Smart Vision Assistant Glasses for Visually Impaired Persons

Thaseen Bhashith[1], Adamya H R[2], Akash S H[3], Dhanyashree B[4], Disha D[5]

[1]Professor,[2]Final year Student,[3]Final year Student, [4]Final year Student, [5]Final year Student
Department of Computer Science and Engineering, Jawaharlal Nehru New College of Engineering
Navule,Shimoga.

---***---

Abstract - In recent years, advancements in computer vision and natural language processing (NLP) have led to the development of highly accessible and assistive technologies. This technology leverages these advancements to create a system that provides real-time object detection and text recognition capabilities, integrated with speech synthesis for audio feedback. The system employs the YOLO (You Only Look Once) algorithm for fast and accurate object detection and an Optical Character Recognition (OCR) module for extracting text from captured images. Text-to-speech (TTS) technology is incorporated to deliver audio outputs, ensuring accessibility for users, especially those with visual impairments. This decentralized system operates on user commands and does not rely on cloud processing, ensuring faster response times and data privacy. By combining computer vision and NLP, this paper offers a cost-effective and portable solution for real-time assistive applications, empowering users to interact effectively with their surroundings through visual data processing and auditory feedback.

Key Words: Computer Vision, YOLOv8, OCR (Optical Character Recognition), Natural Language Processing, Text-to-speech.

1. INTRODUCTION

The goal of this module is to develop a robust text recognition system that can accurately identify and read text from the user's environment, providing audio feedback. This will enable visually impaired individuals to access visual information and improve their independence. In an increasingly visual world, the ability to access and understand textual information is essential for individuals with visual impairments. Traditional assistive technologies often fall short in providing real-time, accurate, and contextually relevant text recognition and audio feedback. This module aims to address this gap by developing a robust and efficient text recognition system that can accurately identify and read text from the user's environment, providing clear and concise audio output. By leveraging advanced computer vision techniques, such as Optical Character Recognition (OCR), and natural language processing, we aim to create a solution that empowers visually impaired individuals to navigate their surroundings with greater independence. The system will be designed to handle printed text, and digital displays. Additionally, it will incorporate intelligent features to adapt to varying lighting conditions, text sizes, and background complexities. The audio feedback component of the system will utilize state-of-the-art text-to-speech (TTS) technology to synthesize natural-sounding speech. Users will have the flexibility to customize the voice, speed, and pitch of the synthesized speech to suit their preferences. Furthermore, the system will be optimized for real-time processing, ensuring minimal latency between text recognition and audio output. The system will also be designed to integrate seamlessly with wearable devices, enhancing portability and ease of use. By incorporating multi-language support, it will cater to a diverse range of users, broadening its accessibility and impact. Additional features like adjustable audio volume and contextual text emphasis will further.

2. RELATED WORK

A paper in IEEE Xplore introduces low-cost smart glasses designed to assist visually impaired individuals in avoiding obstacles. Utilizing ultrasonic sensors, an Arduino Nano, and a buzzer, the system alerts users to potential hazards within a 3-meter range. While currently focused on obstacle detection, it has the potential to evolve into a device capable of image recognition, providing more detailed object information.

Another study in IEEE Xplore discusses advanced smart glasses leveraging Convolutional Neural Networks (CNNs) for real-time object detection and classification. These glasses enable users to identify obstacles, people, and text, integrating image processing techniques such as noise reduction and color correction to enhance recognition accuracy and improve navigation in unfamiliar environments.

Research in IEEE Xplore highlights a portable smart glass prototype equipped with ultrasonic sensors to detect obstacles by measuring distances. A tactile vibration feedback system informs users about obstacles, with varying intensities enhancing the system's utility in real-world scenarios.

A ResearchGate publication presents a low-cost smart glasses prototype aimed at assisting visually impaired individuals in reading hardcopy materials. The system employs a Raspberry Pi 2 for processing, a mounted camera for text capture and enhancement using OCR, and a text-to-speech synthesizer for audio output, with plans for future improvements in text recognition accuracy.

A study in the Indonesian Journal of Electrical Engineering and Computer Science describes smart assistive glasses that integrate artificial intelligence with a Raspberry Pi, a camera, and a sonar sensor for obstacle detection. Features include OCR for text recognition and speech recognition for voice command interaction, enhancing user independence and mobility through safe navigation and seamless access to textual information.

