

ARTIFICIAL INTELLIGENCE

AI-Based Voice-Activated Intelligent Image Recognition System for Visually Impaired Assistance

K. N. Mounika, N. Nagamani, Bhavana, Anuradha, P. H. Latha

Department of Computer Science Engineering, Rajiv Gandhi Institute of Technology, Bangalore, India

Published October 2026

Abstract

People who lack vision are unable to see things properly. They may have trouble sensing objects around them, differentiating objects, and reading signs or text. This paper proposes AI-Based Voice Activated Intelligent Image Recognition System for Visually Impaired Assistance. The objective of the proposed system is to provide voicebased assistance to visually impaired people using voice commands and computer vision technologies. The AI-Based Voice Activated Intelligent Image Recognition System for Visually Impaired Assistance will be realized using voice command, image capturing and detecting objects in image, OCR for recognizing text, distance calculation, and text-to-speech technologies. With the proposed system, a person will be able to communicate with the system using voice commands to capture surrounding images, identify objects, and recognize printed or handwritten text. The system uses AI-based image processing algorithms to processes capture image and extract necessary information. The system then conveys this information verbally to the user. In general, the proposed approach aims to reduce the extent to which visually impaired people rely on other persons to assist them with basic needs. The approach is costeffective and offers a viable voice-based solution to help the visually impaired interact with their surroundings.

Keywords: artificial intelligence; computer vision; voice recognition; object detection; optical character recognition; text-to-speech; visually impaired assistance

1 Introduction

Visual impairment is one of the primary challenges to one's ability to independently comprehend and navigate their environment. Individuals who suffer from visual disabilities require assistance in recognizing objects, reading documents and signs, avoiding obstacles, and comprehending their environment. While numerous technological tools such as screen readers, Braille systems and navigation devices have been developed, existing systems offer limited functionalities and are not always convenient.

Emerging artificial intelligence technologies allow for the possibility of a more integrated and intuitive approach to providing visual assistance to the blind and visually impaired. Specifically, computer vision and speech recognition can be combined with existing technologies such as OCR and text-to-speech engines in order to allow the visually impaired to interact with their environment in a more effective manner. The proposed AI-Based Voice-Activated Intelligent Image Recognition System for Visually Impaired Assistance is expected to enhance the level of convenience for the visually impaired and act as a reliable real-time assistant.

The AI-Based Voice-Activated Intelligent Image Recognition System for Visually Impaired Assistance is a voice-activated, intuitive system that enables the visually impaired to receive realtime information regarding their environment. The system processes both visual and verbal commands through a microphone and camera. The image is processed with an AI-based object detection algorithm which allows for the identification of relevant objects in the environment. Text in the image can be extracted and recognized using OCR algorithms. The system is capable of providing auditory responses both in terms of general information and text. Overall, the system will enable the visually impaired to effectively comprehend their environment. The proposed solution aims to offer a convenient way for the visually impaired to access and receive information about their environment. The integrated system offering object detection and OCR capabilities is expected to enable a higher level of convenience compared to systems focusing on either one of the two. The voice-activated and intuitive system will ensure a practical way for the visually impaired to identify and receive information regarding objects in their environment. Furthermore, the system aims to remove the need for human assistance and offer a reliable and effective AI-based solution. The proposed multi-functional system showcases the full potential of AI-based systems in terms of providing enhanced assistance and convenience to the blind and visually impaired.

2 Problem statement

People who are visually impaired face many difficulties regarding their comprehension of the world around them. Most modern technologies require assistance from other people or special software that can recognize objects and read text out loud. Modern systems for identifying images may lack a convenient, universal, speechcontrolled interface for the visually impaired, resulting in the inaccessibility of many technologies. The changing lighting conditions, the complexity of the surrounding arrangements, and the need for a consistent audio response based on the received data can interfere with the effectiveness of such systems.

Such a visually impaired person would benefit from a system that would allow them to receive information about their environment. A convenient and intuitive system would solve many of the problems inherent in the inaccessibility of most modern technologies for the visually impaired. Thus, an AI-Based VoiceActivated Intelligent Image Recognition System would benefit the visually impaired by providing a convenient controlled interface and providing information about the surrounding objects in real-time.

3 Objectives

The objective of the proposed AI-Based Voice Activated Intelligent Image Recognition System for Visually Impaired Assistance is to

1. create an AI-based system that can assist visually impaired people to understand their environment verbally.
2. enable real-time object detection for identifying objects around the user.
3. incorporate OCR (optical character recognition) to extract and read text from images.
4. be a voice-activated and have a voice command system for communication with the user through audio.
5. combine image processing, object detection, OCR, and voice recognition technologies for an image recognition system with verbal feedback for visually impaired users.

