IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ASSISTING BLIND PEOPLE USING OBJECT DETECTION WITH VOCAL FEEDBACK

[1]Mr. Manjunath N, [2]Ankitha M, [3]Santhrupthi M U, [4]Sneha, [5]Spoorthi HR

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student

[1-2]Department of Computer Science and Engineering,

[1-2]Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India

Abstract: The advancement of mobile technology has opened new avenues for developing innovative solutions to improve the quality of life for individuals with visual impairments. The proposed system presents a mobile application designed to assist blind people in navigating their surroundings by utilizing object detection and providing vocal feedback. The app leverages the power of computer vision algorithms to recognize objects in real-time using the smartphone's camera. Through the integration of state-of-the-art object detection techniques, the app identifies common objects, obstacles, and environmental cues, such as doors, stairs, chairs, or pedestrians. Once the user says describe, the app generates vocal feedback, conveying the object's description. The app's user interface is designed with simplicity and accessibility in mind, featuring intuitive audio prompts and haptic feedback for a seamless user experience. Additionally, the app includes features such as obstacle detection, text-to-speech capabilities, and integration with character recognition, further enhancing its usability and usefulness. The vocal feedback feature provides crucial information about the surrounding environment, enabling users to navigate unfamiliar spaces with ease and avoid potential hazards. the proposed mobile app harnesses the capabilities of object detection and vocal feedback to empower blind individuals in their daily lives. By leveraging the ubiquity of smartphones and the advancements in computer vision, this solution offers a portable and accessible tool for assisting visually impaired individuals, helping them navigate their surroundings and foster a greater sense of autonomy.

Index Terms **- Blind People, Mobile App, Object Detection, Vocal Feedback, Computer Vision, Deep Learning**

I. INTRODUCTION

Visual impairment poses significant challenges for individuals in navigating their surroundings independently and safely. With the proliferation of smartphones and advancements in computer vision, mobile applications have emerged as promising tools to assist blind people in overcoming these challenges. This paper introduces a mobile app that aims to enhance the autonomy and quality of life for visually impaired individuals by leveraging object detection techniques and providing vocal feedback.

The proposed mobile app harnesses the power of computer vision algorithms to detect objects and environmental cues in real-time using the smartphone's camera. By utilizing deep learning models trained on extensive datasets, the app can accurately recognize various objects, obstacles, and contextual elements such as doors, stairs, chairs, and pedestrians. This information is then converted into vocal feedback, delivered through the smartphone's audio output, enabling blind users to perceive and understand their surroundings more effectively.

The primary objective of this app is to provide blind individuals with the essential information they need to navigate their environments independently and with confidence. By leveraging object detection with vocal feedback, the app empowers users to identify and locate objects around them, thus assisting in obstacle avoidance, path planning, and environmental awareness. The real-time processing capabilities of the app ensure prompt feedback, enabling users to make informed decisions and adapt to changing surroundings efficiently.

The user interface of the app is designed with accessibility in mind, ensuring ease of use and customization options to accommodate individual preferences. Intuitive audio prompts and haptic feedback further enhance the user experience, enabling blind users to interact with the app seamlessly. Additionally, the app integrates with existing accessibility tools and features, such as text-to-speech capabilities and navigation assistance, providing a comprehensive solution tailored to the specific needs of visually impaired individuals.

To evaluate the effectiveness and usability of the app, extensive user testing has been conducted, involving blind individuals in real-world scenarios. The feedback received from these tests has been encouraging, with participants expressing increased confidence and independence while using the app. The vocal feedback provided by the app has proven to be invaluable in assisting users in understanding their surroundings, facilitating easier navigation, and mitigating potential hazards.

II. LITERATURE SURVEY

In this paper [1] to achieve the highest level of computational accuracy, this study proposes a potent method for a real-time detector that integrates Deep Learning Neural Networks (DNN). The use of such a framework will ensure the detector's adaptability and dependability by eliminating the sources of distortion. The model depends on integrating the Image Al deep learning framework, the You Only Look Once (YOLO-v3) object detection algorithm, and the DarkNet-53 architecture. The algorithm was trained using the Tensor Flow framework to guarantee trustworthy data processing.

They aimed to give those persons navigation in this paper [2]. It informs the public about the object and gives the object's distance. The programme determines the object's distance on its own. Here, it also offers an audio jack so they can be reminded of the object. Here, we are detecting objects with the SSD technique and measuring their distance with the monodelphi technique. The object will initially be instantly detected by the detecting system through the camera. To process the video, a comparison with the trained models in the cloud database will be made.

This research [3] suggests a system that will identify all conceivable daily things while also prompting a voice to warn users of both nearby and faraway objects. To assess accuracy and performance, two alternative algorithms—Yolo and Yolo_v3—were used in the development of the system in this paper. Yolo uses the Tensorflow_SSD_Mobile Net model while Yolo_v3 uses the Dark Net model. Python module used to translate statements into audio speech to obtain the audio Feedback gTTS (Google Text to Speech). The Python module is used to play the audio in the game. Over 200K photos from the MS-COCO Dataset are used to test both techniques. To assess the algorithm's accuracy in every scenario, both algorithms are examined using webcams in different contexts.

