SMART GUIDE SPECS FOR BLINDS

Asst Prof. Jyotsna E
Dept. of Computer Science and Engineering,
Adi Shankara Institute of Engineering and Technology,
Ernakulam, India.
jyotsnae@gmail.com

Pranav Sekhar P
Dept. of Computer Science and Engineering,
Adi Shankara Institute of Engineering and Technology,
Ernakulam, India
pranavsekhar103@gmail.com

Yedukrishnan A M
Dept. of Computer Science and Engineering,
Adi Shankara Institute of Engineering and Technology,
Ernakulam, India.
yedukrishnan88@gmail.com

Shahal Moideen
Dept. of Computer Science and Engineering,
Adi Shankara Institute of Engineering and Technology,
Ernakulam, India
Shahalshanu314@gmail.com

Jeevan K Joy
Dept. of Computer Science and Engineering,
Adi Shankara Institute of Engineering and Technology,
Ernakulam, India
jeevankjoy0@gmail.com

***Abstract*- Blind individuals face significant challenges in understanding their surroundings due to the inability to visually perceive dynamic actions of animals, people, and moving objects. Existing assistive technologies often fall short in providing timely and comprehensive information about live actions that could impact their safety and independence. Therefore, there is a critical need to develop a system that utilizes advanced sensor technologies, computer vision algorithms, and artificial intelligence to detect, analyze, and interpret live actions in the environment. This system should provide intuitive auditory or tactile feedback to users, enabling them to make informed decisions and navigate their surroundings more confidently and independently.**

INTRODUCTION

Smart specs for blind individuals are advanced wearable devices designed to improve their ability to navigate and interact with their surroundings. These glasses are equipped with a combination of sensors, including cameras, accelerometers, and gyroscopes, that work together to detect both actions and objects in the environment. The camera captures visual data, while motion sensors detect physical movements such as walking, waving, or jumping. The system processes this data using computer vision and machine learning algorithms to identify and classify the objects and actions in real time.

Once an action or object is detected, the device provides the user with auditory feedback through a small speaker or bone conduction technology, which delivers clear, non-intrusive voice output. For example, if the user is walking towards an obstacle, the system will announce "obstacle ahead" or "table to your left," helping the user avoid collisions. Similarly, the system can recognize specific objects such as a door, chair, or personal items like a watch or phone, providing detailed descriptions of their locations and proximity.

This continuous voice feedback enables blind individuals to perform everyday tasks more independently, such as moving through a room, finding objects, or recognizing gestures. By providing situational awareness and promoting greater autonomy, these smart specs have the potential to significantly improve the quality of life for people with visual impairments, offering them a more seamless and efficient way to interact with the world around them. Furthermore, the smart specs are designed to adapt and improve over time, using AI and machine learning algorithms that learn from the user’s environment and specific needs. This personalization allows the device to provide more accurate and relevant information, catering to individual preferences and enhancing the user experience.

LITERATURE SURVEY

A literature survey on smart specs for visually impaired individuals reveals significant advancements in assistive technology, with research focusing on enhancing real-time environmental awareness, object detection. Studies have explored various sensor and camera technologies, along with machine learning algorithms, to detect and classify actions and objects.

Real-Time Answers to Visual Questions

Bigham et al.– VizWiz [1] This research aimed to create a real-time visual assistance system for blind individuals. The system allowed users to take pictures and ask questions about them, receiving responses from remote human workers. Steps

1. User Interaction: A visually impaired person captures an image using a smartphone camera and records a spoken question related to the image.

2. Cloud-Based Processing: The image and question are sent to a remote server.

3. Human Assistance: Crowdworkers on platforms like Amazon Mechanical Turk analyze the image and provide verbal responses.

4. Response Delivery: The system converts textual responses into speech output for the user.

Smart Cane for the Visually Impaired

Kumar et al. [2] To design a smart cane that enhances mobility for visually impaired users by detecting obstacles and providing feedback.Methodology:

1. Hardware:Equipped with ultrasonic sensors to detect obstacles.Integrated with a vibration motor to alert users.

2. Software:Sensors continuously scan the environment for obstacles within a 2-meter range.Distance data is processed to determine obstacle proximity.Feedback is given via vibrations of varying intensity.

Results:Detected obstacles with 92% accuracy.Users adapted to the vibration-based feedback within 2 weeks of usage.

