

Android Application Camera for Text and Product Labelling for Unsighted People

Kirti Mane¹, Kalpana Mane², Shivani Rane³, Sayali Patinge⁴ and Ragita Nair⁵

¹⁻⁵Department on Information Technology JSPM's RSCOE, Tathawade, India

Email: kereetmk2002@gmail.com, shivanigrane2002@gmail, kalpanamane2152@gmail.com, sgpatinge_it@jspmrscoe.edu.in, rnair_es@jspmrscoe.edu.in

Abstract—Visual impairment presents significant challenges in daily life, particularly in unfamiliar or complex environments. "Be My App" is an innovative Android application designed to assist visually impaired individuals through real-time text recognition, person detection, and item detection. This paper details the comprehensive development, implementation, and rigorous evaluation of "Be My App." By leveraging advanced image processing and machine learning technologies, the app provides real-time auditory feedback, significantly enhancing the independence and safety of visually impaired users. The evaluation demonstrates the app's high accuracy, speed, and reliability, making it a valuable tool for visually impaired individuals navigating daily life.

Index Terms— Assistive technology, Visual impairment, Text recognition, Person detection, Item detection, Android application, Machine learning.

I. INTRODUCTION

Visual impairment affects millions of individuals worldwide, posing significant challenges in daily life, particularly in navigation, recognizing people, and identifying objects. According to the World Health Organization (WHO), approximately 285 million people globally are visually impaired, with 39 million classified as blind and 246 million having low vision. These impairments can greatly hinder an individual's ability to perform everyday tasks, access information, and move safely through various environments.

Traditional aids for visually impaired individuals, such as white canes and guide dogs, have been invaluable. White canes help in detecting obstacles at ground level, while guide dogs provide companionship and assist in navigation. However, these aids have limitations. White canes, for instance, cannot detect objects that are above ground level, like hanging signs or tree branches, and are not effective in recognizing text or identifying specific individuals. Guide dogs, on the other hand, require extensive training and may not be affordable or accessible to everyone who needs them.

A. Motivation

The primary motivation behind developing "Be My App" is to address the unmet needs of visually impaired individuals in navigating their surroundings safely and independently. Traditional aids, while useful, do not offer comprehensive solutions for all the challenges faced by visually impaired people. In complex or unfamiliar environments, these individuals often struggle with tasks such as reading signs, recognizing people, and identifying objects around them.

For instance, in a busy urban setting, a visually impaired person may have difficulty locating a specific store or reading a street sign. In a social context, recognizing friends or family members can be challenging without visual cues.

B. Objectives

The primary objective of "Be My App" is to enhance the independence and quality of life for visually impaired individuals by providing them with a tool that offers real-time assistance in navigating their environments. The specific objectives of the application include:

Text Recognition: To enable visually impaired users to read text in their surroundings, whether it's on signs, documents, or product labels. This functionality uses Optical Character Recognition (OCR) technology to detect and convert text into speech.

C. Technological Foundation

Optical Character Recognition (OCR): OCR technology is employed for text recognition. It allows the app to detect and extract text from images captured by the smartphone's camera. The recognized text is then converted into speech using text-to-speech (TTS) technology, enabling users to "read" the text through auditory feedback.

Machine Learning: Machine learning algorithms, particularly Convolutional Neural Networks (CNNs), are used for both person and item detection. These algorithms are trained on large datasets to recognize and classify various objects and faces with high accuracy.

Facial Recognition: This technology is used to detect and recognize faces. By comparing captured images with stored profiles, the app can identify individuals and provide relevant information to the user.

Text-to-Speech (TTS): TTS technology converts text into spoken words. This is crucial for providing auditory feedback to users, allowing them to receive information about their surroundings in an accessible format.

