International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 02 | Feb 2026 www.irjet.net p-ISSN: 2395-0072

Case-Based Reasoning for Vision Friend

Sanika Pramod Shetti[1], Janhavi Ganesh Bhise[2], Anuja Dnyandev Suryawanshi[3], Neha Yuvraj Powar[4]

[1]*Student, Computer Science and Engineering, Dr. D. Y. Patil Polytechnic, Kolhapur, India*
[2]*Student, Computer Science and Engineering, Dr. D. Y. Patil Polytechnic, Kolhapur, India*
[3]*Student, Computer Science and Engineering, Dr. D. Y. Patil Polytechnic, Kolhapur, India*
[4]*Student, Computer Science and Engineering, Dr. D. Y. Patil Polytechnic, Kolhapur, India*

Abstract: *In the modern digital era, visually impaired users encounter multiple challenges in performing daily activities independently. Tasks such as navigation, object identification, reading printed text, and recognizing people often require external assistance. Traditional assistive tools like white canes, basic voice assistants, or manual help are limited in intelligence, adaptability, and real-time responsiveness. This project proposes Vision Friend, an intelligent AI-powered mobile assistance system that integrates computer vision, speech recognition, natural language processing (NLP), and text-to-speech technologies to provide real-time, context-aware support. The system processes live camera feeds and voice inputs to detect objects, recognize faces, extract text from images, and deliver immediate audio feedback. Additionally, Vision Friend incorporates adaptive learning to improve accuracy over time by learning from repeated user interactions. The proposed solution reduces dependency on human assistance, enhances user confidence and independence, and offers a scalable, efficient, and practical accessibility solution.*

Keywords: Object Detection, Face Recognition, Speech Recognition, Text OCR, Computer Vision, Mobile Assistance System, AI-Powered Support, Accessibility Tools.

1. INTRODUCTION

The rapid growth of smartphones and AI technologies has opened new possibilities for assistive systems aimed at improving the quality of life for visually impaired individuals. Despite these advancements, many existing solutions fail to provide comprehensive real-time assistance in dynamic environments. Users often struggle with identifying nearby objects, reading signboards or documents, recognizing familiar faces, and navigating unfamiliar surroundings. Traditional assistive systems are mostly audio-based or rely on physical tools that offer limited functionality and adaptability. Such systems lack intelligence to understand context, handle complex environments, or learn from user behavior. To overcome these limitations, intelligent assistive systems that combine multiple AI techniques are essential. Vision Friend is designed as an intelligent mobile-based solution that integrates object detection, face recognition, speech recognition, and OCR into a single application. By processing visual and auditory inputs simultaneously, the system delivers fast, accurate, and personalized assistance. Unrecognized scenarios are stored and refined through adaptive learning, enabling continuous improvement. This approach significantly reduces response time, minimizes external dependency, and improves overall system efficiency and user satisfaction.

2. LITERATURE SURVEY

Several studies have contributed to the development of assistive technologies using AI and computer vision:

Object detection models like YOLO enable real-time detection with high speed and accuracy, making them suitable for mobile applications.

FaceNet and CNN-based face recognition systems provide reliable identification of known individuals, enhancing social interaction for visually impaired users.

Deep learning-based speech recognition systems such as Deep Speech enable hands-free interaction and reduce user effort.

OCR engines like Tesseract allow extraction of text from images, making printed content accessible through audio output.

Research on adaptive learning and similarity measures highlights the importance of systems that improve over time by learning from user interactions.

These studies collectively indicate that a multi-modal AI-based system can significantly enhance accessibility when integrated efficiently into mobile platforms..

3. PROPOSED SYSTEM

The proposed system, Vision Friend, is an intelligent mobile assistance application that provides real-time support to visually impaired users. The system operates through voice commands or camera activation and processes inputs using multiple AI modules.

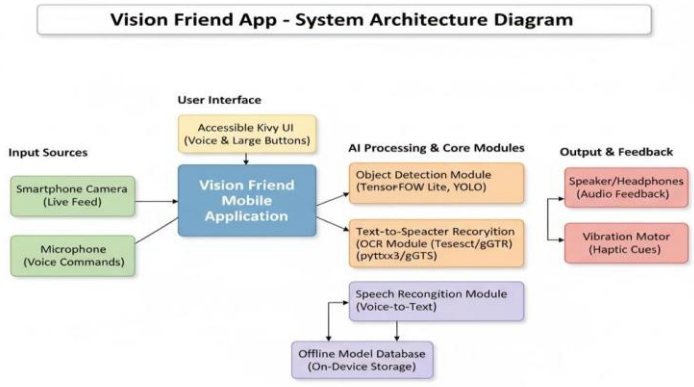
Core Functionalities:

Object Detection: Identifies nearby objects and obstacles to assist in navigation and environment awareness.

