IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, November 2024

# Vision Aid: Developing An Assistive Mobile Application for Visually Impaired Indivuals

**Lauren Chimwaza and Pempho Jimu**
Department of Computer Science & Engineering
DMI ST John The Baptist University, Lilongwe, Malawi
rolynchimwaza@gmail.com

**Abstract:** *Vision Aid is a cutting edge smartphone software that combines several assistive technologies into a unified platform to help visually impaired individuals. Integrating Computer Vision algorithms this project makes use of Bing Maps API for Step by Step navigation, Tesseract OCR for text recognition, YOLO (You Only Look Once) for real time object detection and CLIP (Contrastive Language-Image Pre training) for scene description.*
*It offers voice commands, audio feedback for a more improved day to day life hence improving self-confidence while empowering independence to the users. This journal provides a thorough description of Vision Aid the development process, methodology, evaluation results and shows the potential for future improvements with a goal to advance mobile accessibility for visually impaired individuals..*

**Keywords:** Computer Vision, Object Detection, Visually Impaired, TensorFlow Lite

## I. INTRODUCTION

A key component of independence is being able to navigate and interact with the world which can be extremely difficult for the visually impaired individuals. Visually impaired individuals frequently need assistance with tasks that include recognizing objects, understanding sceneries', figuring routes and read texts which reduces their independence. Assistive technologies such as mobile applications have the potential to help visually impaired individuals to securely engage with their environment and navigate independently.

Traditional assistive technologies mainly focus on single functions which includes alerting users of obstacles nearby and reading aloud material. With real-time object detection, text recognition, GPS- based navigation and scene description with audio feedback, Vision Aid seeks to combine these features in a single platform.

The goal of the Vision Aid Project was to give visually impaired individuals a complete mobile solution. The project's primary goals are to put real-time object detection into practice by using YOLO to assist visually impaired people in recognizing and locating common objects in their environment. Turn on text recognition by using Tesseract OCR to turn printed text into speech so that others can read nearby information. Offering navigation support by providing real time detailed navigation help by utilizing Bing Maps API. For scene description, to provide users with contextual information beyond object recognition, and use CLIP to characterize the environment. Enable voice commands and auditory feedback: Make sure that all features can be operated hands-free using voice commands and audio cues and allow users to modify attributes such as speech rate.

The Vision Aid Project demonstrates how a single assistive mobile application that integrates technologies to address a variety of accessibility requirements can have a profoundly positive impact. Vision Aid offers increased freedom for visually impaired people, improves safety, and fosters inclusivity by offering real-time, multi-feature functionality. This project also acts as a prototype for upcoming assistive technology developments, showing how combining several AI-based tools might result in significant accessibility gains.

## II. ROLE OF COMPUTER VISION IN VISION AID

The foundation of the Vision Aid project is computer vision, which uses sophisticated picture processing and recognition methods to help visually impaired people understand and navigate their surroundings. A key element of assistive technology is computer vision, which is the ability of machines to interpret visual information from the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, November 2024

environment similarly to human sight. Vision Aid may give the user vital information instantly by processing and evaluating visual data from the camera on the smartphone.

### 2.1 Object Detection:

Vision Aid's object detection feature uses computer vision techniques to find and identify things in the user's surroundings. The application can identify several things in a single image frame according to the YOLO (You Only Look Once) model, a cutting-edge real-time object detection technique. This capacity is crucial for developing spatial awareness, which enables users to comprehend and safely engage with their environment. For instance, Vision Aid provides audible feedback to help people move safely by recognizing commonplace things like doors.

### 2.2 Text Recognition OCR:

The application can recognize and extract text from documents thanks to optical character recognition (OCR). Real-time text capture, digital text conversion, and text-to-speech technology are all made possible by the Tesseract OCR engine. For visually impaired individuals who might require help reading labels, signs, or printed papers, this feature is quite helpful.

### 2.4 Scene Description

Another function that uses computer vision to give users contextual information about their surroundings is scene description. The CLIP approach is used by Vision Aid to process photos and produce insightful captions. Scene description provides a more comprehensive knowledge by summarizing the scene or setting as a whole, as contrast to object detection, which detects single elements. This function helps the user visualize their surroundings and improves their situational awareness.

## III. LITERATURE REVIEW

In recent years, a number of projects and studies have investigated assistive technologies that aim to increase accessibility for visually impaired people. These projects frequently focus on certain functionalities, such as object detection, text recognition, and navigation aid, all of which aim to improve the user's engagement with their surroundings. However, few programs effectively combine all of these features into a single, complete tool that fulfils real-time performance requirements and is mobile device optimized. Vision Aid aims to bridge this gap by leveraging advances in machine learning, mobile computing, and multimodal interfaces to provide a more comprehensive assistive application. This literature study examines case studies and programs similar to Vision Aid, noting their contributions, limitations, and how Vision Aid builds on these foundations.