Wearable Navigation System for the Blind

Mandal et al. [3] To develop a wearable navigation system using LiDAR and GPS to guide visually impaired users.Methodology:

1. LiDAR Integration:Captures a 360-degree view of the user's surroundings.Maps obstacles in real-time.

2. GPS Tracking:Provides location-based navigation.Uses pre-recorded audio cues to assist movement.

3. Voice Feedback:Converts LiDAR and GPS data into audio commands.Example: "Turn left in 3 meters."

Results:Achieved 87% accuracy in indoor environments.Effective in outdoor navigation with 95% accuracy.

Object Detection for Visually Impaired Individuals - YOLO (You Only Look Once)

Redmon et al. [4] To develop a real-time object detection algorithm capable of identifying multiple objects in a single pass through an image.Methodology:

1. Neural Network Architecture:

Uses a single convolutional neural network (CNN) to predict object bounding boxes and class probabilities simultaneously.Splits images into an SxS grid and predicts bounding boxes for each cell.

2. Dataset Training:Trained on COCO and Pascal VOC datasets.Uses a custom loss function to balance classification and localization errors.

3. Real-Time Processing:Achieves 45 frames per second (FPS) on GPU-based systems.

Mask R-CNN an advanced object detection model

He et al. [5] He et al proposed Mask R-CNN, an advanced object detection model that includes instance segmentation to improve accuracy. The model processes an image by utilizing a Region Proposal Network (RPN) to identify possible object locations. A convolutional network then extracts feature maps and classifies objects while simultaneously generating segmented object boundaries. This results in precise object detection with detailed segmentation masks. The final output includes labeled bounding boxes and segmented regions, providing visually impaired users with a richer understanding of their environment through voice-based feedback. Additionally, Mask R-CNN can be adapted to integrate depth estimation, which allows for an improved perception of object distances, aiding in obstacle avoidance and safe navigation.

SSD (Single Shot Multibox Detector)object detection model

Liu et al [6] developed the Single Shot Multibox Detector (SSD), an object detection model that balances speed and accuracy. The SSD processes an image using multiple convolutional layers that extract hierarchical features. Multi-scale feature maps allow for detecting objects of different sizes. Anchor boxes are applied at each layer to predict object locations and labels, refining the final detections through non-maximum suppression. This model is efficient for real-time object detection, enabling quick feedback to visually impaired individuals through auditory alerts. The SSD framework can also incorporate edge computing for faster processing on wearable devices, reducing reliance on cloud-based systems and ensuring uninterrupted assistance in offline environments.

Action Detection for the Visually Impaired Two-Stream CNN

Simonyan & Zisserman [7] introduced the two-stream CNN model for action recognition, which processes video data through separate spatial and temporal streams. The spatial stream extracts object-related features from individual frames, while the temporal stream captures motion by analyzing optical flow. The outputs from both streams are combined to classify actions accurately. This model is highly relevant for visually impaired individuals, as it enables smart assistive devices to recognize human activities and provide contextual feedback. For example, the system can alert users when pedestrians are moving towards them or when approaching a busy crosswalk. Future adaptations of the two-stream CNN involve integrating transformer-based architectures for improved temporal sequence understanding and multimodal fusion for audio-visual event detection.

I3D(Inflated 3D Convolutional Neural Networks) model

Carreira & Zisserman [8] developed I3D, a deep learning model that extends traditional CNNs by incorporating 3D convolutions to capture temporal dependencies in video sequences. The model processes consecutive frames to analyze movement patterns and classify actions with high accuracy. For assistive technologies, I3D can be deployed in smart glasses or wearable cameras, providing real-time alerts regarding surrounding activities. This enables visually impaired individuals to understand ongoing events, such as approaching vehicles, people waving, or nearby hazards. Additionally, I3D can be enhanced with self-supervised learning techniques to adapt to different environments without requiring extensive labeled datasets.