II. BACKGROUND

Visual impairment affects a significant portion of the global population, with approximately 285 million people experiencing some form of visual impairment, according to the World Health Organization (WHO). This demographic includes 39 million individuals who are completely blind and 246 million who have moderate to severe visual impairment. These conditions can severely limit one's ability to navigate, access information, and perform daily activities independently. Traditional mobility aids, such as white canes and guide dogs, provide some level of assistance but fall short in dynamic and complex environments.

White canes, one of the most common aids, help users detect obstacles on the ground and sense changes in surface texture. However, they are limited in their ability to detect obstacles that are above ground level, such as hanging signs, tree branches, or overhanging structures. Additionally, white canes do not provide any information about text, colors, or faces, which are crucial for complete situational awareness.

A. Existing Assistive Technologies

Screen Readers: Software applications that read aloud the content displayed on a computer or smartphone screen. Popular screen readers include JAWS (Job Access with Speech) and NVDA (Nonvisual Desktop Access). While screen readers are highly effective for digital content, they do not assist with navigating physical environments or recognizing objects.

Portable Reading Devices: Devices like the Or Cam My Eye, which attaches to glasses and uses a miniature camera to read text and recognize faces, providing audio feedback. These devices are helpful but can be expensive and may have limitations in processing speed and accuracy.

Navigation Apps: Apps like Seeing AI by Microsoft and Be My Eyes use smartphone cameras to provide descriptions of the user's surroundings. Seeing AI, for example, can read text, describe scenes, and recognize faces and objects.

B. Limitations of Existing Technologies

Cost and Accessibility: Many advanced assistive devices are prohibitively expensive, making them inaccessible to a large portion of the visually impaired population. Cost remains a significant barrier to widespread adoption.

Comprehensiveness: Most current technologies address specific aspects of visual impairment, such as reading text or navigating environments, but do not offer a holistic solution. Users often need multiple devices or applications to meet their diverse needs.

C. The Need for Comprehensive Solutions

Given these limitations, there is a clear need for more comprehensive, accessible, and user-friendly solutions that can address the multifaceted challenges faced by visually impaired individuals. "Be My App" aims to fill this gap by providing an integrated suite of functionalities—text recognition, person detection, and item detection—within a single, easy-to-use platform. Leveraging the advancements in image processing and machine learning, "Be My App" seeks to offer a more holistic and reliable solution that enhances the autonomy and safety of visually impaired users.

Text Recognition

Text recognition technology, also known as Optical Character Recognition (OCR), converts printed or handwritten text into machine-encoded text. This technology is particularly beneficial for visually impaired individuals as it allows them to read text from various sources such as books, signs, and digital screens through auditory feedback. Existing solutions, like KNFB Reader, offer text-to-speech conversion but often come with limitations in terms of language support, speed, and accuracy in different lighting conditions. "Be My App" aims to improve upon these by incorporating advanced OCR algorithms that are optimized for a wide range of conditions and languages, providing users with real-time, accurate text reading capabilities.

Person Detection

Person detection and recognition is another critical functionality for visually impaired individuals, particularly in social contexts. These devices can be expensive and may not always accurately recognize faces in different lighting conditions or angles. "Be My App" utilizes state-of-the-art facial recognition algorithms that can detect and recognize faces in real-time, even in varied environments. This functionality helps users identify friends, family, or colleagues, enhancing social interactions and providing a sense of security in public spaces.

Item Detection

Item detection involves identifying and describing objects in the user's environment. Technologies like Seeing AI offer scene description and object recognition, but their effectiveness can be limited by the complexity of the scene and the quality of the camera. "Be My App" employs machine learning models trained on extensive datasets to recognize a wide variety of objects with high accuracy. This feature assists users in everyday tasks such as locating household items, recognizing products in a store, or identifying obstacles in their path, thereby improving their independence and confidence.