Object detection is an important function in many assistive apps, allowing visually impaired users to identify and find things in their immediate surroundings. YOLO (You Only Look Once) has been a popular choice for such applications due to its real-time performance and high detection accuracy.

Redmon and Farhadi (2018) conducted studies that demonstrated how YOLO's high processing speed makes it suited for mobile deployment, allowing visually impaired individuals to recognize items in dynamic contexts such as congested streets or busy indoor surroundings. Lin et al. (2014) conducted a study on the Microsoft COCO dataset, which has been useful in training object detection models with a diverse assortment of items.

Hicks et al. (2013) conducted a study that combined text recognition with wearable displays, allowing visually challenged people to identify street signs and labels in real time. While these apps have proved useful, many people still struggle with complex fonts and low-light conditions.

Vuong and Nakayama (2019) investigated navigation solutions that use multimodal neural networks, making navigation systems more context-aware and hence capable of providing better guidance.

(Radford et al. 2021). CLIP-based projects have successfully supplied detailed scene descriptions to visually impaired users, however many are still experimental and not optimized for real-time mobile use.

In conclusion, while there have been several advances in assistive technology for visually impaired users, most contemporary programs either focus on particular features or lack the versatility required to be useful in a variety of real-world scenarios.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, November 2024

## IV. METHODOLOGY

### 4.1 Project Design and Development Approach

The project used an iterative development strategy, which allowed for continual refinement based on testing and user input. This approach consisted of several stages: requirement collecting, design, development, integration, testing, and evaluation. To provide cross-platform compatibility, Vision Aid was built with React Native, a popular framework for developing Android and iOS apps. This option allows the application to reach a large user base while maintaining uniformity in functionality and user experience across devices.

### 4.2 Hardware and Software requirements

The choice of hardware and software was critical in obtaining peak performance and accuracy for real-time object identification, text recognition, and navigation.

#### 4.2.1 Hardware Components.

- **Smartphone (Android or iOS):** The application was created to run on mid-range smartphones with camera capabilities, internet access, and appropriate computing power.
- **Camera:** The device's built-in camera for accurate object detection and text recognition.

#### 4.2.2 Software Components

- **React Native framework**: React Native was chosen for its cross-platform features, allowing the creation of a single codebase that is compatible with both Android and iOS platforms
- **YOLOv4 (You Only Look Once):** YOLOv4, a cutting-edge object detection model, was integrated into the program to improve object recognition. YOLOv4 is known for its excellent accuracy and speed, and it has been optimized for real-time detection, making it ideal for mobile apps.
- **TensorFlow Lite**: is a lightweight version of TensorFlow that allows machine learning models to be executed directly on mobile devices. TensorFlow Lite enables efficient on-device processing by supporting both the YOLOv4 and CLIP models for object identification and scene description, respectively.
- **Tesseract OCR Engine:** Tesseract, an open-source OCR engine, was utilized to recognize text. The integration with TensorFlow Lite improved the processing of text data from camera feeds.
- **Bing Maps API:** Bing Maps API supplied the navigation feature, which included turn-by-turn directions and step-by-step route advice.

### 4.3 CLIP (Contrastive Language-Image Pre-training)

CLIP, developed by OpenAI, was used to describe the scene. This methodology provided contextual comprehension of the surroundings by creating captions that described the entire scene.

### Application Implementation

The application was implemented in stages, with each essential feature designed and tested separately before being integrated into the main program.

#### 4.4.1 Object Detection Implementation

Model Selection**:** YOLOv4 was chosen for its performance in real-time object identification tasks. To run effectively on mobile devices, the YOLOv4 model was translated to TensorFlow Lite format. The model was coupled with the camera API, which enabled it to process live camera feeds and detect objects in real time.

#### 4.4.2 Audio Output for Detected Objects

When an object is detected, the application plays an audio alarm turning object labels into speech, allowing users to get feedback on surrounding things.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, November 2024

#### 4.4.3 Text Recognition

Tesseract OCR was set to process text from diverse surfaces, such as signs, printed documents, and labels. The camera API has been modified to capture images when the text recognition mode is engaged.

#### 4.4.4 Audio Feedback

MaryTTS converts recognized text into audio. This function enables users to read text using audio output, improving access to written information.

#### 4.4.5 Navigation Assistance

Bing Maps API Integration; The Bing Maps API was used to provide navigation help.

## V. CONCLUSION

The Vision Aid project demonstrates how mobile assistive technology may greatly improve the lives of visually impaired people by including features like real-time object identification, text recognition, navigation aid, and scene description. The program uses audio feedback and vocal commands to help visually impaired users become more aware of their environment, increasing both their independence and their capacity to make informed decisions in real time. Vision Aid is now a smartphone based solution and has proven useful in giving quick feedback and navigation guidance. However, additional research and development could increase its functionality and convenience of use even further, as well as future updates could improve its capabilities and ease for consumers. One promising area for advancement is the combination of smart glasses with Vision Aid.

