Smart Voice Assistant For Visually Impaired People Using Deep Learning Paradigm

Dr. R. Amutha
Department of Electronics and Communication Engineering
Sri Sivasubramaniya Nadar College of Engineering,
Chennai,India
amuthar@ssn.edu.in

Pranav S P
Department of Electronics and Communication Engineering
Sri Sivasubramaniya Nadar College of Engineering,
Chennai,India
pranav2011034@ssn.edu.in

Gokulram A
Department of Information Technology
Sri Sivasubramaniya Nadar College of Engineering,
Chennai,India

***Abstract*—The SVA, a fusion of cutting-edge technologies, blends machine learning and text-to-speech synthesis to create an innovative solution tailored for individuals with visual impairments. A Raspberry Pi is responsible for processing data, while a dedicated camera module captures real-time environmental visuals. These visuals are swiftly processed into frames by the Pi, allowing for deep learning-based object detection. Leveraging pre-trained models, the system identifies objects within the frames and conveys their descriptions through a speaker. This transformative innovation empowers the visually impaired, offering them a cohesive understanding of their environment in real-time, thereby nurturing greater independence and confidence in navigating their surroundings. The compact and portable nature of this setup, coupled with its ability to provide audible information, promises an impactful solution for the visually impaired, enhancing their everyday experiences. Data processing has been shifted to the cloud to address latency issues with the Raspberry Pi module, while concurrent threading of image capturing and text-to-speech synthesis has significantly boosted operational efficiency.**

Keywords—Smart Voice Assistant (SVA)

I. Introduction

The paper majorly deals with developing a voice assistive device which can be worn by the user to provide details about the surrounding like the objects and obstacles coming in the front of the user to assist the visually impaired person in travelling without any external assistance making them independent. The project utilizes deep learning to transform the device from just a simple voice assistant to a travel tool which gives its wearer the confidence to traverse alone thanks to the accurate detection and real-time direction.

People worldwide, who are visually impaired face everyday challenges such as figuring out new places and feeling less independent. Although traditional solutions frequently fall short in terms of accuracy, cost, personalization, and user-friendliness, assistive technologies have arisen as hope-filled alternatives.

II. Literature Review

Hengle et al.[2] introduce the Smart Cap, a novel assistive technology leveraging deep learning and IoT to aid visually impaired individuals. By integrating sophisticated algorithms and connectivity features, the Smart Cap offers real-time assistance in navigation and obstacle avoidance. The system's ability to interpret environmental data and provide intuitive feedback demonstrates a significant step forward in wearable assistive devices. Through its seamless integration of advanced technologies, the Smart Cap holds promise for enhancing the independence and safety of visually impaired users in various environments.

A recent study by Căilean et al. (2023) explores the use of visible light communication (VLC) in smart backpacks designed for visually impaired people. VLC technology transmits data using visible light, potentially offering navigation assistance and obstacle detection for those with visual impairments. The research details the design and evaluation of a prototype backpack, suggesting this could be a promising new approach to assistive technology, improving mobility and independence for visually impaired individuals.[1]

The research paper "VeRO: Smart Home Assistant for Blind with Voice Recognition" by Triyono et al. (2021) proposes a system called VeRO to assist blind people in their homes. This system utilizes voice commands to control smart home devices like lights and appliances. Additionally, VeRO can access information and entertainment for the user, such as news updates and internet radio streaming. By offering voice-controlled interaction with their surroundings, VeRO contributes to increased accessibility and independence for blind individuals within their homes.[14]

In their 2015 study, Csapó et al.[3] offer a comprehensive survey of assistive technologies available for blind users on mobile platforms. Their review illuminates the wide array of applications and tools crafted to tackle the distinct challenges confronted by visually impaired individuals. Through an in-depth analysis of existing solutions, the authors furnish valuable insights into the prevailing landscape of mobile-based assistive technology, pinpointing avenues for further exploration and advancement. This seminal work serves as an indispensable reference for researchers and developers venturing into the realm of assistive technology for the visually impaired.