In this study [4], they suggested using computer vision machine learning algorithms to identify objects and use them to help the blind and visually handicapped. This study explains how convolution neural networks are trained on the ImageNet dataset to be able to identify items and describe identified objects to people who are blind. Any device that has a camera can use this implementation, including PCs, tablets, and mobile phones. Different phones are used to test the developed system. The experiment was conducted, and a total of 8 real objects were positioned. Different phones detect objects differently and with varying degrees of precision. iPhone 8, Samsung Galaxy A70, and MI A6 could only accurately identify six items out of a possible six. While mostly phones from different companies detected 5 objects with an accuracy of 62.5 %.

They suggested a low-cost, adaptable technology in this work [5] that may be applied to mobile devices to convert the visual environment to the audio environment, which could aid the blind or visually handicapped in understanding their surroundings. This system requires two components: a sound conversion application that can choose and play the appropriate sounds based on the detected results and a web-based object detection application that uses the SSD-MobileNetV2 object detection algorithm to identify objects such as people, cars, and other things in front of the camera. With the help of a Samsung S8 smartphone, the experiment was carried out. The results of the experiment were examined from a variety of angles, including the detection of a single object belonging to the same category, several objects belonging to the same category, and multiple objects belonging to multiple categories. For up to 80 trained items, the SSD-MobileNetV2 object detection model is offered. However, three types of items (including people, cars, and dogs) were selected for analysis in our initial experiment. We simply need to assess whether the system can provide the sound feedback appropriate for the discovered object because the accuracy and efficacy of sound feedback depend on the accuracy of object detection.

III. PROPOSED TECHNIQUE

The app integrates with the Azure Computer Vision API, which provides a range of pre-trained models and functionalities for object detection and recognition. The API offers methods for detecting objects in images, extracting relevant information, and generating descriptive tags or captions. The app leverages the real-time object detection capabilities of the Azure Computer Vision API to detect and recognize various objects in the user's environment. By utilizing the API's object detection methods, the app can accurately identify objects such as people, walls, doors, furniture, and other obstacles in real-time. The object detection algorithm included in the app utilizes the object localization feature of the Azure Computer Vision API. This functionality enables the app to precisely identify the location of objects in the user's surroundings by drawing bounding boxes around the objects within the image or video stream. Alongside object localization, the app utilizes the object classification capabilities of the Azure Computer Vision API to identify the specific type or category of each detected object. This enables the app to provide more detailed and context-specific vocal feedback to the user. For example, it can differentiate between a person, a car, or a tree. In addition to general object detection, the app incorporates specific functionality for medicine strip detection. By employing custom-trained models or leveraging the API's general object detection capabilities, the app can identify and recognize different types of medicinal strips. This feature allows blind users to easily locate and identify their medication. The app employs the text-to-speech functionality provided by Azure or recorded audio files to generate clear and understandable vocal feedback for the user. This feedback includes descriptive information about the detected objects, their locations, and any other relevant details necessary for the user's navigation and understanding of the environment. The app integrates with various Azure services, such as Azure Cognitive Services, to optimize object detection and vocal feedback generation. By utilizing the power of Azure's cloud-based resources, the app can achieve higher accuracy, real-time processing, and scalability. The app's user interface is designed with accessibility in mind, ensuring ease of use and intuitive interactions. Blind users can easily navigate through the app, customize vocal feedback settings, and access additional features such as navigation assistance or text-to-speech functionality.

- Object Detection: The system should be able to detect and recognize different types of objects in real-time, including people, walls, doors, furniture, and other obstacles. In the object localization stage, the algorithm identifies the location of objects in the image or video by drawing a bounding box around the object. In the object classification stage, the algorithm identifies the type of object in the bounding box, such as a person, car, or tree.
- Medicine Strip Detection: The system should be able to detect and recognize different types of medicinal strips. Medicine strip detection algorithms use a combination of computer vision and deep learning techniques to detect and classify medicine strips in an image or video.
- Vocal Feedback: The system should provide vocal feedback to the user that is clear and understandable. Vocal feedback systems typically use text-to-speech software or recorded audio files to provide spoken feedback to the user. The feedback may include information about the location of objects in the environment, instructions on how to perform a task, or other relevant information.

IV. Conclusion

The development of a mobile app that assists blind people using object detection with vocal feedback represents a significant advancement in leveraging technology to enhance the independence and safety of visually impaired individuals. By integrating the Azure Computer Vision API and implementing key functionalities such as real-time object detection, medicine strip detection, and vocal feedback generation, the app offers valuable support for blind users in navigating their surroundings and understanding their environment.

The app's ability to detect and recognize different types of objects in real-time, including people, obstacles, and environmental cues, empowers blind individuals to confidently navigate their surroundings and avoid potential hazards. The object localization and classification feature further enhance the app's accuracy and provide detailed information about the objects' locations and types. Moreover, the integration of medicine strip detection addresses a specific need for blind users, enabling them to easily identify and locate medicinal strips, contributing to their independence in managing their healthcare needs. The vocal feedback functionality plays a crucial role in the app, delivering clear and understandable information to blind users about their environment. Whether through text-to-speech software or pre-recorded audio files, the vocal feedback provides descriptions, instructions, and other relevant details, facilitating the user's understanding and decision-making process.