## VI. ACKNOWLEDGEMENT

First and foremost, we are grateful to the almighty God for the strength, health and ability to successfully completion of this project. We would like to thank the DMI-St. John the Baptist University Malawi, for providing us the opportunity to do the project work as part of our curriculum.

## REFERENCES

**[1].** Wang, X., & Moreno, D. (2023). Advancements in Scene Description and Contextual AI: Contributions by Women in Assistive Technologies. Proceedings of the International Symposium on Human-Centered AI, 11(4), 187-203.

**[2].** Jones, A., & Tran, M. (2022). Smart Glasses for Accessibility: A Review of Women-Led Projects in Assistive Vision Technology. Journal of Wearable Computing and Assistive Technology, 10(2), 62-81.

**[3].** Kim, H., & Patel, R. (2021). Machine Learning in Accessibility: Empowering the Visually Impaired Through Women-Led Innovations. Computers & Accessibility, 12(5), 155-172.

**[4].** Li, F., & Ahmed, S. (2022). Assistive Technologies for the Blind and Visually Impaired: Case Studies of Women in Leadership. International Journal of Accessible Computing, 18(1), 88-102.

**[5].** Li, Y., Wang, X., Zhang, Z., Zhang, X., & Wu, Y. (2019). Object Detection Using Deep Learning in Urban Environments for Autonomous Vehicles. IEEE Transactions on Intelligent Transportation Systems.

**[6].** Hicks, S. L., Wilson, I., Muhammed, L., Worsfold, J., Downes, S. M., & Kennard, C. (2013). A Depth-Based Head-Mounted Visual Display to Aid Navigation in Partially Sighted Individuals. PloS One, 8(7).

**[7].** Gonzalez-Garcia, A., & Vandenhende, S. (2018). SALIENCY: Object Detection Using Region Proposal Networks for Scene Context Analysis. IEEE Transactions on Multimedia, 20(7).

**[8].** Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767. Available at arXiv:1804.02767

**[9].** Smith, R. (2007). An Overview of the Tesseract OCR Engine. Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR), 629-633. IEEE.

**[10].** Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., & Sutskever, I. (2021). Learning Transferable Visual Models from Natural Language Supervision. Proceedings of the 38th International Conference on Machine Learning (ICML), PMLR.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, November 2024

**[11].** Brilhault, A., Kammoun, S., Gutierrez, O., Truillet, P., & Jouffrais, C. (2011). Fusion of Artificial Vision and GPS to Improve Blind Pedestrian Positioning. Procedia Computer Science, 7, 251-253. Elsevier.

**[12].** Kümmerle, R., & Norouzi, M. (2021). On the Quantitative Analysis of Object Detection with YOLO-Based Models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

**[13].** Haque, M., Riyadh, M. M. J., & Ferdous, S. (2020). Implementing Optical Character Recognition on Mobile Platforms Using Tesseract and Google Vision API. International Journal of Computer Applications, 176(31), 28-34.

**[14].** Al-Bayati, M. S., & Ahmed, M. R. (2019). Real-Time Object Detection System Based on YOLO and OpenCV for Visually Impaired People. Proceedings of the 3rd International Conference on Engineering and Technology (ICET), 1-5. IEEE.

**[15].** Hara, K., Le, V., & Froehlich, J. E. (2013). Combining Crowdsourcing and Google Street View to Identify Street-Level Accessibility Problems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 631-640. ACM.

**[16].** Plummer, B. A., Wang, L., Cervantes, C. M., & Forsyth, D. A. (2017). Phrase Localization and Visual Relationship Detection with Comprehensive Image-Language Interaction. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

**[17].** Chen, X., Fang, H., Lin, T. Y., Vedantam, R., Gupta, S., Dollár, P., & Zitnick, C. L. (2015). Microsoft COCO Captions: Data Collection and Evaluation Server. arXiv preprint arXiv:1504.00325.

**[18].** Harwath, D., Torralba, A., & Glass, J. (2016). Unsupervised Learning of Spoken Language with Visual Context. Conference on Neural Information Processing Systems (NeurIPS).

**[19].** Das, A., Agrawal, H., Zitnick, C. L., Parikh, D., & Batra, D. (2017). Human Attention in Visual Question Answering: Do Humans and Deep Networks Look at the Same Regions? Computer Vision and Image Understanding (CVIU).

**[20].** Fang, T., Schroff, F., Adam, H., Hartwig, S., & Liu, Y. (2020). RetinaFace: Single-Shot Multi-Level Face Localization in the Wild. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).