Arora et al.[5] unveil a real-time multi-object detection system finely tuned for blind users, harnessing advanced computer vision techniques. Through the deployment of machine learning algorithms, the system adeptly discerns and categorizes objects in the user's surroundings, furnishing valuable auditory feedback. This strategy amplifies the user's spatial comprehension and streamlines independent navigation, thereby bolstering mobility and safety. The incorporation of real-time object detection charts a promising trajectory for the evolution of assistive technologies catering to the visually impaired.

In 2020, Adil et al. [4] introduce an innovative IoT-based voice-controlled blind stick designed to enrich navigation for visually impaired individuals. By amalgamating voice recognition technology with a conventional mobility aid, the authors present a hands-free solution for navigation and obstacle detection. The system's integration with IoT facilitates instantaneous communication and feedback, heightening the user's situational awareness and overall mobility. This research marks a notable stride forward in crafting intuitive and inclusive assistive devices tailored for the visually impaired.

In their research, Aladrén et al.[6] direct their focus towards navigation assistance for the visually impaired, leveraging RGB-D sensors with expanded range. The study delves into the fusion of sensor technology to furnish intricate spatial data and detect impediments in the user's path. By broadening the scope of detection capabilities, the system fortifies the user's adeptness in traversing intricate environments securely. This investigation underscores the pivotal role of sensor-driven guidance systems in fostering autonomous mobility among visually impaired individuals, paving the way for forthcoming strides in assistive technology.

Agarwal et al. [7] introduce an ultrasonic stick designed to aid the visually impaired in navigation tasks. This assistive device utilizes ultrasonic sensors to detect obstacles and provide haptic or auditory feedback to the user, enhancing spatial awareness. By offering real-time feedback based on environmental cues, the ultrasonic stick enables users to navigate safely and independently in various settings. The simplicity and effectiveness of this technology make it a valuable tool for enhancing mobility and autonomy for visually impaired individuals.

Dang et al. [8] propose a virtual blind cane system that utilizes a line laser-based vision system and an inertial measurement unit (IMU). By combining laser-based object detection with motion sensing technology, the system provides users with real-time feedback about their surroundings. This innovative approach improves upon traditional white cane methods by offering more accurate and comprehensive environmental information. By leveraging advanced sensor technologies, the virtual blind cane enhances the mobility and safety of visually impaired individuals, particularly in dynamic and crowded environments.

In their 2017 discussion, Sosa-García and Odone delve into "hands-on" visual recognition systems customized for visually impaired users. Employing computer vision algorithms, these systems recognize objects and scenes in real-time, providing auditory or haptic feedback to enhance users' environmental understanding and facilitate independent navigation. The integration of computer vision technology holds immense potential in improving the autonomy and quality of life for visually impaired individuals.[11]

Kanwal et al.[12] introduce a navigation system that integrates vision and depth sensors to assist visually impaired individuals in wayfinding. By combining visual and depth information, the system offers detailed spatial awareness and navigation guidance, surpassing traditional aids by providing more accurate environmental information. This fusion of vision and depth sensing technologies represents a significant leap forward in the development of assistive navigation systems for the visually impaired.

Sudol examines Looktel, a computer vision application designed to aid visually impaired individuals in object recognition and scene understanding. Leveraging smartphone cameras and computer vision algorithms, Looktel provides real-time object recognition and scene description, empowering users to navigate and interact with their surroundings independently. Looktel's development signifies a valuable contribution to assistive technology for the visually impaired community.[13]

In their 2016 study, Dhananjeyan et al.[9] present a navigation system tailored for visually impaired individuals, which integrates GSM and RFID technologies to aid in wayfinding. Utilizing RFID tags placed strategically in key locations, the system communicates location information to users via GSM communication, offering real-time navigation guidance and alerts. This integration of GSM and RFID technologies shows promise in addressing the unique navigation challenges faced by the visually impaired, enhancing their independence and safety in unfamiliar environments.

