VisionMate-Lite

A Lightweight Assistive Vision System

COMP5523 Computer Vision and Image Processing

Name: HUNG Chi Ming
SID: 24125591G

The Challenge of Visual Impairment

The Problem

- Millions of people are affected by visual impairment
- Worldwide, highlighted in LDC with digital divide

This creates significant challenges in daily navigation and environmental interaction.

Limitations of Current Tech

- Often require expensive, specialized hardware.
- Rely on cloud connectivity, raising privacy concerns and failing offline.
- Complex setup procedures limit practical, real-world deployment.

"What if we could build something simpler that runs on hardware most people already have?"

Our Solution: VisionMate-Lite

Runs on Standard Hardware

Operates on any standard laptop/portable device using only the built-in webcam and speakers.

Offline & Privacy-First

All processing is 100% local. No internet or cloud connectivity is required, ensuring privacy.

Practical & Focused

Provides real-time object detection alerts and on-demand text reading (OCR).

System Architecture

Modular & Robust Design

The system is built with five core, independent components:

- **Camera Interface:** Manages webcam access with error handling.
- **Processing Engines:** YOLOv8n (Detection), Tesseract (OCR), and MobileNetV2 (Scene).
- **Audio Manager:** Centralizes all audio feedback (pyttsx3).
- **Error Handler:** Monitors all modules to ensure robust, continuous operation.

Core Feature: Real-Time Detection

Detection Pipeline

- **Model:** YOLOv8n (Nano variant) chosen for its balance of speed and accuracy.
- **Optimization:** Processes every 3rd frame to reduce CPU load while maintaining responsiveness.
    - **Threshold:** Confidence set at 0.5 to minimize false positives.

Proximity Algorithm

- **Heuristic Approach:** No depth sensor or stereo camera needed.
- **Logic:** An object is "CLOSE" if its bounding box area is > 15% of the total frame area.
- **Alerts:** A 5-second cooldown prevents repetitive announcements of the same object.

Core Feature: Reading & Context

On-Demand OCR (Tesseract)

Activated by the user, a multi-stage preprocessing pipeline improves accuracy:

1. Grayscale Conversion
2. Gaussian Blur (Noise Reduction)
3. CLAHE (Contrast Enhancement)
4. Adaptive Thresholding

Scene Classification

- **Model:** Lightweight MobileNetV2, pre-trained on Places365.
- **Function:** Provides environmental context (e.g., "Entering kitchen area").
- **Optimization:** Runs every 30 seconds to balance performance and awareness.

Performance: Exceeding Targets

The system meets or exceeds all key performance targets on CPU-only hardware, with 14-33% improvements.

Performance: Latency & Reliability

428ms

Average Detection Latency

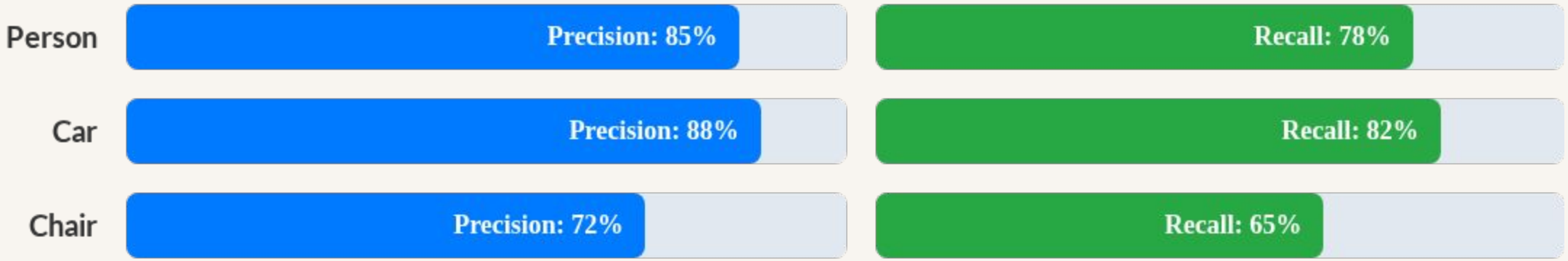
95th Percentile: 612ms. This indicates stable and consistent real-time processing.

100%

Error Recovery Rate

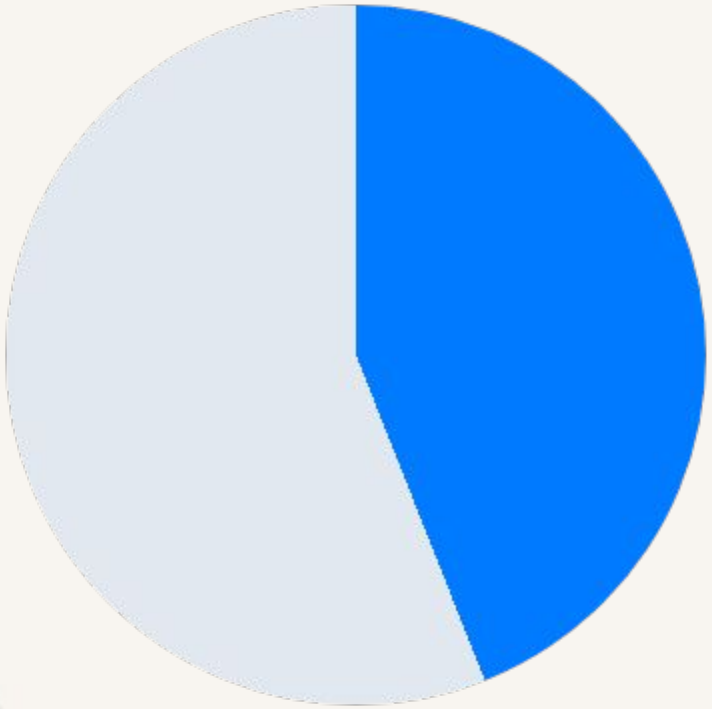
Succeeded in all 15/15 camera and 10/10 TTS failure tests, demonstrating high robustness.

Detection Accuracy (on 45 COCO Images)



Overall Average: ***82% Precision*** *and* ***75% Recall****. Performance is strong for 'Person' and 'Car', and more challenging for 'Chair' due to high variance.*

OCR & Manual Test Results



OCR: 44% Detection

Manual Test Highlights

Person Detection: 100% Success (25/25)

Car Detection: 100% Success (20/20)

Chair Detection: 95% Success (19/20)

Multiple Objects: 93.3% Success (28/30)

Error Recovery: 100% Success (25/25)

Cross-Platform: 100% Success (Win/macOS)

Limitations & Future Work

Current Limitations

- Formal evaluation on only 3 object classes.
- Door/Scene detection tested manually, no formal metrics.
- Proximity detection is a heuristic, not true distance.
- Performance degrades in poor lighting (< 50 lux).

Future Work

- Add more object classes via transfer learning.
- Integrate monocular depth estimation (e.g., MiDaS).
- Mobile deployment (iOS / Android).
- Voice command integration.
- GPS integration for outdoor navigation.

Thank You