Vision Assist: Documentation

Neural Nexus: ADIT Calicut

Abstract

VisionAssist is an innovative AI-powered mobile application designed to significantly enhance the independence, safety, and quality of life for visually impaired individuals. The app leverages cutting-edge technologies such as object detection, facial recognition, depth estimation, and natural language processing to provide real-time assistance in various aspects of daily life. Here's a detailed breakdown of the key features and benefits:

Key Features:

1. **Object Identification:**
    - **Functionality:** The app uses the smartphone's camera to capture the user's surroundings and employs advanced object detection algorithms to identify and classify objects in real-time. Each detected object is announced to the user via audio feedback.
    - **Benefit:** This feature helps users understand what objects are around them, enhancing their ability to interact with their environment and increasing their confidence and independence.
2. **Spatial Description:**
    - **Functionality:** The smartphone's camera captures the surroundings. The visual data is then sent to OpenAI's API (or a Llama model) to generate natural language descriptions of the spatial arrangement of the detected objects.
    - **Benefit:** This enhances spatial awareness, allowing users to navigate their environment more safely and effectively by providing detailed information about the distance and location of objects.

3. **Facial Recognition:**
    - **Functionality:** Using facial recognition technology, the app can recognize and announce the presence of familiar individuals. Users can train the app to recognize specific people by uploading images and assigning names.
    - **Benefit:** This feature strengthens social interactions and helps users navigate social settings with confidence, knowing who is around them.
4. **Activity Recognition:**
    - **Functionality:** The app uses machine learning models to recognize and describe common activities and human gestures. By analyzing live video data, it interprets what people around the user are doing and provides this information audibly.
    - **Benefit:** Provides contextual awareness, which is particularly useful in dynamic environments, helping users understand ongoing activities and make informed decisions.
5. **Voice Feedback and Interaction:**
    - **Functionality:** The app features a voice-first interface that allows users to interact through simple voice commands and receive audio feedback. Text-to-speech technology is used to communicate information effectively, ensuring that all instructions and descriptions are accessible.
    - **Benefit:** Ensures that the app is fully usable without visual input, making it accessible and user-friendly for visually impaired individuals.
6. **Text Reading:**
    - **Functionality:** The app uses optical character recognition (OCR) technology to detect and extract text from images, documents, and signs. The extracted text is then converted to speech, allowing the user to listen to the content.

- **Benefit:** Provides access to printed and digital text, allowing users to read documents, signs, menus, and other written materials independently, enhancing their access to information and daily convenience.

7. **Real-Time Currency Recognition:**
    - **Functionality:** The app uses the smartphone's camera to capture images of currency notes and applies currency recognition algorithms to identify the denomination. The app then provides audio feedback to the user, announcing the value of the currency.
    - **Benefit:** Helps users manage financial transactions independently, ensuring they use and receive the correct amounts.

1. Introduction to Problem

Visually impaired individuals face numerous challenges in navigating their environment, identifying objects, recognizing faces, and performing everyday tasks. These challenges limit their independence and quality of life. Traditional aids like canes and guide dogs provide some assistance but have limitations. Modern technology, particularly AI and computer vision, offers the potential to create more sophisticated and versatile tools to help visually impaired individuals navigate their surroundings and perform daily activities more independently.

2. Literature Review

Previous research and development efforts in assistive technology for the visually impaired have focused on various solutions:

1. **Canes and Guide Dogs:** Traditional aids that provide tactile feedback or guide visually impaired individuals.
2. **Electronic Travel Aids (ETAs):** Devices that use ultrasonic waves to detect obstacles and provide feedback.

3. **Mobile Applications:** Apps like Seeing AI and Be My Eyes that offer object recognition and assistance through crowd-sourced help.
4. **AI and Computer Vision:** Recent advancements in machine learning and computer vision have enabled real-time object detection, facial recognition, and depth estimation.

Despite these advancements, there remains a gap in providing a comprehensive, user-friendly, and accessible solution that integrates multiple functionalities into a single mobile application.

