

CANVAS AI: A Gesture-Driven Visual Question Answering Interface

By: Arun Sundar , Akash Sriram Prasad and Aparajith Krishna

Dr. Radhika.P , Assistant Professor

Department of Electronics and Communication Engineering, SRM Institute of Science and Technology,
Kattankulathur

INTRODUCTION/ABSTRACT

This project presents CANVAS AI, a novel software interface that combines computer vision, gesture recognition, and generative AI to create an interactive canvas for natural human-computer communication. Users can draw using hand gestures and interact with an AI system capable of analyzing canvas content and responding through speech. This innovation merges intuitive control with visual intelligence for educational and creative applications

MOTIVATION

Traditional input tools like keyboards or mice pose limitations in accessibility and intuitive control, especially in educational or sketch-based environments. This project is motivated by the need for a touchless, intelligent interaction system where users can draw freely and ask contextual questions.

PROBLEM STATEMENT/OBJECTIVE

- To develop a local software platform that:
- Captures hand gestures using a webcam.
 - Converts fingertip motion into drawing on a virtual canvas.
 - Analyzes the canvas content (e.g., questions, math problems, diagrams).
 - Provides AI-generated spoken answers using vision-language models.

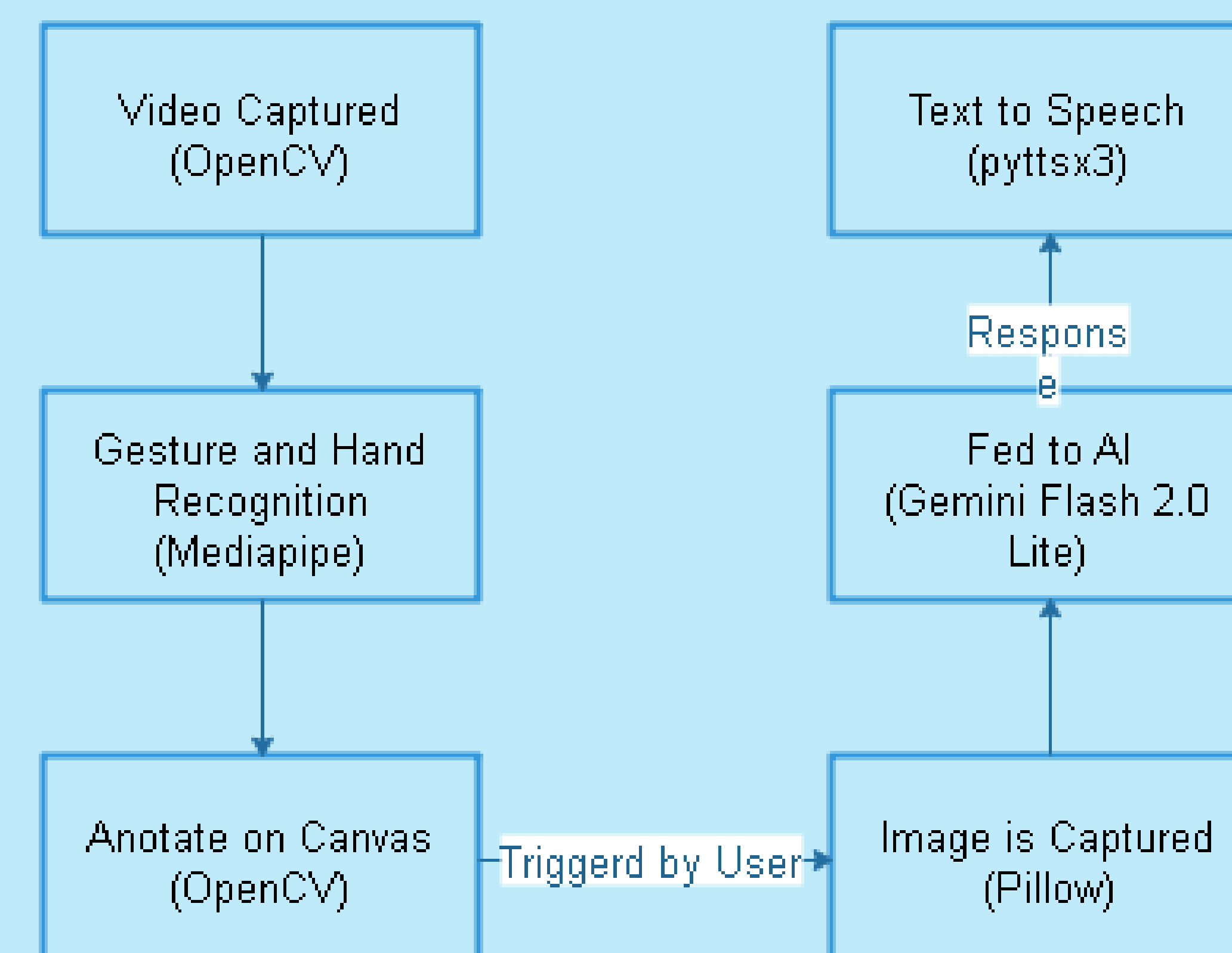
METHODOLOGY

- **Gesture Detection:** *Using OpenCV and Mediapipe to track hand landmarks, especially the index fingertip.
- **Canvas Drawing Module:** Real-time rendering of fingertip movement on a virtual canvas with color, thickness, and erasing options.
- **Screenshot Capture:** The drawn canvas is saved as an image when user triggers a query.
- **Vision-Based Question Analysis:** *Google Gemini API (or similar) processes the canvas image and extracts semantic meaning.
- **Speech Output:** Text responses are converted to speech using a TTS engine (e.g., pyttsx3 or gTTS).

DESIGN

- **Frontend:** Python (Tkinter/Streamlit or OpenCV GUI)
- **Gesture Tracking:** Mediapipe Hands
- **AI Backend:** Gemini Vision API or Open Source Alternatives (e.g., BLIP, MiniGPT-4)
- **Speech:** pyttsx3 / gTTS
- **Optional:** Whisper/Coqui for speech-to-text if enabling voice input in future

BLOCK DIAGRAM



RESULTS & INFERENCE

- Successfully tracks fingertip for smooth canvas drawing.
- AI accurately interprets questions drawn on canvas (MCQs, equations, diagrams).
- Fast and intelligible voice feedback (~2-4 seconds response latency).
- Effective for demo-level educational tools, accessible interaction, and creative sketching.
- Lightweight enough to run on mid-range systems (e.g., GTX 1650).

CONCLUSION

CANVAS AI introduces a touchless and intelligent way to interact with machines. It bridges the gap between gesture-based inputs and visual AI reasoning. The system holds potential for applications in education, design, and accessibility solutions.

REFERENCES

1. Google Gemini Vision API Documentation
2. OpenCV & Mediapipe Hand Tracking Guides
3. BLIP: Bootstrapped Language-Image Pretraining
4. MiniGPT-4 GitHub Repository
5. pyttsx3 and gTTS Python Libraries
6. SRM Course Material & Mentor Discussions

TEAM MEMBERS

- ARUN SUNDAR (RA2311004010150)
- AKASH SRIRAM PRASAD (RA2311004010177)
- APARAJITH KRISHNA (RA2311004010170)