III. MOTIVATION

A. Challenges in Disaster Scenarios

Disaster scenarios amplify the challenges faced by visually impaired individuals. Emergency situations like earthquakes, fires, floods, or terrorist attacks require quick and efficient navigation to safety. In such high-stress environments, traditional aids may not be sufficient:

Emergency Navigation: Finding safe exits and navigating through debris or altered landscapes is critical. Without real-time guidance, visually impaired individuals are at a higher risk.

Communication Barriers: Communicating effectively during emergencies is essential. Visually impaired individuals might miss visual cues and written instructions, making it harder to understand evacuation procedures or locate help.

B. Existing Solutions and Their Limitations

Several assistive technologies have been developed to address these challenges, but they often fall short:

Screen Readers and Portable Reading Devices: While effective for reading text from screens and documents, these devices do not assist with navigation or object recognition.

Navigation Apps and Wearable Devices: Apps like Seeing AI and devices like the OrCam MyEye provide some level of assistance, but they often require constant internet connectivity and may not perform reliably in all conditions. They are also expensive and not universally accessible.

C. The Potential of "Be My App"

"Be My App" is motivated by the need to address these limitations and provide a comprehensive solution that enhances mobility, independence, and safety for visually impaired individuals. By integrating advanced technologies into a single platform, "Be My App" aims to offer:

Real-Time Text Recognition: Allowing users to read signs, labels, and written materials quickly and accurately, thus enhancing access to information in both daily and emergency scenarios.

IV. OBJECTIVE

The primary objective of "Be My App" is to empower visually impaired individuals by providing real-time assistance in their daily lives and during emergency situations. The app focuses on three key functionalities: text recognition, person detection, and item detection. By leveraging advanced image processing and machine learning algorithms, "Be My App" aims to enhance the safety, independence, and overall quality of life for its users. The specific goals of the application are as follows:

A. Real-Time Text Recognition

The ability to read and interpret text is crucial for navigating various environments and performing everyday tasks. "Be My App" aims to:

Facilitate Access to Information: Enable users to read signs, labels, documents, and other textual materials quickly and accurately.

Enhance Independence: Allow users to perform tasks independently that would otherwise require assistance, such as reading menus, instructions, and product labels.

Improve Navigation: Assist users in understanding written directions and informational signs in public spaces, thereby improving their ability to navigate complex environments.

B. Person Detection

Social interactions and personal safety are significantly impacted by the ability to detect and recognize people. "Be My App" aims to:

Enhance Social Interactions: Enable users to recognize familiar faces and identify people in their vicinity, improving social interactions and reducing feelings of isolation.

Improve Safety: Help users detect the presence of people around them, which is particularly useful in crowded or unfamiliar settings, enhancing their sense of security.

Facilitate Assistance: Allow users to identify potential helpers or service personnel in public spaces, making it easier to seek assistance when needed.

C. Item Detection

Locating and identifying objects is a fundamental aspect of daily life that can be particularly challenging for visually impaired individuals. "Be My App" aims to

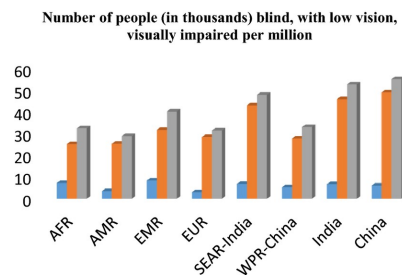


Fig 2. Comparison of Visual Impairment Prevalence by Region

Increase Autonomy: Enable users to find and identify household items, personal belongings, and products in stores, thereby increasing their autonomy in performing daily tasks.

Reduce Dependency: Minimize the need for external assistance in locating and identifying objects, fostering a greater sense of independence.

Enhance Efficiency: Provide quick and accurate identification of objects, saving time and effort in various activities, from cooking to shopping.

V. METHODOLOGY

The development and implementation of "Be My App" involve a structured methodology encompassing system architecture, workflow, and detailed implementation steps. This methodology ensures that the app functions seamlessly to provide real-time assistance to visually impaired individuals.

