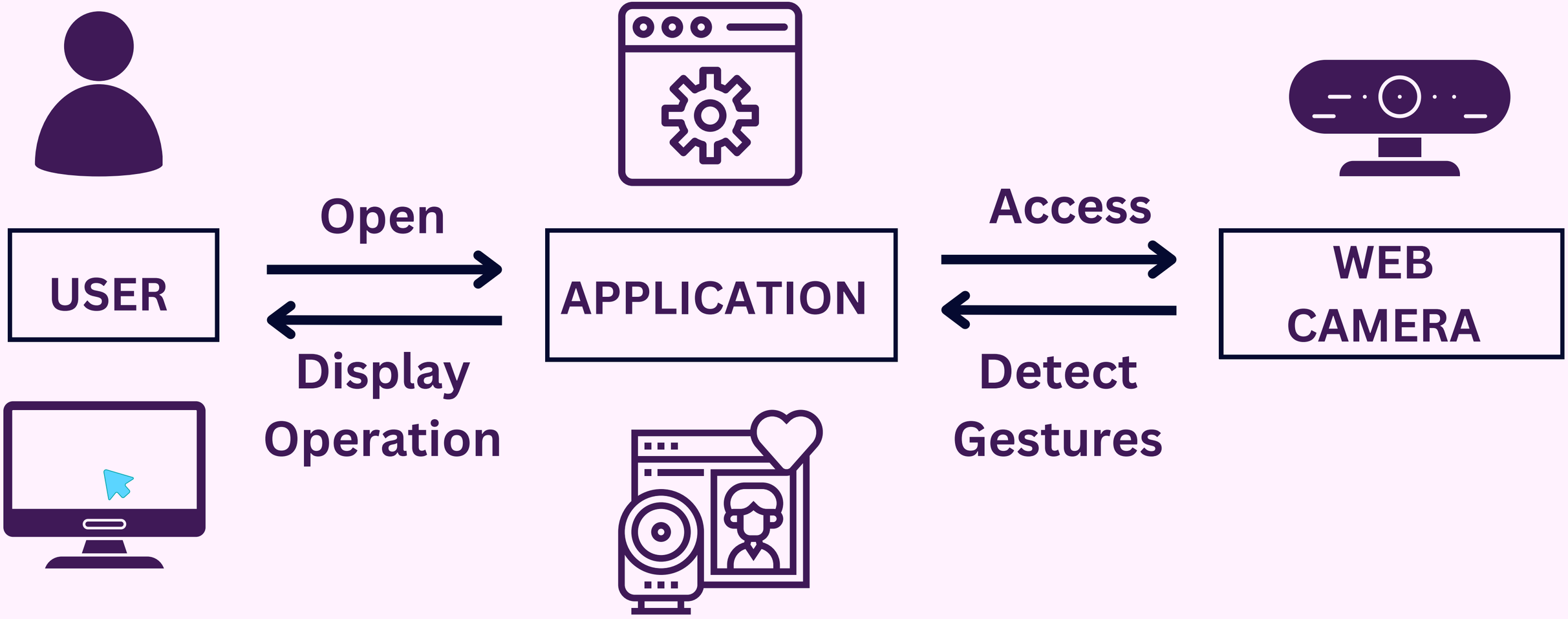
Parallel Processing

- Uses Python's and Go's multiprocessing to run gesture and voice modules independently, reducing latency and improving efficiency.

Resource Monitoring

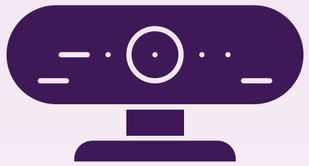
- Uses psutil to track CPU, memory, and I/O usage, visualizing trends with Plotly for performance analysis.

