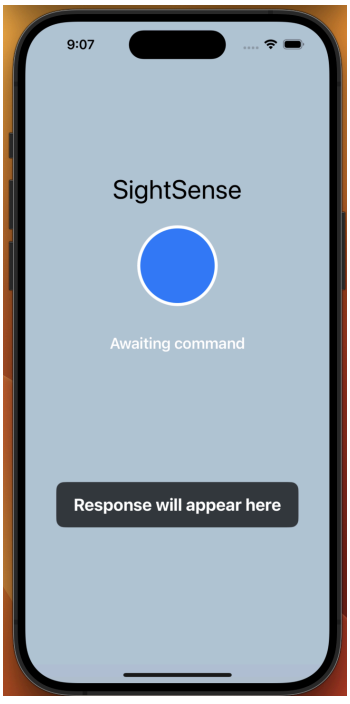
SightSense

An AI-Powered Vision Assistant for the Visually Impaired

DeltaHacks 11 Project Submission

- **Date**: January 12th, 2025
- **Team Members**: Adem, Aiden, Khush, Sammy
- **Host**: McMaster University
- **Link to Devpost:** [SightSense Devpost]



1. Executive Summary

SightSense is an innovative mobile application designed to assist visually impaired individuals in:

- Navigating their environment
- Identifying objects
- Accessing visual information

Using cutting-edge AI technologies, SightSense enhances users' independence and quality of life through its three core features:

1. Object Detection and Location Guidance
2. Scene Description
3. Text Reading

---

2. Project Overview

2.1 Objectives

- Develop a user-friendly, voice-controlled mobile app for visually impaired users.
- Integrate advanced AI models for real-time object detection, scene description, and text reading.
- Create a system for guiding users to locate specific objects in their environment.

2.2 Target Audience

Visually impaired individuals seeking greater independence in their daily activities.

---

3. Core Features

3.1 Object Detection and Location Guidance

Summary:
Helps users locate specific objects in their environment using voice commands. The app provides **audio guidance** to navigate toward the desired object.

User Interaction:

- Users give voice commands to specify the desired object.
- The app provides audio feedback for navigation.

Technical Components:

- YOLOv8 for real-time object detection
- MediaPipe for hand tracking
- Depth Estimation Model for proximity sensing
- Custom Algorithm to convert spatial data to verbal instructions

Challenges Overcome:

- Achieving real-time processing on mobile devices
- Providing accurate 3D spatial hand movement guidance

3.2 Scene Description

Summary:
Provides concise descriptions of the user's surroundings, highlighting key elements and spatial relationships for better navigation and awareness.

User Interaction:

- Users request scene descriptions via voice commands.
- The app delivers audio playback of generated descriptions.

Technical Components:

- OpenAI GPT-4 Vision for image analysis and description generation
- Text-to-Speech Conversion for delivering verbal descriptions

Challenges Overcome:

- Balancing descriptive detail with conciseness
- Prioritizing practical and navigational information over aesthetic details

---

3.3 Text Reading

Summary:
Enables users to access written information by capturing text through the camera and reading it aloud.

User Interaction:

- Users initiate text reading via voice commands.
- Extracted text is read aloud for the user.

Technical Components:

- EasyOCR for text extraction from images
- Natural Language Processing for text formatting and coherence
- Text-to-Speech Conversion for verbal playback of text

Challenges Overcome:

- Managing diverse text layouts and imperfect captures (e.g., blur, skew)
- Structuring text for coherent audio playback

---

4. Technical Implementation

4.1 Backend Development

Framework: FastAPI
Key Features:

- RESTful API endpoints for handling image processing requests
- Integration of multiple AI models
- Efficient data handling and processing

4.2 AI Models Integration

- Object Detection: YOLO (YOLOv8)
- Hand Tracking: MediaPipe
- Natural Language Processing: SentenceTransformer, OpenAI GPT-4 Vision
- Optical Character Recognition: EasyOCR
- Depth Estimation: Hugging Face depth estimation model

4.3 Mobile App Development

Platform: iOS
Language: Swift

Key Features:

- Voice command interface
- Camera integration for real-time image capture
- Text-to-Speech for conveying information to users

4.4 System Architecture

- Mobile app captures images and voice commands
- Backend server processes requests and runs AI models
- Results are sent back to the app and conveyed to the user via speech

---

5. Challenges and Solutions

5.1 Swift Learning Curve

- **Challenge**: Limited experience with Swift
- **Solution**: Intensive self-study and pair programming sessions

5.2 AI Model Integration

- **Challenge**: Efficiently integrating multiple complex AI models
- **Solution**: Optimized backend processing and model selection

5.3 Real-Time Performance

- **Challenge**: Ensuring responsive performance on mobile devices
- **Solution**: Implemented efficient data transfer protocols and local processing where possible

5.4 User Interface Design

- **Challenge**: Creating an intuitive voice-based UI for visually impaired users
- **Solution**: Extensive user testing and iterative design improvements

---

6. Future Enhancements

- Improve object detection accuracy and speed
- Expand language support for global accessibility

- Develop an Android version of the app
- Incorporate user feedback for feature enhancements
- Explore partnerships with organizations for the visually impaired

---

7. Conclusion

SightSense represents a significant advancement in assistive technology for the visually impaired. By harnessing the power of AI and mobile technology, we've created a tool that enhances independence and quality of life.

Through its core features—object detection, scene description, and text reading—SightSense opens up new possibilities for visually impaired individuals to navigate and interact with their environment confidently.

---

8. Contact the Team

We welcome your feedback, questions, and collaboration opportunities. Feel free to reach out to any of our team members:

Adem Cehajic
Role: [e.g., Backend Developer]
Email: ademc1191@gmail.com
Github: https://github.com/Adem-Cehajic
LinkedIn: www.linkedin.com/in/adem-cehajic

Aiden Ly
Role: [e.g., AI Integration Specialist]
Email: atl.aidenly@gmail.com
GitHub: https://github.com/aiden-ly
Linkedin: www.linkedin.com/in/aiden-ly

Khush Patel
Role: [e.g., Mobile App Developer]
Email: khush.patel1@uwaterloo.ca
GitHub:https://github.com/KhushP27
Linkedin:www.linkedin.com/in/khush-patel

Sammy Tourani
Role: [e.g., UI/UX Designer]
Email: sammytourani@gmail.com
GitHub:https://github.com/SammyTourani
Linkedin: www.linkedin.com/in/sammy-tourani

Project Repository: https://github.com/Adem-Cehajic/Deltahacks11

We're excited to hear from you and discuss how SightSense can make a difference!