SYNOPSIS

Synapse OS: A Multimodal Neuro-Symbolic Architecture for Context-Aware Desktop Automation

Submitted by

Rohith B

USN: 1MS25SCS032-T

Department of Computer Science and Engineering

MS Ramaiah Institute of Technology, Bangalore

Under the Course: Advanced Computer Intelligence

M.Tech in Computer Science and Engineering

Academic Year: 2025 - 2026

Abstract

Traditional operating systems function as passive tools, responding only to explicit user input while remaining unaware of contextual factors such as user presence, attention, or environmental conditions. This project proposes **Synapse OS**, a novel *Perceptive Operating System* that integrates **Computer Vision** and **Large Language Models (LLMs)** to create an adaptive, intelligent, and secure desktop environment.

The architecture is built upon a **dual-sensory input system** comprising:

1. **Visual Perception Layer** – Employing OpenCV-based algorithms for face detection, pose estimation, and object recognition, this layer observes the user's presence, gaze direction, and surroundings to detect anomalies such as unauthorized viewing ("shoulder surfing").

2. **Semantic Reasoning Layer** – A Large Language Model-driven kernel interprets natural language commands and visual cues to make high-level contextual decisions.

Key functionalities include **Dynamic Privacy Shielding**, **Context-Aware Media Control**, and **Autonomous File Management**. Through multimodal sensory integration, Synapse OS demonstrates an advancement towards **Ambient Computing**, enabling computers to act with awareness, proactivity, and autonomy in response to real-world contexts.

Keywords: Multimodal AI, Computer Vision, Large Language Models, Ambient Intelligence, Human-Computer Interaction, Adaptive Operating Systems.

1. Introduction

The rapid evolution of artificial intelligence has led to the emergence of *context-aware computing*—a paradigm where systems adapt behavior based on environmental stimuli and user context. However, current operating systems largely remain reactive, requiring manual control for every operation.

Synapse OS aims to revolutionize this interaction by combining **vision-based perception** and **language-based reasoning**, effectively giving the desktop environment a sense of "situational awareness." By interpreting visual scenes and natural language commands, Synapse OS enables intelligent automation that mimics human-like adaptability.

Such a system bridges the gap between **neuro-symbolic AI** (logical reasoning through language models) and **perceptual AI** (computer vision), achieving a balanced cognitive framework. The project envisions a new computing paradigm—**Ambient Operating Systems**—that proactively interacts with users, improving security, efficiency, and experience.

2. Objectives

The primary objectives of this project are:

1. To design and implement a multimodal OS-level automation framework integrating vision and language inputs.
2. To develop a Visual Perception Layer using OpenCV for:
    - Face detection and recognition
    - Pose estimation and attention monitoring
    - Object and bystander detection for privacy control
3. To develop a Semantic Reasoning Layer using a fine-tuned Large Language Model capable of:
    - Understanding natural language commands
    - Inferring contextual meaning from multimodal data
4. To implement key intelligent behaviors, including:
    - Dynamic Privacy Shielding (automatic screen blur when unauthorized faces are detected)
    - Context-Aware Media Control (pause/resume when user gaze shifts or leaves frame)
    - Autonomous File Organization (categorizing files based on semantic content)
5. To evaluate system performance in terms of accuracy, latency, and user satisfaction.

3. Methodology

The methodology involves the integration of Computer Vision pipelines with Language Model-based reasoning through a modular and event-driven architecture.

3.1 System Architecture Overview

1. **Input Layer:** Webcam video feed (OpenCV capture), voice/text input.
2. **Visual Perception Layer:** Modules for face detection, pose estimation, object recognition (YOLOv8, MediaPipe).
3. **Semantic Reasoning Layer:** LLM kernel (OpenAI, LangChain, HuggingFace) for reasoning and context interpretation.

4. **Decision & Action Layer:** Executes OS-level operations like file organization, media control, privacy enforcement.
5. **Feedback Loop:** Continuous learning and adaptive refinement.

3.2 Workflow

1. Capture visual and verbal input.
2. Extract contextual features (e.g., user absent, bystander detected).
3. LLM kernel infers appropriate action.
4. Execute system command autonomously.
5. Log and analyze for iterative improvement.

4. System Requirements

Hardware Requirements

- Processor: Intel i5 or higher
- RAM: 8 GB minimum (16 GB recommended)
- Storage: 256 GB SSD
- Camera: HD Webcam
- GPU: Optional (for accelerated inference)

Software Requirements

- OS: Windows / Linux (Ubuntu recommended)
- Libraries: Python 3.10+, OpenCV, MediaPipe, Dlib, PyTorch/TensorFlow, LangChain, Transformers
- Tools: VS Code / PyCharm

5. Expected Outcomes

- Prototype desktop environment with adaptive, intelligent, and context-aware features.
- Real-time fusion of vision and language inputs for autonomous operation.
- Enhanced privacy and productivity through perceptive automation.
- Research contribution in multimodal neuro-symbolic AI and ambient computing.

6. Conclusion

Synapse OS represents a pivotal advancement toward intelligent and perceptive computing environments. By merging Computer Vision and Language Reasoning, it transcends conventional user-OS interaction boundaries, creating a responsive and secure digital workspace. The system's multimodal foundation establishes a scalable path toward true **Ambient Intelligence**, where the computer operates not just as a tool but as a cognitive collaborator.