2 Way Collaborative Interface – HGR



Hand Gesture Recognition

Hand Gestures



Web Camera

Real-Time Hand Image Input

Image Preprocessing

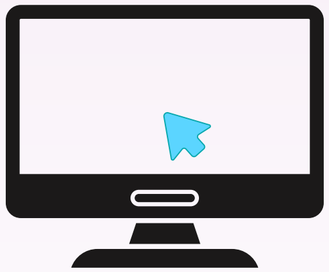
Image Processing

Tracking Hand Movement Region

FingerTip Detection

Recognized Meaningful Gesture

Align mouse functions wrt Hand Gestures



PC Mouse works on Hand Gestures

Techstack – Hand Gesture Recognition

INPUT

WEB CAMERA
REAL TIME
IMAGES FRAMES

PROCESS

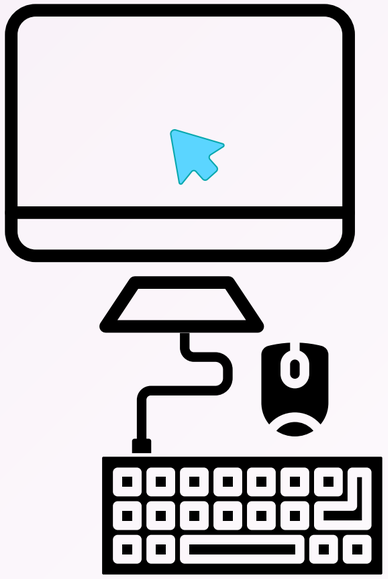
EDITOR
Pycharm or VSCode

COMPUTER INPUT INTERACTION
PyautoGUI