In 2018, Cardillo introduces an electromagnetic sensor prototype designed to assist visually impaired individuals during autonomous walking. By utilizing electromagnetic sensors, the device detects obstacles and provides feedback to users through auditory or tactile cues, enabling real-time obstacle detection and navigation assistance. This technology represents a significant advancement in assistive technology, particularly in outdoor environments, where it enhances the mobility and safety of visually impaired individuals.[10]

III. Proposed Method

A. Existing Technology

The literature study explores two approaches primarily, to advance assistive technologies for the visually impaired. The first integrates a microwave radar into traditional white canes for enhanced obstacle detection during autonomous walking. Simultaneously, recent advancements in audio and tactile feedback-based technologies, seamlessly integrated into standard mobile devices, encompass electronic travel aids and text-to-speech applications. Challenges include accessibility gaps, technical limitations, power consumption, adaptation and training requirements, reliability issues, and privacy concerns. Addressing these challenges is crucial for successful integration into the visually impaired community.

- User Training: Introducing a new technology, especially one involving complex sensors like a microwave radar, may require significant training for users to fully understand and utilize the system effectively. The learning curve could be a drawback.

- Interference and False Positives: Electromagnetic sensors, as used in the design proposed by Emanuele Cardillo et al. (2018), including microwave radars, may be susceptible to interference from other electronic devices, leading to false positives or inaccurate obstacle detection. Ensuring a high degree of accuracy is crucial for user safety.

B. Methodologies

- The pre-trained object detection model is loaded. The model is a Single Shot Detector (SSD) based on the MobileNet architecture and is trained on the COCO dataset.

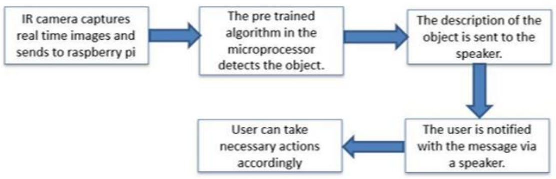

Fig 1: Process Flow Diagram

- The minimum confidence threshold for object detection is set to 0.5, which means that only objects with a confidence score greater than 0.5 will be detected. The labels for the objects that the model can detect are defined in a dictionary called labels. The known focal length and actual width of the object for distance estimation are assigned to specific variables.
- The text-to-speech engine is initialized and a queue is defined to hold the speech texts. A separate thread is started to run the speech engine. A function gets the speech text from the queue, says it using the text-to-speech engine, and then waits for the next speech text.
- In a loop, a frame is captured from the webcam using the function. The frame is converted to a blob using the cv2.dnn.blobFromImage() function. The blob is passed through the model and the detections are obtained. For each detection with a confidence score greater than the threshold, a bounding box is drawn around the object.
- The distance of the object from the camera is estimated using the known focal length, actual width of the object, and the width of the object on screen. If an object is detected for more than 2 seconds (120 frames), a speech text is generated informing the user about the object detected and the distance it is from the user.

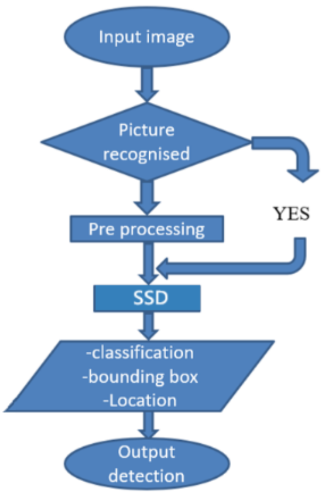

Fig 2: Flowchart of the program

- Due to latency issues with the Pi module's response time, data transfer to a server for cloud processing has been implemented, effectively reducing latency. Additionally, optimizing efficiency has been achieved by running the image capturing component and the text-to-speech synthesizer concurrently in threads.