Face Recognition: Recognizes known individuals to support social interactions.
Text OCR: Reads printed or handwritten text from images and converts it into speech.
Speech Recognition & NLP: Interprets user voice commands and extracts contextual meaning.
Audio Feedback: Delivers processed information through speech synthesis.
Adaptive Learning: Stores new scenarios and refines models for future interactions.
The system continuously learns from user interactions, ensuring improved accuracy, faster response, and better personalization overtime.

Vision Friend Processing Cycle (Expanded)

1. Detect: The camera captures live video frames, which are analyzed using object and text detection models.
2. Recognize: The system identifies detected objects, faces, or interprets speech inputs for context understanding.
3. Respond: Based on processed data, the system provides immediate audio feedback or alerts.
4. Learn: New or unclear scenarios are stored in the offline database and used to refine models for future use.



System Architecture Diagram – Detailed

Explanation The Vision Friend system architecture consists of the following components:

1.Input Sources :

Smartphone Camera (Live Feed): Captures real-time images and video for object detection, face recognition, and OCR.

Microphone (Voice Commands): Accepts user voice inputs for commands and interaction.

2. User Interface:

Accessible UI: Designed with large buttons, voice-based navigation, and minimal visual dependency to ensure ease of use for visually impaired users

3. Vision Friend Mobile Application (Core Layer)

This is the central control unit that coordinates all processing modules and manages data flow between inputs and outputs.

4. AI Processing & Core Modules

Object Detection Module (YOLO): Detects objects in real-time and identifies obstacles or important items.

Face Recognition Module (CNN/FaceNet): Matches detected faces with stored profiles to identify known people.

Text-to-Speech & OCR Module (Tesseract OCR): Extracts text from images and converts it into audible speech.

Speech Recognition Module: Converts voice commands into text and passes them to NLP for interpretation.

5. Offline Model Database

Stores learned scenarios, recognized patterns, and user-specific data.

Enables functionality even with limited or no internet connectivity.

6. Output & Feedback Layer

Speakers/Headphones: Provide clear audio feedback and instructions.

Vibration Motor (Haptic Feedback): Offers tactile alerts for critical notifications such as obstacles.

This modular architecture ensures scalability, reliability, and efficient performance in real-world environments.

4. COMPARATIVE ANALYSIS

Parameter	Existing System	Proposed System
System Type	Traditional / Audio-only aid	Intelligent AI-based mobile assistance
Query Handling	Manual or basic voice	Speech Recognition and NLP-based processing
Response Time	Slow to moderate	Fast and real-time
Similar Scenario Handling	Limited	Efficient detection and recognition
Learning Capability	No self-learning	Self-learning using adaptive models
Effort Required	High (external help)	Low
Scalability	Limited	High

5. ADVANTAGES

- Provides faster response by detecting and recognizing elements in real-time.
- Reduces dependency on human assistance or physical aids.
- Ensures consistent and reliable feedback for similar scenarios.
- Handles large varieties of environmental inputs efficiently.
- Improves overall user independence and quality of life.

6. CONCLUSION

Vision Friend is an intelligent AI-powered mobile assistance system designed to address the limitations of traditional assistive tools for visually impaired users. By integrating object detection, face recognition, speech recognition, OCR, and adaptive learning, the system delivers accurate, fast, and context-aware assistance. The detect–recognize–respond–learn cycle ensures continuous improvement and personalization. Overall, Vision Friend reduces user effort, enhances independence, and improves quality of life. Its scalable and self-learning architecture makes it suitable for modern mobile environments and provides a strong foundation for future advancements in accessibility-focused AI solutions.

7. REFERENCES

[1] YOLO: Real-Time Object Detection — Joseph Redmon, 2016.

[2] FaceNet: A Unified Embedding for Face Recognition — Florian Schroff et al., 2015.

[3] Deep Speech: Scaling up End-to-End Speech Recognition — Awni Hannun et al., 2014.

[4] Tesseract OCR Engine — Ray Smith, 2007.

[5] Coarse-to-Fine Vision Systems — Jing Lu et al., 2015.

[6] Computer Vision: Algorithms and Applications — Richard Szeliski, 2010.

[7] S. K. Pal and S. C. K. Shiu, “Soft Computing,” 2004.

[8] J. Lu et al., “Knowledge-Based Systems,” 2015.

[9] I. Watson, “Applying AI Techniques,” 1999.