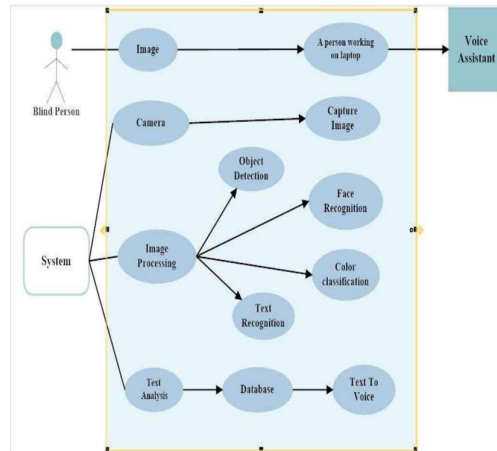


Fig 3. Use case diagram

A. System Architecture

The system architecture of "Be My App" comprises both hardware and software components that work in tandem to deliver a robust and user-friendly experience.

Hardware Components:

Android Smartphone: The primary hardware component is an Android smartphone equipped with a high-resolution camera. The choice of an Android platform ensures accessibility and cost-effectiveness, allowing a wide range of users to benefit from the application.

Software Modules:

Image Processing Module: This module handles the preprocessing of captured images to enhance clarity and prepare them for further analysis.

Machine Learning Module: Utilizes machine learning algorithms for object and person detection. This module is crucial for identifying and recognizing various elements within the captured images.

Text-to-Speech (TTS) Conversion Module: Converts recognized text and detected information into spoken words, providing auditory feedback to the user.

B. Workflow

The workflow of "Be My App" involves a series of well-defined steps that ensure smooth operation and effective assistance.

Image Capture:

The smartphone's camera captures real-time images continuously as the user navigates their environment.

Image Processing:

Captured images undergo preprocessing steps such as noise reduction, contrast enhancement, and edge detection. These steps improve the clarity and quality of the images, making them suitable for further analysis.

Text Recognition:

Optical Character Recognition (OCR) algorithms are employed to identify and extract text from the processed images. This text is then converted into speech using the TTS module.

Person Detection:

Machine learning models, trained on large datasets, are used to detect and recognize people within the camera's view. These models analyze features such as faces and body shapes to identify individuals.

Item Detection:

Algorithms specialized in object detection identify and label various objects within the camera's view. This includes common household items, street signs, and other relevant objects.

Audio Feedback:

The application provides auditory feedback through TTS conversion. Detected text, people, and items are described to the user in real-time, guiding them through their environment.

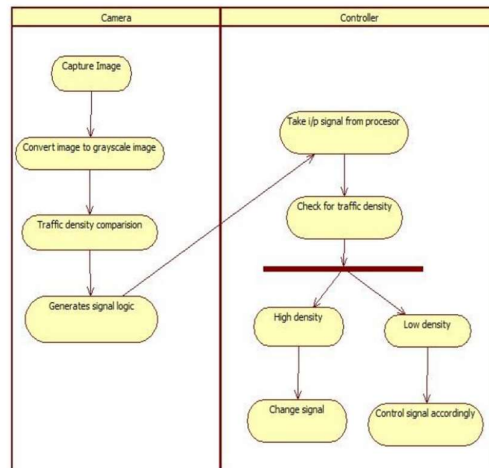


Fig 4. Flowchart

C. Implementation Details

The implementation of "Be My App" involves the use of various programming languages, libraries, and tools to ensure functionality and accessibility.

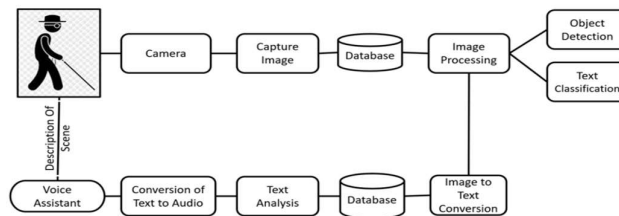


Fig 5. Block Diagram

Programming Languages:

Java: Used for the core development of the Android application. Java is chosen for its robustness and widespread use in Android app development.