IV. Implementation and Results

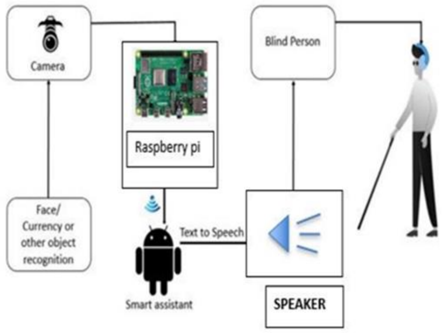

Fig 3: Implementation of the Device

A. Implementation

The initial concept aimed to integrate a Raspberry Pi, Power bank, and Pi camera module onto a cap worn by the user. However, this design faced numerous drawbacks and limitations. Firstly, it was susceptible to weather conditions like rain, posing a risk of damage to Raspberry Pi hardware .Secondly, excessive heat radiation was an issue since incorporating a cooling fan into the cap design was impractical. Additionally, the cap's weight increased significantly, and the camera stability was compromised.

Fig 4: Initial Design as a Smart Cap

In response, we transitioned to a bag model worn on the chest, positioning the camera forward for an unobstructed view. This relocation equally distributes weight across the shoulders and allows for the integration of a cooling fan, addressing heat concerns. Moreover, the bag design offers protection from adverse weather conditions safeguarding the hardware components.

Fig 5: Upgraded Design

B. Results

The device can effectively detect and recognise multiple objects captured by the camera, promptly informing the user with minimal latency. The following images depict bounding boxes along with class names and the distances of the objects identified after being processed by the device.

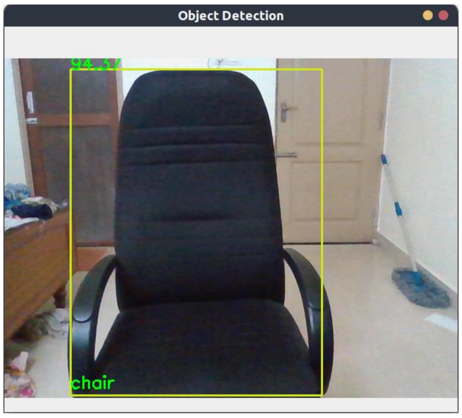

Fig 6: Output from the Detector

V. CONCLUSION

A significant development in assistive technology is the Smart Voice Assistant for Visually Impaired People, which was created utilizing a Deep Learning Paradigm. The system improves visually impaired people's accessibility, independence, and inclusion by combining voice command technology with deep learning algorithms. With its customizable, hands-free interface, users can browse, retrieve information, and complete tasks more confidently and easily. The importance of inclusive design principles in developing technology that satisfies a wide range of user needs is demonstrated by this project. As time goes on, more improvements to the Smart Voice Assistant might potentially improve people's quality of life and enable those who are blind or visually impaired to fully engage in the digital world.

To mitigate latency concerns associated with the Raspberry Pi module's response time, data transmission to a cloud server for processing has been adopted. Moreover, enhancing operational efficiency has been realized through concurrent threading of the image capturing process and the text-to-speech synthesis component. These adaptations mark significant strides in improving the functionality and performance of the Smart Voice Assistant, reinforcing its effectiveness in aiding individuals with visual impairments.

ACKNOWLEDGMENT

We express our heartfelt gratitude to Dr. V.E. Annamalai, Principal of SSN College of Engineering, for providing us with the necessary facilities to carry out our project successfully.

We would like to express our profound gratitude and deep regards to Dr. P Vijayalakshmi, Head of the Department of ECE, for her exceptional guidance, monitoring, and unwavering encouragement throughout the project's duration.

We also extend our thanks to Dr. R Amutha, Professor of ECE department at SSN College of Engineering, for her invaluable guidance, insights, and support, which helped us in completing this task through every stage.