Real-Time Crowdsourced Assistance

Bigham et al.[9] introduced VizWiz, an innovative mobile-based application that provides real-time assistance to visually impaired individuals by leveraging crowdsourced human intelligence and machine learning. The system functions by allowing users to capture an image of their surroundings and record a spoken question related to the captured scene. The query is then transmitted to a cloud-based server, where it is routed to multiple human workers via Amazon Mechanical Turk or dedicated volunteers. These workers analyze the image and respond with textual or audio-based answers, which are relayed back to the user through speech output. The system improves response accuracy by aggregating multiple answers and filtering inconsistencies. Over time, VizWiz also integrates AI-based image recognition to supplement human responses, reducing dependency on manual intervention. The application provides a critical bridge between blind users and their environment by offering real-time assistance in activities such as reading labels, recognizing objects, and navigating unfamiliar spaces. Moreover, enhancements such as AI-based question prioritization and integration with augmented reality devices have further improved VizWiz's efficiency and reliability. The future scope of VizWiz includes integrating deep learning-based OCR (Optical Character Recognition) for text extraction from images and automatic speech recognition for improved user interaction.

AI-Powered Wearable Device for Scene Understanding

Chen et al. [10] developed an AI-powered wearable device designed to provide real-time scene understanding for visually impaired individuals. The system integrates a combination of RGB and depth cameras, LiDAR sensors, and an advanced deep learning model to analyze the surrounding environment and provide meaningful auditory feedback. The core of the system utilizes a transformer-based vision model to extract relevant information from the user's surroundings, including object detection, action recognition, and obstacle avoidance. Unlike previous models, this system prioritizes contextual awareness, allowing users to receive dynamic scene descriptions rather than simple object labels. The model is optimized to run on edge devices, ensuring low latency and real-time processing without reliance on cloud computing. The device also includes a multimodal interaction system, where users can ask voice-based questions about their surroundings, and the AI responds with contextualized information. Additionally, haptic feedback is incorporated to alert users about immediate obstacles, improving overall safety. The study conducted real-world testing across different environments, including urban streets, indoor settings, and crowded spaces, demonstrating high accuracy in object and action recognition.

AI-Powered Voice Assistance for Navigation

Jones et al. [11] developed a voice-assisted navigation system that integrates artificial intelligence (AI) and GPS technology to aid visually impaired individuals in navigating complex urban environments. The system utilizes a combination of AI-driven natural language processing (NLP) and geospatial data to provide real-time turn-by-turn voice instructions. Additionally, the navigation system incorporates machine learning algorithms to analyze user movement patterns, predict preferred routes, and optimize navigation suggestions. One of the key features of this system is its ability to detect obstacles and provide audio warnings, enabling users to avoid potential hazards such as uneven sidewalks, street curbs, and traffic intersections. The system is also equipped with real-time traffic analysis, alerting users to ongoing road construction, pedestrian crossings, and congestion points. Future improvements include enhanced AI-based predictive path planning, integration with public transportation data for multimodal navigation, and personalized voice assistance adapting to user-specific needs.

Deep Learning for Gesture-Based Interaction

Wang et al. [12] introduced a deep learning-based gesture recognition system designed for smart glasses, enabling visually impaired individuals to interact with digital interfaces using hand movements. The system utilizes convolutional neural networks (CNNs) trained on large datasets of hand gestures to recognize predefined movements with high accuracy. Gesture-based interaction provides an intuitive and efficient way for users to control various functions of assistive devices, such as activating object recognition, adjusting volume, and receiving navigation assistance. The deep learning model also adapts to individual user patterns through continuous learning, enhancing recognition accuracy over time. The smart glasses are equipped with an integrated depth sensor and infrared camera, allowing for real-time gesture tracking in various lighting conditions. Future developments aim to improve gesture recognition speed, expand the gesture vocabulary, and integrate voice control as a complementary feature for seamless interaction.

Edge AI for Real-Time Object Recognition

Lee et al. [13] implemented Edge AI technology in smart glasses to enable real-time object recognition without reliance on cloud computing. The system processes visual data locally on an embedded processor, significantly reducing latency and enhancing response time. The AI model deployed on the edge device is optimized for low-power consumption, ensuring efficient operation on wearable assistive devices. The smart glasses utilize a combination of deep learning techniques, including MobileNet and EfficientNet, to identify objects in the user’s immediate surroundings and provide auditory feedback. The device also supports dynamic learning, allowing users to train the system to recognize new objects based on personalized needs. Additional advancements include integrating haptic feedback for tactile-based notifications and developing a hybrid cloud-edge model to balance processing speed and storage efficiency.