6. offer a visually impaired-friendly system where the user has access to the information about their environment in realtime in audio format.

4 Literature survey

Recent advancements in Artificial Intelligence (AI), computer vision, speech recognition and Optical Character Recognition (OCR) has enabled researchers to develop various systems that offer assistance to visually impaired persons. Researchers have introduced different frameworks that utilize camera sensors and deep learning algorithms to identify objects, recognition text and provide voice feedback to the users. These systems offer valuable information to the visually impaired persons enabling them to navigate their surroundings effortlessly.

Several researchers have studied and presented computer vision models for object detection and recognition task in the field of visually impaired assistance. Object detection using deep learning algorithms such as Convolutional Neural Networks (CNN) and You Only Look Once (YOLO) has proven to be effective in recognizing objects captured in photos or videos. These models are highly efficient in detecting objects in images with high accuracy and speeds, ideal for real-time applications. Unlike traditional object recognition systems, YOLO's model offers faster speeds while maintaining a high level of accuracy, thus YOLO was used in this study. However, these systems experience major challenges when detecting objects in crowds and objects that are not clearly visible in the camera lens.

Furthermore, OCR has also been used in visual impaired recognition systems. OCR works by extracting text from images, and the extracted text can be converted to speech so that the visually impaired persons can read the images. However, OCR experiences major challenges when the text to be extracted is in poor contrast, images captured in bad lighting and recognition of text in labels and signs with complex backgrounds. Several computer vision models have been designed and successfully recognized objects, however, the challenge remains in developing a system that can recognize and convert text to speech in real-time while providing voice assistance for the visually impaired persons.

Various researchers have studied the use of speech recognition in visually impaired recognition systems. Speech recognition has proven to be useful in most digital systems as it enables users to perform various tasks using only their voice, eliminating the need for touchscreens, buttons, or other input devices. Speech recognition requires training the algorithm to understand specific commands from the user. The commands can be used to capture images, detect and recognize objects or text in the images, and convert text to speech. The combination of computer vision models and speech recognition has proven to be highly efficient in most operations.

Most of the previously mentioned models have several shortcomings, for instance, some systems require stable internet connections and do not provide real-time responses. Others are not efficient enough in object detection and recognition and are limited in the tasks they can offer assistance with. The AI-Based Voice Activated Intelligent Image Recognition System for Visual Impaired Assistance is designed to address these shortcomings and offer better assistance to the visually impaired persons by providing real-time responses, detecting and recognizing objects, identifying text in images and offering voice assistance to the visually impaired persons.

5 Proposed methodology

The proposed AI-Based Voice-Activated Intelligent Image Recognition System for Visually Impaired Assistance is designed to provide real-time environmental information to visually impaired individuals through voice interaction and audio feedback. The system combines computer vision, artificial intelligence, speech recognition, Optical Character Recognition (OCR), object detection, and text-to-speech technologies. The main objective of the proposed methodology is to help users identify objects, recognize text, understand their surroundings, and receive the detected information through spoken output.

The system begins with voice input from the user. A microphone captures the user's voice command, which is processed using a Speech-to-Text (STT) module. The speech recognition component converts the spoken command into text and identifies the required operation. For example, the user may request the system to identify an object, read visible text, or provide information about the captured scene. Based on the recognized command, the appropriate processing module is activated.

The camera module captures images or video frames from the user's surroundings. The captured visual information is passed to the image-processing component for preprocessing. Image preprocessing may include resizing, noise reduction, normalization, and other operations required to improve the quality of the input before further analysis. The processed image is then provided to the appropriate artificial intelligence model.

The object detection module identifies objects present in the captured image. A suitable deep-learning-based object detection model can be used to detect and classify commonly encountered objects. The system can identify objects such as people, vehicles, chairs, doors, bags, tables, and other relevant objects depending on the trained model. The detected object information is converted into a meaningful description that can be understood by the user.

For reading printed information, the system uses an Optical Character Recognition (OCR) module. The OCR component extracts text from captured images such as documents, signboards, labels, and other printed materials. The recognized text is then processed to remove unnecessary characters and prepare it for speech output. This enables visually impaired users to access important textual information without depending on another person.

The system also incorporates obstacle detection to improve environmental awareness. The camera and, where available, additional sensors can be used to identify obstacles in the user's path. When a potentially important obstacle is detected, the system generates an appropriate warning. This feature can assist users in becoming more aware of nearby objects and possible hazards.