Research in Electronics presents an advanced smart glasses system that uses deep learning and computer vision to assist visually impaired individuals. The system enhances low-light images, detects objects in real-time using advanced encoderdecoder structures, and includes text recognition and tactile feedback, offering comprehensive assistance through an AI-powered platform.

A paper in Sage Journal explores the development of multi-functional glasses for visually impaired individuals, combining cameras, haptic feedback, and audio cues for navigation and obstacle detection. The system employs semantic segmentation for object recognition, alongside localization through GPS and sensors, ensuring accurate and user-friendly navigation with an ergonomic design.

Finally, research in the International Journal of Science and Research Archive showcases a prototype utilizing YOLOv5 for real-time object detection, powered by a Raspberry Pi setup. The system integrates OCR for text recognition and auditory feedback via Python scripts, providing efficient assistance through advanced object detection and classification technologies.

2.1. METHODOLOGY

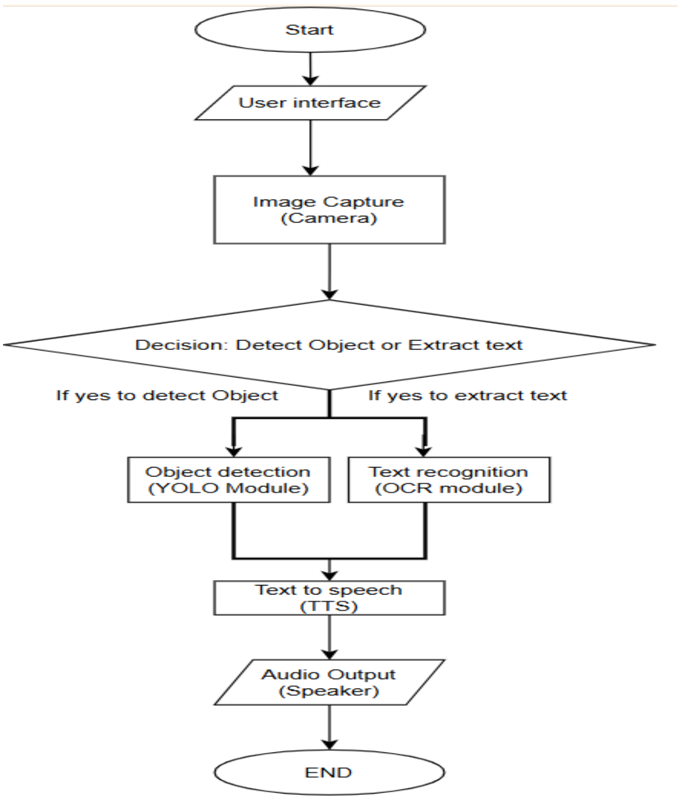

Figure 1. Process Flow of Methodology

The system is designed to assist visually impaired users by integrating various technologies, including object detection, text recognition, and text-to-speech (TTS) capabilities. The process starts when the user interacts with a graphical user interface (GUI), selecting the function they wish to perform, such as detecting objects or extracting text. This interaction triggers the system to capture a real-time image using a built-in camera, which becomes the input for further processing, as illustrated in Figure 1. After the image is captured, the system follows a decision-making process. If the user opts for object detection, the captured image is passed to the YOLO (You Only Look Once) algorithm. YOLO processes the image in real-time, detecting and identifying various objects within the frame. It highlights these objects with bounding boxes and labels them with their respective names. On the other hand, if text recognition is selected, the image is sent to the Optical Character Recognition (OCR) module, which extracts any text from the image, such as signs or documents, and prepares it for further processing. Once either object detection or text extraction is complete, the output is transferred to the Text-to-Speech (TTS) engine. The TTS engine converts the recognized text or identified object names into spoken feedback, which is then played through the system's speakers. This allows users to receive auditory information about the objects or text detected in their environment, facilitating interaction for those with visual impairments.

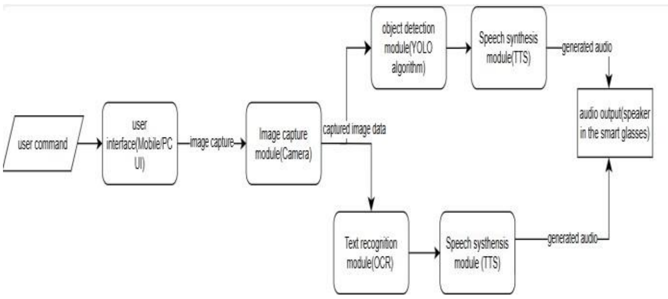

Figure 2. System Architecture

The system architecture for the Smart Vision Assistant Glasses technology integrates multiple components to deliver a seamless assistive experience, particularly for users with visual impairments. It leverages real-time image processing, object detection, text recognition, and speech synthesis to offer an intelligent and user-friendly solution. The components work cohesively to transform user commands into meaningful outputs, as illustrated in Figure 2. System Architecture.