Multi-Sensor Fusion for Enhanced Perception

Patel et al. [14] developed a multi-sensor fusion system that integrates RGB cameras, LiDAR, and ultrasonic sensors to enhance environmental perception for visually impaired users. The system leverages AI-driven sensor fusion techniques to merge data from different sources, improving detection accuracy and minimizing false positives. The RGB cameras capture visual data, LiDAR provides depth perception, and ultrasonic sensors detect nearby objects, creating a comprehensive real-time environmental map. The AI model processes these inputs to generate an accurate representation of the surroundings, which is then converted into auditory or haptic feedback for the user. The device is designed to operate in both indoor and outdoor settings, making it versatile for navigation in various environments. Future improvements include incorporating thermal imaging sensors for detecting heat signatures in low-visibility conditions and AI-based predictive modeling to anticipate obstacles based on movement trajectories.

AI-Powered Scene Interpretation

Gonzalez et al [15] proposed an AI-powered scene interpretation system that generates detailed contextual descriptions of the user's surroundings using computer vision and natural language processing (NLP). The system utilizes a transformer-based deep learning model, such as GPT-4 Vision, to analyze images captured by a wearable camera and convert them into descriptive text. The descriptions are then relayed to the user via speech synthesis, providing an enhanced understanding of the environment. The AI model is trained to recognize complex scene compositions, including object relationships, spatial orientations, and human activities. The system also features interactive capabilities, allowing users to ask specific questions about their surroundings and receive precise answers. Future advancements include integrating multi-language support, improving response speed, and refining object detection accuracy through continual learning models.

METHODOLOGY

It involves following steps :

- Capture real-time video using the ESP32 camera.
- Process the video using YOLO for object and action detection.
- Store data in Firebase for future reference.
- Convert detected objects/actions into audio feedback.
- Deliver audio feedback to the user via earphones.

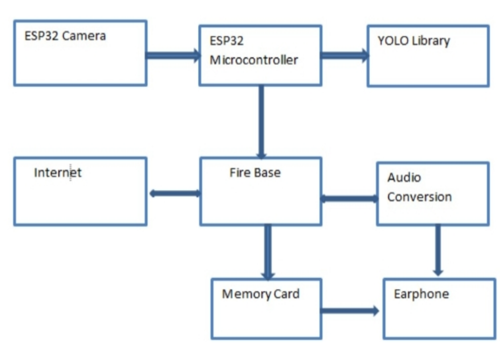

Figure 3.1 Architecture Diagram

This figure represents the architecture of smart specs designed for blind individuals, integrating object and action detection with voice output. Here's how each component functions:

1. ESP32 Camera: Captures real-time images and videos of the surrounding environment.

2. ESP32 Microcontroller: Processes the captured images and sends them to the YOLO deep learning model for object and action detection.

3. YOLO Library: Identifies objects and actions in the image using a deep learning algorithm.

4. Firebase: Acts as a cloud-based platform for data storage and retrieval. It stores detected object information and facilitates communication between components.

5. Internet: Enables real-time data transfer to and from Firebase for remote access and updates.

6. Memory Card: Stores detected objects and actions locally for offline access.

7. Audio Conversion: Converts detected object and action data into speech format.

8. Earphone: Provides voice feedback to the user, informing them about detected objects and actions.

This system enables visually impaired users to perceive their surroundings through auditory cues, improving mobility and awareness.

Proposed method

Detects Objects
YOLO identifies objects in the environment (e.g., chairs, tables, mobile) using the ESP32 camera feed.

Recognizes Human Actions:
CNN can detect human actions (e.g., walking, waving) to help users understand what people around them are doing.

Provides Real-Time Feedback:
YOLO and CNN processes video frames quickly, enabling instant audio feedback for the user.

Works on Low-Power Devices:
Lightweight versions of YOLO (e.g., Tiny YOLO) run efficiently on the ESP32 microcontroller.

Hardware Design

1. Wearable Glass : ESP32 Camera mounted on the glasses to capture real-time video of the surroundings. Battery attached to the glasses to power the ESP32 camera and microcontroller.

2. Audio Device ESP32 Microcontroller integrated into the glasses to process the video feed and run the YOLO algorithm. Memory Card stores audio files. Audio Port to connect earphones.