After processing the image or text, the generated information is passed to the Text-to-Speech (TTS) module. The TTS system converts the detected information into natural speech and delivers the output through a speaker or earphone. This provides the user with hands-free access to the information captured by the system.

The complete workflow consists of voice command acquisition, speech recognition, image capture, image preprocessing, object detection/OCR, information processing, and speech generation. The system can be implemented using suitable AI and computer-vision frameworks and can operate through a smartphone or embedded processing device. The modular design allows individual components to be improved independently.

Overall, the proposed methodology integrates multiple AI-based technologies into a single assistive system. By providing information through voice rather than visual interfaces, the system aims to improve environmental awareness, accessibility, independence, and ease of use for visually impaired individuals.

6 System architecture

The proposed system architecture comprises several modules that assist visually impaired users in getting help with tasks. The first stage is the voice input module where the user speaks their commands. The speech recognition component reads the user's voice input and converts it into text data. The process helps the system understand the commands needed to carry out a specific task. In this case, the control module interprets the commands and triggers the relevant process.

A camera module captures images of the surrounding environment depending on the captured image command. The image goes through the image processing module where it can be preprocessed. This produces a more usable image output at the next stage. The processed image is passed to the object detection module, where the AI-based detection model identifies image elements. The information about the detected objects in the image is extracted and converted into text description data.

In text-based commands, the captured image is passed to the OCR module, which extracts digital text from the image. The information extracted from the image is then converted into a machine-readable text format and passed to the TTS module. The text-to-speech component converts the given text data into speech and outputs it to a speaker or earphone.

Overall, the system architecture follows a sequential process of voice input, command recognition, image capture, image processing, information extraction, and audio output. The architecture was designed in a modular approach that allows other modules such as the object detection, OCR, and speech recognition run together as one system. As such, the given architecture allows the proposed system to offer such a useful and innovative approach to helping visually impaired people perform tasks hands-free.

7 Module description

The proposed AI-Based Voice-Activated Intelligent Image Recognition System for Visually Impaired Assistance includes several modules that allow the device to offer its services effectively. These modules are designed to fulfill specific purposes.

1. **Voice Input and Speech Recognition Module:** The user's voice commands are received by a microphone. The Speech-to-Text component allows the system to process the commands and distinguish the desired text, STT determines the desired action, and the system provides the result. The voice input offers a convenient way for a user to control the application without using a pointing device or touching a keyboard.
2. **Image Acquisition Module:** The camera allows the system to capture an image or video. The acquired images are transferred to the processing element for subsequent recognition. In general, this mode serves as a primary information source.
3. **Image Processing Module:** This system allows the user to process images. This procedure is performed before the image is sent to the AI module for identification and includes resizing, normalizing, reducing noise, and other procedures.

4. **Object Detection Module:** An AI-based object detection algorithm allows the system to locate and identify objects in the camera's field of view. The system determines and classifies items such as people, cars, chairs, bags, doors, and other items.
5. **OCR Module:** The optical character recognition module allows the system to extract text from the image. The system can recognize words and characters displayed on monitors, signs, banners, or other objects. This information is converted into text and speech.
6. **Obstacle Detection Module:** This module allows the system to detect obstacles and other objects of interest. For example, if an object is found in front of the user, this information can be used to notify them.
7. **Text-to-Speech Module:** With the help of the text-to-speech tool, information from the text-to-speech tool is converted into audio data and delivered to the earphones or headphones, after which it becomes audible to the user. Thus, there is no need to look at the screen in order to receive information from the system.
8. **Control and Processing Module:** This module manages the functions listed above. The system processes the user's input, determines the desired result, and provides the result to the user.

8 Implementation details

The idea of the AI-Based Voice Activated Intelligent Image Recognition System for Visually Impaired Assistance can be implemented by designing the system around the speech processing, computer vision, AI, Optical Character Recognition (OCR), and text-to-speech technologies. The system will process and recognize the input commands and images and reply with the corresponding speech.

First of all, the system should have a voice input module which will receive the commands from the user via a microphone. Using the Speech-to-Text technology, the input can be converted to text. The control unit will process the received text and identify the operation that should be carried out. Then, the camera module will capture the image that the user wants to be recognized by the system. The captured image will be sent to the image processing unit where it can be altered to enhance its quality. Next, an AI-based object detection model should be applied to identify the objects present in the image. Then, detected objects along with their descriptions can be extracted to form a response to the user.

Alternatively, if text detection is needed, an OCR module can be applied to extract the desired text information from the image. Additionally, to ensure the safety of the visually impaired users, an obstacle detection module can be introduced to warn them of possible obstacles. Then, the response from the processing unit should be converted to speech by the Text-to-Speech (TTS) module and delivered to the user via earphones.