1. User Command: The process begins with a user-initiated command issued through the user interface (UI), which can be accessed on a mobile device or PC. This command specifies the functionality to be activated, such as object detection, text recognition, or both. The interface provides an intuitive way for users to interact with the system.

2. Image Capture Module: The captured input data is handled by the Image Capture Module, using a camera to obtain real-time visual information. This module is crucial for providing raw input to subsequent components such as the object detection and text recognition modules. The camera captures high-resolution images to ensure accurate detection and extraction processes.

3. Object Detection Module (YOLO Algorithm): The object detection module employs the YOLO (You Only Look Once) algorithm, specifically the YOLOv8 model, to analyze the captured images. YOLOv8 is chosen for its efficiency and accuracy in real-time object detection. It processes the input images and identifies various objects within the scene, returning a list of detected items.

4. Text Recognition Module (OCR): In parallel with object detection, the Text Recognition Module utilizes Optical Character Recognition (OCR) technology to extract textual

content from the captured image. The OCR processes the visual data to identify text and convert it into machine-readable format. This feature is particularly useful for recognizing printed or handwritten text in the environment, assisting users in reading signs, documents, or labels.

5. Speech Synthesis Module (TTS): To ensure inclusivity and accessibility, the system integrates a Text-to-Speech (TTS) module. This module converts both the detected objects and extracted text into spoken feedback. Users can adjust the speech rate and volume to match their preferences, ensuring the output is clear and understandable. The TTS module enhances hands-free interaction by providing auditory responses.

6. Audio Output: The generated audio from the TTS module is relayed to the built-in speaker system within the smart glasses. This enables users to receive immediate spoken feedback without requiring external devices, ensuring a self-contained assistive solution.

7. Graphical User Interface (GUI): A dedicated Graphical User Interface provides a visual representation of the processed data. It displays the annotated images with bounding boxes for detected objects, extracted text, and options for customization. Users can select languages for text recognition, adjust settings for speech synthesis, and monitor system outputs through the GUI.

8. Error Handling: The system includes robust error handling mechanisms to ensure smooth operation. This ensures that issues such as camera errors, processing failures, or incomplete outputs are detected and resolved promptly without disrupting the user experience.

3. EXPERIMENTAL RESULTS

The proposed technology aims to present the results obtained from the implementation and testing of the system, highlighting key findings and their significance. The results are based on extensive tests evaluating the system's performance in real-time object detection, text recognition, and text-to-speech functionalities. An image of the developed device is included to provide a visual representation of the system setup, emphasizing its practical application and usability. Figure.3 depicts a user wearing assistive glasses equipped with a mounted webcam, which is connected via a wired setup to a laptop running the project. This configuration enables the system to perform real-time object detection, text recognition, and text-to-speech functionalities. The figure highlights the integration of hardware and software to provide an accessible solution for visually impaired individuals.

Figure 3. User can wear assistive glasses connected via a wired webcam to the software-running in the laptop.

The results of object detection are presented in Figure 4 demonstrating the system's effectiveness in identifying various objects in real time. Using the YOLO algorithm, the system processes input from a live camera feed and detects people and vehicles, essential for navigation and obstacle awareness. In Figures, common objects like bottles, handbags, and cups are identified with bounding boxes and confidence scores, showcasing the versatility of the model in diverse environments. The detection process involves frame-by-frame analysis, applying pre-trained weights for object recognition and annotating the detected objects with labels and scores. This capability ensures reliable performance in both dynamic outdoor scenarios and static indoor settings, enhancing usability for the visually impaired. The outputs exemplify the precision and adaptability of the implemented system.