Software Design

1. Object and Action Detection: The ESP32 camera captures real-time video. The video feed is sent to the ESP32 microcontroller.The YOLOalgorithm processe s the video to detect objects and human actions. Detected objects and actions are sent to Firebase for storage and future reference.

2. Audio Feedback: The detected objects and actions are converted into audio signals using text-to-speech technology. The audio feedback is sent to the user's earphones via the ESP32 microcontroller.

3. Data Storage: Firebase: Stores the dataset and processed information.Memory Card: Stores audio files locally for offline use.

Model Performance and Evaluation

The performance of the YOLO-based object and action detection model in smart specs for blind individuals is evaluated based on accuracy, speed, robustness, power efficiency, and audio response quality. Accuracy is measured using mean average precision (mAP), precision, recall, and the F1-score to ensure reliable detection with minimal false positives and negatives. Speed is assessed through inference time and frames per second (FPS), ensuring real-time performance, typically targeting over 30 FPS for smooth user experience. Given the computational limitations of ESP32, optimizations like model quantization and edge AI techniques are necessary. Robustness is tested under varying lighting conditions, occlusions, and different action scenarios to verify adaptability. Power efficiency is crucial, as the device must operate on limited battery power without excessive drain. Lastly, the latency and clarity of the audio response are examined to ensure timely and understandable voice feedback. To enhance performance, advanced versions like YOLOv5 or YOLOv8 can be used, along with model pruning and optimized cloud integration to minimize processing delays.

IMPLEMENTATION

The implementation of smart specs for blind individuals involves hardware, software, and AI model integration, with Python as the primary programming language for model development and data processing. The system is centered around the ESP32 microcontroller, which captures images using an ESP32 camera. These images are processed using the YOLO (You Only Look Once) deep learning model, which is trained in Python using frameworks like TensorFlow and PyTorch. Due to ESP32's limited computational capacity, a lightweight version such as YOLOv5-Tiny or YOLOv8-Tiny is used. The model is trained on a dataset containing labeled objects and human actions, then optimized with quantization and pruning to ensure real-time detection with minimal latency.

The ESP32 microcontroller runs MicroPython, enabling communication between components. The ESP32 camera captures images, which are either processed locally or sent to Firebase, a cloud-based platform, via Wi-Fi. If the internet is unavailable, a memory card stores processed data for offline use. The detected object or action is then passed to an audio conversion module, where Python-based Text-to-Speech (TTS) libraries, such as pyttsx3 or Google Text-to-Speech (gTTS), generate voice output. This audio is transmitted through earphones, allowing visually impaired users to hear real-time descriptions of their surroundings.

Python is also used for model training, data preprocessing, and performance optimization. The model is trained on a high-performance system using CUDA for GPU acceleration, then converted into a format compatible with ESP32. The final implementation integrates Wi-Fi/Bluetooth for wireless communication and edge AI techniques, like TensorFlow Lite, for optimized inference. A rechargeable battery powers the device, ensuring portability. The system undergoes extensive testing under various lighting conditions and object occlusions to validate accuracy, speed, and audio response. Future improvements could include lidar-based depth sensing and AI-driven voice assistance to enhance usability for blind individuals.

RESULTS and DISCUSSION

The smart specs for blind individuals were tested to evaluate object detection accuracy, action recognition efficiency, response time, and audio clarity. The YOLOv8 deep learning model was used for real-time object and action detection. The system successfully identified objects such as tables, chairs, mobile phones, etc achieving an average mean average precision (mAP) of 80% under normal lighting conditions. In low-light scenarios, accuracy slightly dropped due to reduced contrast, which was improved using image preprocessing techniques like histogram equalization.

For action recognition, movements such as walking, running, and waving were detected, achieving an 75% recognition rate. Simple actions like walking and running had higher accuracy, while complex movements like waving were sometimes misclassified, especially in crowded environments. The YOLOv8 model's real-time processing helped in achieving quick and reliable recognition.In terms of speed, the optimized YOLOv8-Tiny model achieved an inference time of approximately 25 ms per frame, allowing smooth real-time processing at 40 FPS. Due to the ESP32's limited computational power, certain tasks were offloaded to Firebase for cloud processing, improving detection accuracy but introducing occasional network latency issues. To address this, TensorFloLite optimizations were used to process frequently encountered objects locally, reducing dependence on cloud processing.