The app's integration with Azure services, including the Azure Cognitive Services, ensures optimal performance, real-time processing, and scalability. The cloud-based resources contribute to higher accuracy and overall efficiency, further enhancing the user experience. The user interface of the app prioritizes accessibility, offering intuitive interactions and customization options to suit individual preferences. Through a user-friendly design, blind users can easily navigate the app, adjust vocal feedback settings, and access additional features like navigation assistance or text-to-speech functionality.

V. Future Work

While the mobile app for assisting blind people using object detection with vocal feedback has significant potential, there are several avenues for future work and improvements to enhance its functionality and effectiveness. Some areas of focus include:

- Enhanced Object Recognition: Continued research and development can improve the app's object recognition capabilities by expanding the range of recognized objects. This could include identifying specific brands of products, differentiating between similar objects, or recognizing complex environments such as outdoor landscapes or crowded areas.
- Fine-grained Description: To provide more detailed information to blind users, the app can be enhanced to generate fine-grained descriptions of objects. This could include additional attributes such as color, size, texture, or specific features that provide a more comprehensive understanding of the objects in the user's environment.
- Environmental Context: Integrating environmental context can further enhance the app's usefulness. This could involve incorporating additional sensors, such as GPS or depth sensors, to provide spatial awareness and context-specific information about the user's surroundings. For example, the app could provide guidance about nearby landmarks, points of interest, or potential obstacles in the immediate vicinity.
- Integration with Indoor Navigation Systems: Collaborating with indoor navigation systems can greatly improve the app's utility by assisting blind users in navigating complex indoor environments. By integrating with existing infrastructure, the app could provide turn-by-turn directions, elevator and staircase information, and room identification to enhance indoor navigation and wayfinding.
- Multi-language Support: Expanding the app's vocal feedback capabilities to support multiple languages would enhance its accessibility for users worldwide. This would involve integrating additional language models and text-to-speech functionalities to ensure accurate and natural-sounding vocal feedback in various languages.
- User Feedback and Iterative Improvements: Gathering continuous feedback from blind users and incorporating their suggestions and insights into app updates is crucial for ongoing improvements. Conducting user studies, usability tests, and actively engaging with blind communities can provide valuable insights into the app's strengths and areas for refinement.
- Integration with Wearable Devices: Exploring integration with wearable devices, such as smart glasses or haptic feedback devices, can offer a more seamless and intuitive user experience. Wearable devices could provide real-time audio feedback or haptic cues to further augment the user's understanding of their environment and facilitate hands-free interaction.

VI. ACKNOWLEDGMENT

We would like to express my sincere gratitude to Sri. Vasu Sir, Chairman of VVIET, Mr. Kaveesh, Hon. Secretary of VVIET and all management members of VVIET, who have made my seminar experience possible and full cherished memories for the years to come. We acknowledge and express my sincere thanks to our beloved Principal Dr. Manjunath T.S, VVIET, Mysuru who is source of inspiration to us and who has been to our college a brighter vision in technical education.

We would like to express my deepest sense of gratitude towards Dr. Madhu B K, Professor, Head of the Department, Computer Science & Engineering, VVIET, Mysuru for the valuable suggestions, support, and encouragement. We would also thank our Project coordinator Dr. Archana B, Associate Professor, Department of Computer Science & Engineering for her excellent support in terms of inspiration and resources provided.

We would like to extend my heartfelt gratitude to Prof. Manjunath N, Assistant Professor, Department of Computer Science & Engineering for the valuable guidance and advice. His willingness to motivate me contributed tremendously to our project work. We also would like to thank him for his timely guidance and useful suggestions, which helped us in completing the project work, in time.

We would also like to thank all teaching and non-teaching staffs of the Computer Science Department who has directly or indirectly helped us in completion of the project work. My thanks and appreciations also go to my family and friends who have willingly helped me out with their abilities.

REFERENCES

[1] Tarimo, William, Moustafa M. Sabra, and Shonan Hendre. "Real-Time Deep Learning-Based Object Detection Framework." In 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1829-1836. IEEE, 2020.

[2] Vijlyakumar, K., K. Ajitha, A. Alexia, M. Hemalashmi, and S. Madhumitha. "Object Detection for Visually Impaired People Using SSD Algorithm." In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1-7. IEEE, 2020.

[3] M. Mahendru and S. K. Dubey, "Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 734-740, DOI: 10.1109/Confluence51648.2021.9377064.

[4] J. Nasreen, W. Arif, A. A. Shaikh, Y. Muhammad, and M. Abdullah, "Object Detection and Narrator for Visually Impaired People," 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2019, pp. 1-4, DOI: 10.1109/ICETAS48360.2019.9117405.

[5] H. Nguyen, M. Nguyen, Q. Nguyen, S. Yang and H. Le, "Web-based object detection and sound feedback system for visually impaired people," 2020 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), 2020, pp. 1-6, DOI: 10.1109/MAPR49794.2020.9237770.