We would also like to thank all the staff members of the Department of Electronics and Communication Engineering and our family for their constant support throughout the project

REFERENCES

[1] A. -M. Căilean, S. -A. Avătămăniței and C. Beguni, "Design and Experimental Evaluation of a Visible Light Communications-Based Smart Backpack for Visually Impaired Persons' Assistance," 2023 31st Telecommunications Forum (TELFOR), Belgrade, Serbia, 2023, pp. 1-4, doi: 10.1109/TELFOR59449.2023.10372606. keywords: {Wireless communication;Navigation;Visual impairment;Prototypes;Lighting;Transforms;Telecommunications;blind assistance;indoor guiding;obstacle detection;smart backpack;visible light communications;visually impaired assistance.},

[2] A.Hengle, A. Kulkarni, N. Bavadekar, N. Kulkarni and R. Udyawar, "Smart Cap: A Deep Learning and IoT Based Assistant for the Visually Impaired," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 1109-1116, doi: 10.1109/ICSSIT48917.2020.9214140.

[3] Ádám Csapó, György Wersényi, Tony Stockman (2015), 'A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research', J. Multimodal User Interf. 9 (4) pp.275–286 .

[4] Adil, Md & Rafa, Taiyabashadaka & Ferdoush, Jannatul & Mahmud, Mohammad & Pathak, Abhijit (2020), 'An IoT based Voice Controlled Blind Stick to Guide Blind People', International Journal of Engineering Inventions. 9. pp.9-14.

[5] Adwitiya Arora, Atul Grover, Raksha Chugh & S. Sofana Reka (2019), 'Real Time Multi Object Detection for Blind Using Single Shot Multibox Detector', Wireless Pers Commun 107, pp.651–661.

[6] Aladrén, G. López-Nicolás, L. Puig and J. J. Guerrero, "Navigation Assistance for the Visually Impaired Using RGB-D Sensor with Range Expansion," in IEEE Systems Journal, vol. 10, no. 3, pp. 922-932, Sept. 2016. doi: 10.1109/JSYST.2014.2320639

[7] Ankit Agarwal,Deepak Kumar,Abhishek Bhardwaj (2015), 'Ultrasonic Stick for Blind',International Journal Of Engineering And Computer Science ISSN, Volume 4 Issue 4,pp.2319-7242.

[8] Dang, Q., Chee, Y., Pham, D., & Suh, Y., "A Virtual Blind Cane Using a Line Laser-Based Vision System and an Inertial Measurement Unit", Sensors 2016, 95. doi:10.3390/s16010095

[9] Dhananjeyan, S., Sundaram, K. M., Kalaiyarasi, A., and Kuppusamy, P. G., "Design and Development of Blind Navigation System using GSM and RFID Technology", in Indian Journal of Science and Technology, doi:10.17485/ijst/2016/v9i2/85809

[10] Emanuele Cardillo et al. (2018), 'An Electromagnetic Sensor Prototype to Assist Visually Impaired and Blind People in Autonomous Walking',IEEE Sensors Journal, volume 18, no. 6, pp. 2568-2576.

[11] J. Sosa-García , F. Odone , Hands on"visual recognition for visually impaired users, ACM Trans. Access. Comput. (TACCESS) 10 (3) (2017) 1–30 .

[12] Nadia Kanwal, Erkan Bostanci, Keith Currie, Adrian F. Clark, "A Navigation System for Visually Impaired: A fusion of Vision and Depth Sensor", in Applied Bionics and Biomechanics. 2015. doi: 10.1155/2015/479857.

[13] Sudol, J. (2013). Looktel—computer vision applications for the visually impaired (Doctoral dissertation, UCLA).

[14] Triyono, L., et al. "VeRO: Smart home assistant for blind with voice recognition." *IOP Conference Series: Materials Science and Engineering*. Vol. 1108. No. 1. IOP Publishing, 2021.