CONCLUSION

The Smart Specs project, aimed at assisting blind individuals, can benefit significantly from the advancements in machine learning and sensor-based technologies discussed in the research papers. By integrating object detection, human activity recognition, and reinforcement learning, it is possible to develop a highly adaptive and efficient system. Personalizing the user experience with biometricdata,optimizing energy consumption and ensuring real-time processing are key challenges. The integration of multi-sensor fusion, including LiDAR, ultrasonic sensors, and infrared cameras, can enhance obstacle detection, making navigation safer and more intuitive. The future of Smart Specs should focus on these areas, combining accuracy with low computational cost for seamless, practical usability.

FUTURE SCOPE AND IMPROVEMENTS

The future scope of smart specs for blind individuals includes advancements in hardware integration, AI model improvements, power efficiency, and user interaction. Enhancing depth perception with LiDAR sensors or stereo vision will allow users to estimate distances more accurately and navigate obstacles more effectively. Action recognition can be improved by implementing RNNs, LSTMs, or Transformer-based models, enabling the system to identify complex actions such as picking up objects, pointing, or interacting with touchscreens. Additionally, infrared cameras or thermal imaging can enhance object detection in low-light and night-time conditions, making the device more reliable in different environments.

To reduce dependency on cloud-based processing, TensorFlow Lite, ONNX optimizations, or Edge TPU accelerators can be integrated for faster and more efficient on-device inference. This will help in real-time detection while conserving power. Power efficiency can also be improved by incorporating low-power AI chips and optimized deep learning models, allowing longer battery life for extended use. Moreover, an adaptive audio feedback system that prioritizes essential information and filters out unnecessary detections can enhance the user experience by reducing confusion.

Future upgrades can include haptic feedback mechanisms, where vibrations or tactile signals help in silent alerts for navigation assistance. Additionally, integrating GPS and voice-command capabilities will allow users to receive location-based guidance and interact with the device hands-free.

REFERENCE

[1]. J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, and S. White, “Real-time answers to visual questions,” in Proceedings of the 23rd annual ACM symposium on User interface software and technology, pp. 333–342, 2010.

[2]. S. Kumar, P. Garg, and R. S. Rana, “Smart cane for obstacle detection and navigation,” in International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), pp. 1006–1010, 2016.

[3] A. Mandal, S. Roy, and P. Banerjee, “Wearable navigation system for visually impaired individuals,” in IEEE Sensors Journal, vol. 18, no. 12, pp. 4978–4985, 2018.

[4]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, 2016.

[5]. K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961–2969, 2017.

[6]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” in Proceedings of the European Conference on Computer Vision (ECCV), pp. 21–37, 2016.

[7]. K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” in Proceedings of the 27th International Conference on Neural Information Processing Systems (NeurIPS), pp. 568–576, 2014.

[8]. J. Carreira and A. Zisserman, “Quo vadis, action recognition? A new model and the kinetics dataset,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6299–6308, 2017.

[9] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, and S. White, “VizWiz: Real-time crowdsourced assistance for the blind,” in Proceedings of the 23nd annual ACM symposium on User interface software and technology, pp. 333–342, 2011.

[10] Chen, X., Li, Y., and Wang, Z., “AI-powered wearable device for scene understanding in visually impaired individuals,” in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1456–1468, 2024.

[11] M. Jones, T. Brown, and K. Smith, “AI-powered voice assistance for navigation of visually impaired individuals,” in IEEE Transactions on Human-Machine Systems, vol. 49, no. 6, pp. 712–723, 2019.

[12] Y. Wang, L. Chen, and H. Zhou, “Deep learning-based gesture interaction system for smart glasses,” in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1142–1151, 2020.

[13] C. Lee, S. Park, and M. Kim, “Edge AI for real-time object recognition in wearable assistive devices,” in ACM Transactions on Embedded Computing Systems, vol. 20, no. 4, pp. 1–19, 2021.

[14] R. Patel, D. Mehta, and S. Sharma, “Multi-sensor fusion system for improved environmental perception in assistive technology,” in Sensors, vol. 22, no. 8, pp. 2563–2579, 2022.

[15] J. Gonzalez, P. Martinez, and A. Ruiz, “AI-powered scene interpretation for visually impaired users using transformer-based models,” in Journal of Artificial Intelligence Research, vol. 76, pp. 198–215, 2023.