GUI
Tkinter

HAND GESTURE MODULE
MediaPipe

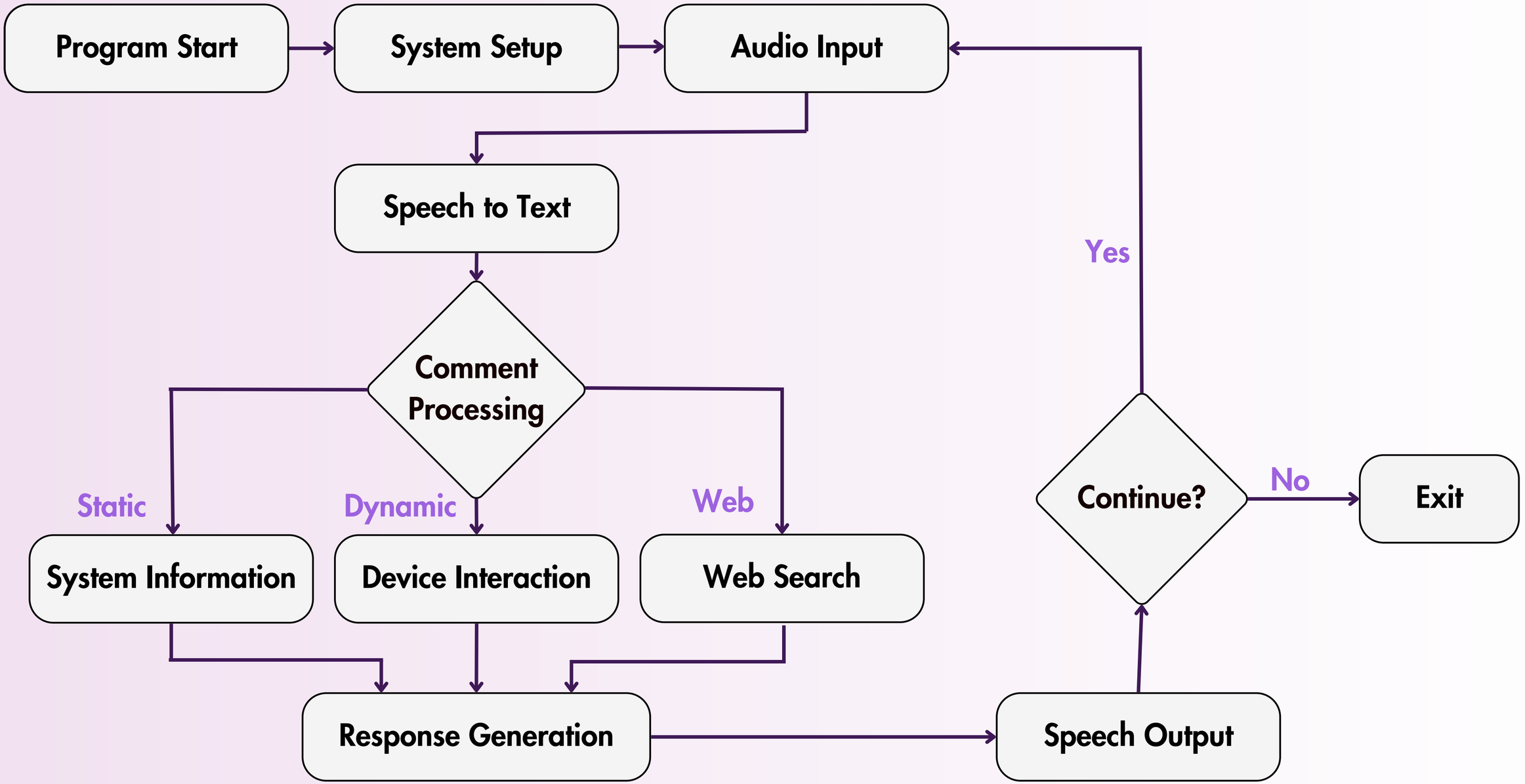
OPEN SOURCE COMPUTER VISION
OpenCV

OUTPUT



HAND GESTURE BASED MOUSE ACTION

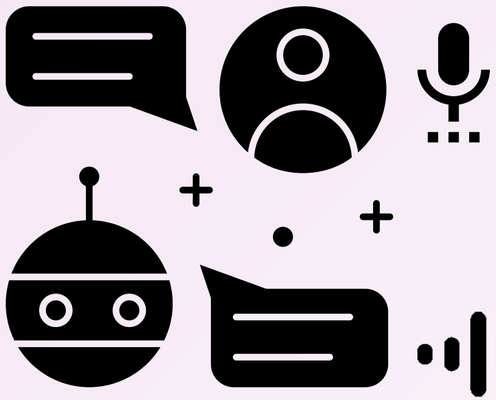
Voice Command Module



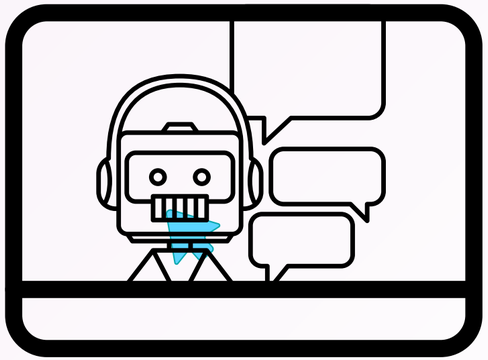
Techstack – Voice Command Module

- pyttsx3 – For text-to-speech (TTS) conversion.
- speech_recognition – For speech-to-text conversion.
- webbrowser – To open web pages.
- pynput.keyboard – For keyboard control and automation.
- pyautogui – For GUI automation (mouse and keyboard).