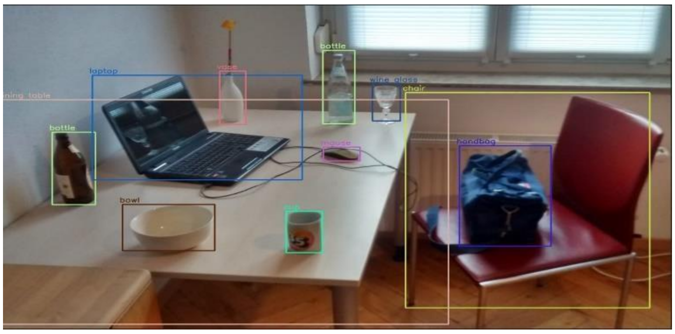

Figure 4. Real-Time Detection of Everyday Objects

.

The system's text detection capabilities are demonstrated in Figure 5, showcasing its versatility in various contexts. Figure 5 highlights the detection of quoted text displayed on a board, effectively capturing words for processing and audio feedback. Meanwhile, Figure illustrates the recognition of text from a page held by a person, indicating the system's practical application for reading printed materials. The implemented OCR algorithms enhance these results by converting detected text into digital format, which is then vocalized using text-to-speech technology. Image pre-processing steps such as sharpening and binarization ensure precision in extracting even small fonts or complex layouts. This feature empowers visually impaired users to access textual content from diverse sources, thereby fostering inclusivity and independence

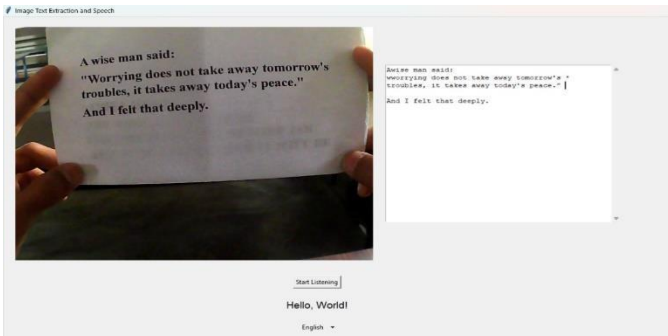

Figure 5. Text Detection from a Page Held by a Person.

4. CONCLUSIONS

The Smart Vision Glasses technology represents a groundbreaking solution for enhancing the daily lives of visually impaired individuals. By integrating real-time object detection and text extraction, it offers a promising tool for aiding navigation, reading, and interaction with the environment. Through advanced technologies such as YOLO for object detection and Tesseract OCR for text recognition, the system provides users with immediate feedback, enabling them to engage more effectively with their surroundings. Additionally, the inclusion of voice assistance further amplifies the system's utility, allowing for hands-free interaction. As the technique progresses, enhancements such as multi-language support, real-time translations, and advanced gesture controls will be incorporated, making the system even more versatile. This innovation not only addresses immediate challenges faced by visually impaired people but also holds the potential for a wider range of applications in accessibility and assistive technologies. The impact of this technology extends beyond just convenience; it opens new doors to independence and confidence for those affected by visual impairments

ACKNOWLEDGEMENT

We would like to thank our institution JNNCE, Computer science and Engineering department for giving us this opportunity to put for our ideas. Also, we would like to extend our heartfelt gratitude towards the faculty who supported us throughout our journey and our parents who were the backbones and helped us complete this technology in success

REFERENCES

[1]. Agarwal, R., Ladha, N., & others. (2017). Low-Cost Ultrasonic Smart Glasses for Blind. IEEE Xplore.

[2]. Ahmed, W., Rawat, T., & others. (2023). Smart Glasses for Blind People Using Obstacles Detection. IEEE Xplore.

[3]. Purushotham, P., & others. (2024). Ultrasonic Sensor for Blind People. IEEE Xplore.

[4]. Hassan, E. A., & Tang, T. B. (2016). Smart Glasses for the Visually Impaired People. ResearchGate.

[5]. Gollagi, S. G., Bamane, K. D., & others. (2023). An Innovative Smart Glass for Blind People Using Artificial Intelligence. Indonesian Journal of Electrical Engineering and Computer Science.

[6]. Mukhiddinov, M., & Cho, J. (2021). Smart Glass System Using Deep Learning for the Blind Visually Impaired. Electronics 2021.

[7]. Bastola, A., Enam, M. A., Bastola, A., Gluck, A., & Brinkley, J. (2023). Multi-Functional Glasses for the Blind and Visually Impaired: Design and Development. Sage Journal.

[8]. Sweatha, R., & Priya, S. S. (2024). YOLOv5 Driven Smart Glasses for Visually Impaired. International Journal of Science and Research Archive.