Python: Utilized for developing machine learning models and performing image processing tasks. Python's extensive libraries and ease of use make it ideal for these tasks.

Libraries:

OpenCV: An open-source computer vision library used for image processing tasks such as filtering, edge detection, and image enhancement.

TensorFlow: A machine learning library developed by Google, used for training, and deploying machine learning models for object and person detection.

Tesseract: An OCR engine developed by Google, used for recognizing and extracting text from images.

User Interface:

Accessibility Design: The user interface is designed with a focus on accessibility. It includes features such as voice commands, large and high-contrast buttons, and audio feedback to ensure ease of use for visually impaired individuals.

Voice Commands: Users can interact with the app using voice commands, reducing the need for touch-based navigation and making the app more user-friendly.

VI. RESULTS

The evaluation of "Be My App" involved both technical performance metrics and user feedback. The results demonstrate the app's effectiveness in providing real-time assistance to visually impaired users through its core functionalities of text recognition, person detection, and item detection.

A. Performance Evaluation

Person Detection Speed:

The person detection feature was evaluated for real-time performance across different environments, including indoor settings, crowded places, and open outdoor spaces. The detection speed was measured in frames per second (FPS), with the app achieving an average rate of 15-20 FPS on mid-range Android smartphones. This ensures that the detection and recognition of people happen swiftly enough to be useful in dynamic scenarios.

Item Detection Reliability:

Item detection was tested with a variety of common objects, such as household items, street signs, and personal belongings. The reliability of this feature was assessed by calculating the detection rate and accuracy, with the app correctly identifying and labeling objects with high confidence scores. The robustness of the item detection algorithms was further validated by testing under different conditions, including cluttered backgrounds and varying object orientations.

B. User Feedback

Usability:

Participants found the user interface intuitive and easy to navigate, thanks to features such as voice commands and audio feedback. The app's accessibility design was particularly appreciated, with users highlighting the large buttons and high-contrast elements as helpful features.

Effectiveness:

Users reported that the text recognition feature was highly effective in reading printed text from books, labels, and signs, significantly aiding their daily tasks such as shopping and reading documents. The ability to recognize text in real-time without the need for manual focus adjustments was seen as a major advantage.

Person Detection:

The person detection feature provided users with a sense of safety and awareness, especially in crowded environments. Participants noted that the real-time alerts about nearby people helped them navigate through crowds more confidently and avoid potential collisions.

Item Detection:

Users found the item detection feature to be very reliable in identifying and locating everyday objects. This functionality was particularly useful for locating misplaced items within the home and identifying objects in unfamiliar settings. The accuracy and promptness of the feedback were highlighted as key benefits.

