SenseVision Pro: A Real-Time Image Captioning App for the Visually Impaired

Vishnu Nandan
Dept. of Computer Science and Engineering
Christ College of Engineering
Thrissur, India
vishnunandan24cs@cce.edu.in

Nithin N
Dept. of Computer Science and Engineering
Christ College of Engineering
Thrissur, India
nithin005cn@gmail.com

Swathy K S
Dept. of Computer Science and Engineering
Christ College of Engineering
Thrissur, India
swathysubhish@gmail.com

Nimmy Shaji
Dept. of Computer Science and Engineering
Christ College of Engineering
Thrissur, India
nimmyshaji9b@gmail.com

***Abstract*—SenseVision Pro is a mobile application designed to assist blind and visually impaired individuals by providing real-time descriptions of their surroundings. The app uses the phone's camera to capture images, sends them to a local server for processing, and reads aloud AI-generated captions. This paper discusses the architecture, features, and development process of the app, which is intended to improve accessibility for visually impaired individuals by enabling them to better navigate their environment.**

***Index Terms*—Blind Assistive Technology, Image Captioning, Text-to-Speech, Mobile App, Real-Time Processing, AI, Accessibility.**

I. Introduction

SenseVision Pro is designed to bridge the gap between blind individuals and their environment by providing real-time image descriptions. The app captures images using the phone's camera and processes them through an AI-powered system hosted on a local server. This provides users with an audio summary of what the camera sees, enabling them to navigate the world with more confidence.

II. Problem Statement

Visually impaired individuals face significant challenges in understanding their environment. While assistive devices exist, many lack real-time image recognition and audio feedback. Existing solutions either require expensive equipment or fail to provide continuous, actionable information about the user's surroundings.

III. Existing System

Current systems for visually impaired individuals mostly rely on tactile interfaces or audio feedback from pre-recorded information. Some apps use computer vision to identify objects, but they are often unreliable or not optimized for blind users. These systems fail to offer a seamless experience with real-time image processing and continuous feedback.

IV. Proposed System

SenseVision Pro addresses the limitations of existing systems by offering:

- **Real-time image processing:** The app uses the phone's camera to capture images every 5-10 seconds which can be modified based on performance of the model.
- **AI-powered captioning:** Images are sent to a local Flask server running GIT-Base for real-time image captioning.
- **Continuous feedback:** Captions are read aloud to the user through the phone's text-to-speech engine.
- **Seamless user interface:** The app provides feedback on the processing status, letting users know whether the image has been sent, processed, or if an error occurred.

The app also features an auto-capture mode, which continuously captures images and reads captions aloud every few seconds, making it an effective tool for daily use.

V. Conclusion

SenseVision Pro is an early-stage project aimed at improving the quality of life for visually impaired individuals. By leveraging mobile technology and AI, it offers an accessible solution to real-time image captioning. Although the current version implements only a subset of the intended features, it shows great promise in enhancing the daily experiences of blind users. Future work will focus on refining the app's

capabilities, including enhancing the accuracy of image captioning and expanding its feature set.

VI. References

[1] Android Documentation for CameraX: https://developer.android.com/training/camerax

[2] Flask Documentation: https://flask.palletsprojects.com/en/2.0.x/

[3] GIT-Base Documentation: https://github.com/FroggingFamily/colossalai/tree/main/colossalai/modules/gitbase

[4] Kotlin Documentation: https://kotlinlang.org/docs/home.html

[5] Android Text-to-Speech Documentation: https://developer.android.com/reference/android/speech/tts/TextToSpeech

[6] TensorFlow Documentation (for AI-related work): https://www.tensorflow.org/

[7] OpenCV Documentation (if used for image processing): https://docs.opencv.org/