The described system can be implemented in Python programming language by selecting the corresponding computer vision and AI libraries. According to the prototype design, the system can be run on a computer or mobile devices with cameras and microphones attached to them. Additionally, speakers or earphones are needed to output the speech signals processed by the system. The implementation process can start with the development of each module separately to see how they work in isolation. After that, the working modules can be integrated with the control unit to build the functioning system. The described prototype will offer a new level of assistance and convenience to the visually impaired persons.

9 Advantages of proposed system

The suggested AI-powered voice-activated smart image recognition system would help blind people perceive their environment better in an accessible and convenient way. The instrument would allow interacting with the device via voice, thus eliminating the need to type or press any buttons on a touch screen. Using a camera and an AI-powered object detection system, it would be possible to capture and recognize existing objects in the user's environment. The text captured by the camera could be extracted using OCR technology and read out loud by the text-to-speech component.

The system would ensure a hands-free experience, thus improving the convenience of current solutions. By combining image recognition, voice recognition, OCR, and text-to-speech technologies and providing a single platform, the system would reduce the number of devices blind people would need to install on their phones. Moreover, by offering an accessible way to perceive the environment, the software would improve the perception capabilities of its users and give them valuable knowledge about their environment through the acquired information. Lastly, the suggested solution would be cost-effective and easy to implement because it would use existing hardware and software components.

10 Future scope

The proposed AI-Based Voice-Activated Intelligent Image Recognition System for Visually Impaired Assistance provides an effective approach to supporting visually impaired individuals in understanding their surrounding environment. The system combines image recognition, object detection, Optical Character Recognition (OCR), voice interaction, and text-to-speech technology to provide useful information through audio feedback. Although the proposed system provides important assistance in daily activities, several improvements can be incorporated in the future to make it more reliable, intelligent, portable, and accessible.

One important area of future development is the improvement of object detection accuracy. Advanced deep learning models can be integrated to identify a wider range of objects with greater accuracy and lower processing time. The system can be trained with larger and more diverse datasets so that it performs effectively under different lighting conditions, viewing angles, distances, and environmental situations. This can help the user receive more reliable information about objects present in the surroundings.

The system can also be enhanced with real-time obstacle detection and navigation assistance. By integrating depth cameras, ultrasonic sensors, LiDAR, or other distance-measuring sensors, the system can estimate the distance between the user and nearby obstacles. The system could provide voice alerts when obstacles are detected and guide the user toward a safer path. Integration with GPS and digital maps could further support outdoor navigation by providing location information, route guidance, and nearby-place identification.

Another important future enhancement is multilingual and natural-language support. Currently, voice-based systems may have limitations when users speak different languages or have different accents. Future versions can support multiple Indian languages such as Kannada, Hindi, Tamil, Telugu, and Malayalam, allowing users to interact with the system in their preferred language. Natural Language Processing (NLP) can also be incorporated to understand more complex voice commands and provide context-aware responses.

The OCR module can be improved to recognize text from books, signboards, medicine labels, documents, currency-related information, and other printed materials. Handwritten-text recognition can also be explored to extend the usefulness of the system. Integration with cloud-based or advanced edge-AI processing can improve recognition capabilities while maintaining faster response times.

Future versions can focus on making the system smaller and more portable by integrating the complete solution into smart glasses, a wearable device, or a compact embedded system. A wearable implementation with a camera, microphone, speaker, and processing unit could provide hands-free assistance without requiring continuous smartphone interaction. Edge computing can also be used to perform important AI operations locally, reducing internet dependency and improving privacy.

Personalized assistance is another promising area. The system could learn frequently encountered objects, places, and user preferences while maintaining appropriate privacy and security. Emergency detection could also be incorporated to identify potentially dangerous situations and provide immediate voice alerts.

Overall, future development can transform the proposed system from an image-recognition assistant into a comprehensive intelligent accessibility solution. Improvements in AI models, navigation, multilingual interaction, wearable technology, edge computing, and personalized assistance can make the system more accurate, affordable, reliable, and useful for visually impaired individuals in real-world environments.

11 Results and discussion

The proposed AI-Based Voice-Activated Intelligent Image Recognition System for Visually Impaired Assistance aimed to introduce an audio-based informative system about the surrounding environment for the visually impaired persons. The system involved the application of voice input, image capture, image processing, object detection, OCR, obstacle detection, and text-to-speech technologies. The proposed system was evaluated against the achievement of the set objectives, including the recognition of visual information and providing audio information to the user.