3. Proposed Solution

VisionAssist is an AI-powered mobile application designed to enhance the independence and safety of visually impaired individuals. The app leverages advanced technologies such as object detection, facial recognition, depth estimation, and natural language processing to provide real-time assistance in various aspects of daily life. Key features include:

- **Object Identification:** Recognizes and names objects in the user's environment.
- **Spatial Description:** Describes the spatial arrangement and positioning of objects.
- **Facial Recognition:** Identifies familiar people in the user's vicinity.
- **Activity Recognition:** Describes the activities people are engaged in.
- **Voice Feedback and Interaction:** Provides audible descriptions and guidance.
- **Text Reading:** Reads text from various sources aloud.
- **Real-Time Currency Recognition:** Detects and identifies different denominations of currency.

4. Requirements

4.1 Technology Stack

- **Programming Languages:** Python, JavaScript
- **Frameworks:** TensorFlow, PyTorch, React Native
- **APIs:** OpenAI API, TensorFlow.js
- **Libraries:** OpenCV, Tesseract (for OCR)

4.2 Hardware

- **Smartphone:** With a high-resolution camera, sufficient processing power, and internet connectivity.
- **Optional:** Depth sensors for enhanced depth estimation.

4.3 Software

- **Mobile Operating System:** Android or iOS
- **Development Environment:** Android Studio, Xcode, or a cross-platform development tool like Expo for React Native.

4.4 Deployment Environment

- **Mobile Application Store:** Google Play Store for Android and Apple App Store for iOS.
- **Backend Server:** For model updates and potential cloud-based processing.

5. User Requirements

- **Accessibility:** User-friendly interface with large buttons and voice commands.
- **Real-Time Feedback:** Immediate audio feedback for detected objects and spatial descriptions.
- **Customization:** Ability to customize feedback preferences and personal settings.

- **Reliability:** Consistent and accurate object detection and description.
- **Privacy:** Secure handling of personal data and images.

6. Design Documentation

- **User Interface (UI):** Designed with high contrast, large fonts, and intuitive navigation. Includes voice command options for ease of use.
- **System Architecture:**
    - **Client-Side:** Mobile application capturing real-time video, processing frames locally or sending to a backend server for processing.
    - **Server-Side:** Optional backend server for processing heavy computations or storing user data.

7. Implementation Details

- **Object Detection:** Implemented using YOLOv3 model fine-tuned on a diverse dataset of objects.
- **Depth Estimation:** Using MiDaS model for monocular depth estimation.
- **Facial Recognition:** Using a pre-trained model like FaceNet, fine-tuned for recognizing familiar faces.
- **Natural Language Processing:** Using OpenAI's GPT-4 API or a LLaMA model to generate descriptive text based on detected objects and their spatial relationships.
- **Integration:** Combining outputs from detection and depth models to generate comprehensive spatial descriptions.

8. Testing

- **Unit Testing:** For individual components like object detection, depth estimation, and text-to-speech.

- **Integration Testing:** Ensuring seamless interaction between different components.
- **User Testing:** Conducting trials with visually impaired users to gather feedback and make necessary improvements.
- **Performance Testing:** Ensuring real-time performance and low latency in providing feedback.

9. Deployment

- **Build and Package:** Prepare the app for deployment on Android and iOS platforms.
- **Submit to App Stores:** Follow guidelines for submitting apps to Google Play Store and Apple App Store.
- **Monitor and Update:** Continuously monitor app performance and user feedback, providing updates and improvements as needed.

10. Future Scope

- **Navigation Assistance:** Integrate turn-by-turn navigation with obstacle avoidance.
- **Indoor Navigation:** Use Bluetooth beacons or Wi-Fi positioning for indoor navigation.
- **Expanded Language Support:** Provide support for multiple languages for voice feedback.
- **Crowdsourced Hazard Reporting:** Allow users to report hazards and share with other users.
- **Enhanced Depth Estimation:** Incorporate dual-camera smartphones or external sensors for more accurate depth estimation.

11. Conclusion

VisionAssist aims to provide a comprehensive, accessible solution to enhance the independence and safety of visually impaired individuals. By integrating advanced AI and computer vision technologies, the app offers real-time object identification, spatial description, facial recognition, and more. With continuous improvements and user feedback, VisionAssist strives to become an essential tool for visually impaired users, significantly improving their quality of life.