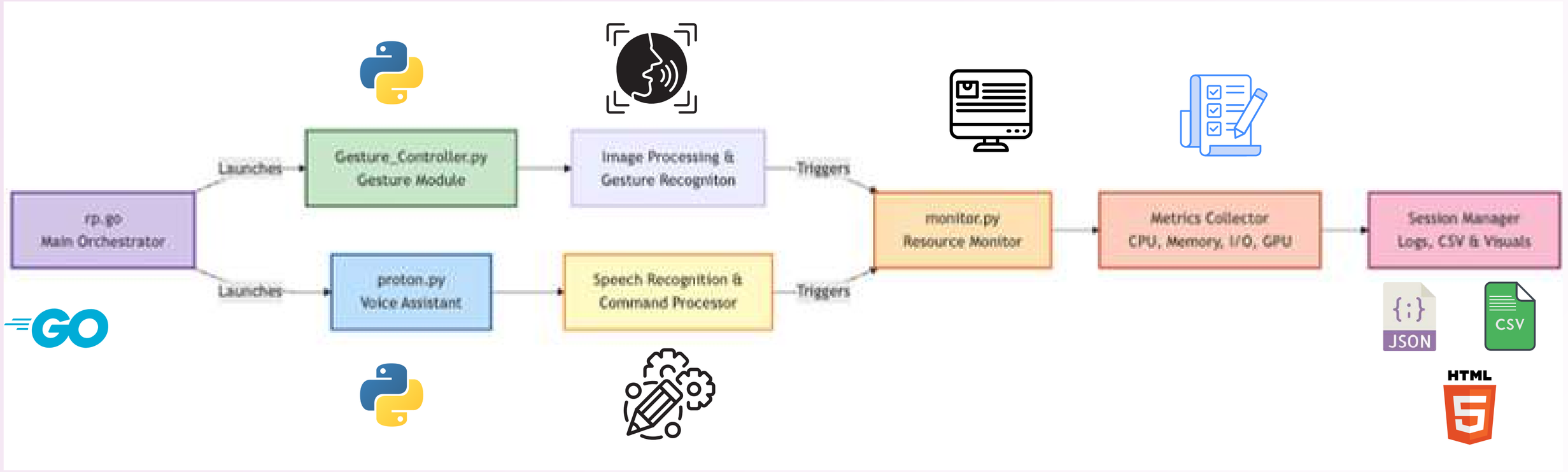
INPUT



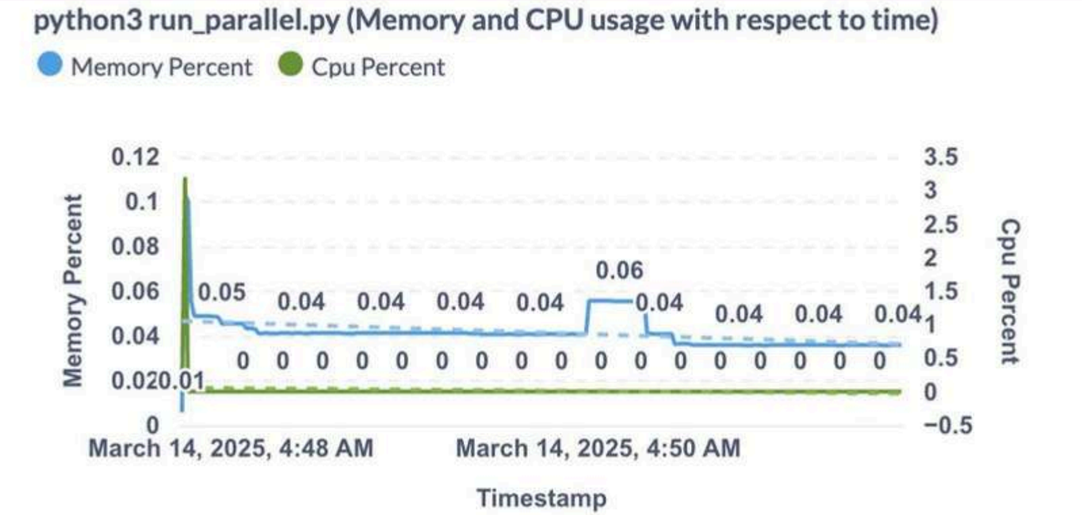
OUTPUT

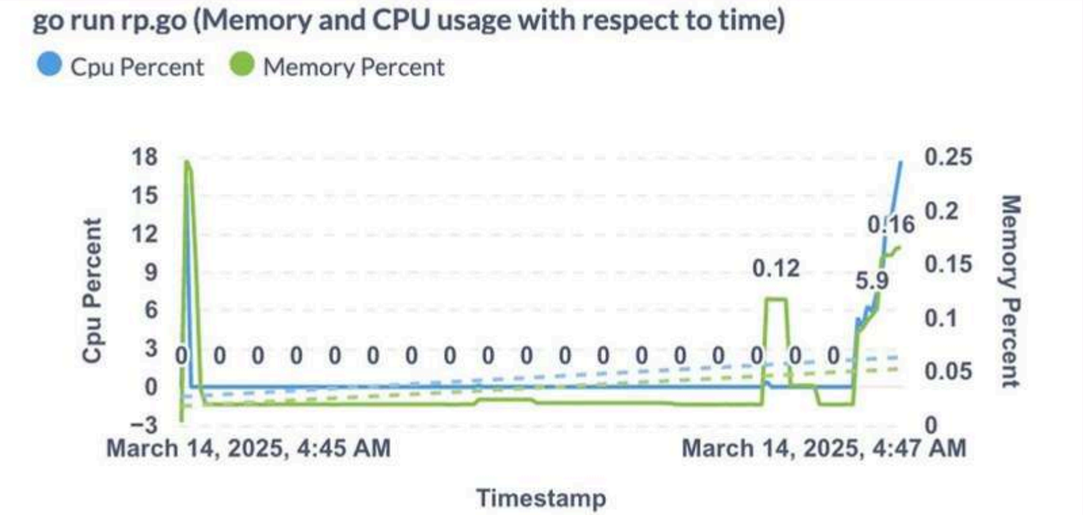


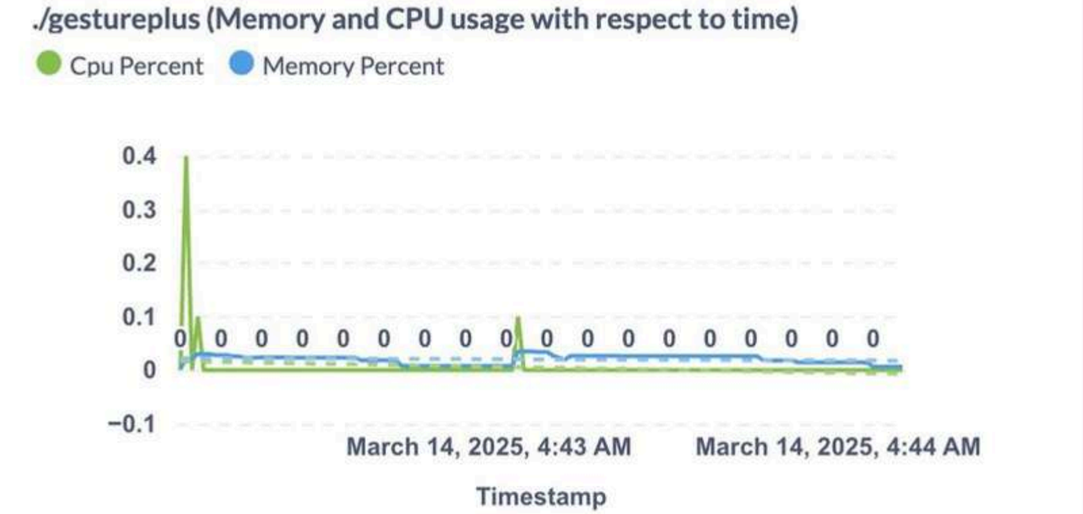
Overview of System



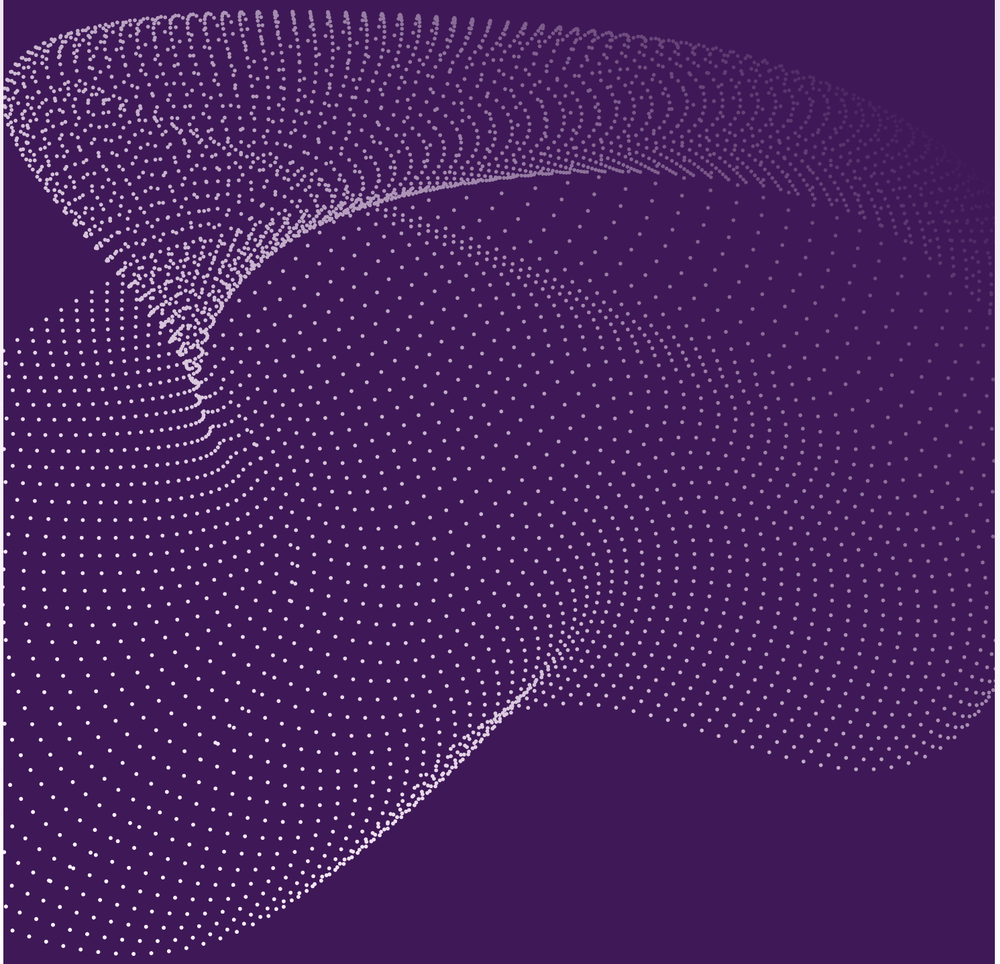
Results


python3 run_parallel.py (Memory and CPU usage with respect to time)
Memory Percent
Cpu Percent
Memory Percent
Cpu Percent
Timestamp
March 14, 2025, 4:48 AM
March 14, 2025, 4:50 AM


go run rp.go (Memory and CPU usage with respect to time)
Cpu Percent
Memory Percent
Cpu Percent
Memory Percent
Timestamp
March 14, 2025, 4:45 AM
March 14, 2025, 4:47 AM


./gestureplus (Memory and CPU usage with respect to time)
Cpu Percent
Memory Percent
Timestamp
March 14, 2025, 4:43 AM
March 14, 2025, 4:44 AM

Future work



- **ASL Sign Language Integration:** Incorporating American Sign Language (ASL) recognition for better accessibility, enabling communication for the hearing-impaired community.
- **VR Hardware Support:** Enhancing applications for gaming, simulations, and immersive environments by enabling natural interactions in virtual spaces.
- **Customizable Gestures & Voice Commands:** Enabling users to define their own gestures and voice commands instead of relying solely on predefined inputs.
- **AI-Powered Chatbot Improvements:** Enhancing context-aware interactions and natural language understanding for a more refined user experience.
- **Scalability & Industry Applications:** Making GesturePlus more adaptable to various industries and use cases through improved flexibility and intelligence.

Conclusion

In this research we were able to

- Create a Multi-modal system that can use both gestures and voice to command the system
- Address the limitations of other researches and resolve them
- Make a resource-efficient system that can perform without any bugs or major performance issue
- Lastly, creating a resource tracking module that can show at the resource usage at the end of the session

Thank You!