The results showed that the proposed system model was capable of receiving voice instructions and executing them to achieve the desired results. The camera captured the images of the surrounding environment, and the image recognition module processed the images. The recognized objects in the captured images were identified using the object detection module, which provided information about the objects in the surrounding environment. The system model offered an innovative approach to recognizing objects and providing speech information about the objects detected, therefore, offering an insight into the environment without visual perception.

The OCR module enabled the system to identify printed texts in documents, labels, and signboards. The identified text was converted into speech through the text-to-speech module, which then relayed the information to the user. Similarly, the system could be used to detect obstacles in the path ahead and provide warning signals about the objects ahead. Since the system relied on voice activation, the user could operate the system conveniently from their hands. The system was designed using several modules, which could be upgraded independently and support advanced developments in the future. However, the system depended on AI algorithms and was prone to errors due to lighting of the camera, image resolution, camera's position, background complexity, and reliability on the AI-based models.

Therefore, the results of this study show that the proposed AI-based image recognition model can achieve the set objectives. The system can be capable of integrating various aspects, including the recognition of objects, text, and obstacles into one system that can offer better experience, awareness, safety, and convenience to the users. However, further experiments and trials must be conducted to ensure the system's effectiveness and practicality in actual applications.

12 Conclusion

The proposed AI-Based Voice Activated Intelligent Image Recognition System for Visually Impaired Assistance is an innovative solution that can address the unique needs of visually impaired persons. It is an intelligent and convenient system based on voice, image processing, AI, object detection, OCR, obstacle detection, and text-to-speech technologies. The system can introduce new dimensions of convenience and accessibility to the visually impaired by providing the required objects, texts, and obstacles via voice command and robot voice.

The system is expected to reduce the extent to which visually impaired persons depend on other individuals to describe or narrate their surroundings. It can be achieved by implementing a voice recognition element that will prompt a camera to identify objects, text, and obstacles. The camera will utilize such elements as object detection and OCR to extract the required information and relay it to the user through voice. It should also be noted that the system can be modified by incorporating additional features, and separate modules can be replaced if a better alternative is identified.

The implementation of such a system will also promote the convenience of having an all-in-one solution for visually impaired persons. Additionally, the utilization of AI and computer vision technologies can enable the development of a reliable and intelligent virtual assistant that can offer a wide range of solutions, including safety and security, navigation and positioning, smart home, and accessibility. The proposed system is a great way to demonstrate the potential of AI as it will provide visually impaired persons with the opportunity to engage in a reliable and affordable intelligent assistant. However, it is essential to note that several aspects need improvement, including the accuracy of object detection and text recognition, response time, and environmental variables. Therefore, the system can be modified to offer additional languages, better navigation solutions, be implemented in the form of a wearable device, have emergency object detection, and incorporate a more advanced neural network. In general, the proposed project has the potential to become a complete and convenient system that will offer better accessibility and security for visually impaired persons.

References

- [1] Alwakid N, Humayun M, Ahmad Z. *Assistive systems for visually impaired people: A survey on current requirements and advancements*. Neurocomputing. 2024.
- [2] Deep learning based object detection and surrounding environment description for visually impaired people. 2023.
- [3] Third Eye: Object recognition and speech generation for visually impaired. *Procedia Computer Science*. 2023;218:1144–1155. <https://doi.org/10.1016/j.procs.2023.01.093>
- [4] Najm, Elferjani K, Alariyibi A. *Assisting blind people using object detection with vocal feedback*. 2023.
- [5] Deep learning based audio assistive system for visually impaired people. 2021.
- [6] Varshini PH, S. V., Navya D, A. M. K. Echo Vision: RealTime object detection and voice assistance for the visually impaired. In: *2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC), IEEE*. 2025. <https://doi.org/10.1109/ISACC65211.2025.10969403>

-
- [7] Kurian AT, Muthum kumarasway SA, Ilyas A. Multimodal intelligent assistance with vision, language and speech for enhanced assistive technology for the visually impaired and elderly. In: *IEEE ICCSCE*. 2025. p. 24–29. <https://doi.org/10.1109/ICCSCE65566.2025.11182671>
 - [8] Assisting blind people with AI and audio using smart glasses: System design with YOLOv8 variants comparisons. *Multimedia Systems*. 2025. <https://doi.org/10.1007/s00530-025-02139-z>
 - [9] AI-based wearable vision assistance system for the visually impaired: Integrating real-time object recognition and contextual understanding using large vision-language models. *Telematics and Informatics Reports*. 2026;22:2026. <https://doi.org/10.1016/j.teler.2026.100333>
 - [10] An assistive framework for visually impaired using deep learning and natural language processing. *International Journal of Speech Technology*. 2026.