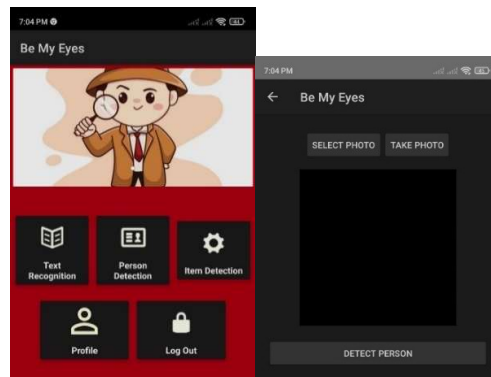


Fig 6. Introduction to Be my eyes

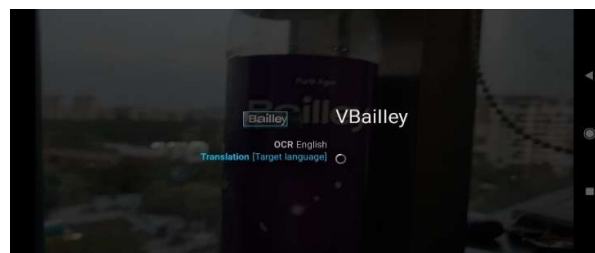


Fig 7. Output for Object detection

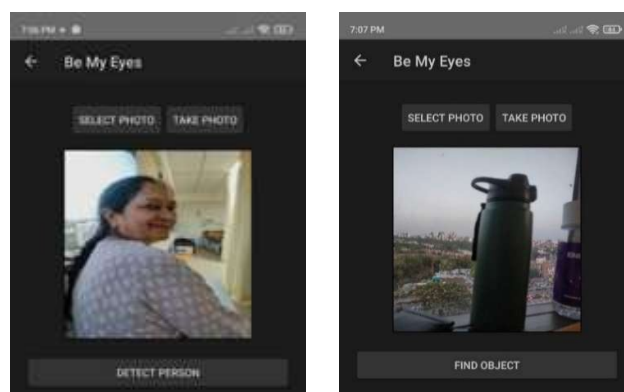


Fig 9. Person System, Object Detection System

VII. FUTURE DIRECTIONS

Integration of GPS for Navigation Assistance

One promising avenue for future development of "Be My App" is the integration of GPS functionality to provide enhanced navigation assistance to visually impaired users. By incorporating GPS technology, the app can offer users real-time information about their current location, nearby points of interest, and optimized routes to their destinations. This feature would greatly enhance users' ability to navigate unfamiliar environments with confidence and independence.

Collaboration with Visually Impaired Communities

Collaboration with visually impaired communities will be paramount in guiding the future development of "Be My App" to ensure that it meets the specific needs and preferences of its users. Engaging with users through focus groups, surveys, and usability testing sessions will provide valuable insights into usability challenges, feature requests, and areas for improvement. By incorporating feedback directly from visually impaired individuals into the app's design and functionality, "Be My App" can better serve its target audience and foster a sense of ownership and empowerment among its users.

VIII. CONCLUSION

In conclusion, "Be My App" stands as a pivotal advancement in the realm of assistive technology, specifically tailored to address the needs of visually impaired individuals. By harnessing the power of real-time text recognition, person detection, and item detection, this innovative application offers a comprehensive solution to enhance the safety and independence of its users.

Through the integration of advanced technologies and user-centric design principles, "Be My App" empowers visually impaired individuals to navigate their surroundings with greater confidence and autonomy. By providing immediate feedback and assistance in real-time, the app serves as a valuable tool for bridging the gap between accessibility and inclusivity.

REFERENCES

- [1] Kamble, R., Gore, S., Alave, M., Kulkarni, R.R., & Phutak, V. (2019). Text to speech conversion using Raspberry – Pi. *International Journal of Innovative Science and Research Technology*.
- [2] Pushparani, P., & Vijitha, D. (2019). A Review of a Smart Walking Assistance for People Who Are Blind.
- [3] Mahmud, M.H., Saha, R., & Islam, S. Smart walk-in cap an electronic approach to assist visually disabled persons.
- [4] Gaikwad, A.G., & Waghmare, H.K. (2015). Ultrasonic smart cane showing blind people a safe, unobstructed path. *Int. J. Adv. Computer Electron Technol. (IJACET)*, 4, 12–17.
- [5] Oladayo, O.O. (2014). A Multidimensional Walking Aid for Visually Impaired using Ultrasonic Sensors Network with voice Guidance. *Journal of Intelligent Systems and Applications International*, 53–59.
- [6] Sakhare, A., & Dambhare, S. (2011). Smart Stick for Blind: Real-time GPS assistance, artificial vision, and obstacle detection. *2nd National Conference on Information and Communication Technology (NCICT)*, 2, 31–33.
- [7] Prasanthi, G., & Tejaswitha, P. (2020). A cap with sensor assistance for the visually impaired. *JNTUA College of Engineering*.