

Trineta v1.0: An Aid to Visually Impaired Persons

Tejaswini Borekar¹, Shravani Gavli², Aditi Khambe³, Dr. Ashish Vanmali⁴

^{1, 2, 3}Department of Information Technology,

⁴Department of Electronics & Telecommunication Engineering,
Vidyavardhini's College of Engineering & Technology,
Vasai, Maharashtra, India.

Abstract - In this paper, we introduce a revolutionary solution named Trineta designed to navigate the challenges faced by visually impaired persons in their daily lives. Trineta represents a significant leap forward in providing a seamless and accessible one-touch solution for the vision loss community, aimed at fostering independence and self-reliance. This innovative system combines the strengths of the Internet of Things (IoT) and Artificial Intelligence (AI) to identify the surrounding objects and their distances from the person and translates the same as an audio output. The core functionality of Trineta involves capturing real-time images through an integrated camera, coupled with the implementation of the cutting-edge YOLO (You Only Look Once) algorithm for efficient object detection. An onboard depth sensor measures the distance of the object. This information is then translated into audio output using text-to-speech conversion, providing visually impaired persons with valuable insights about their surroundings. Trineta stands out from existing market solutions with its low cost, user-friendly design, and real-time feedback capabilities.

Keywords - Artificial Intelligence; Depth Sensor; Object Detection; Pi-camera; Raspberry Pi; YOLO.

I. INTRODUCTION

As per the World Report of Vision by the World Health Organization (WHO), approximately 40 million people in the world are blind. Additionally, another 250 million people suffer from different forms of visual impairment [1]. In India, these numbers are estimated to be around 5 million and 70 million, respectively. These people face a lot of challenges in their day-to-day life such as reading, cooking, and navigating unfamiliar environments. This also takes a toll on their mental health, leading them to frustration, anxiety, and depression.

Visually impaired individuals encounter significant challenges in their daily lives, hindering their independence and safety while navigating their surroundings. Traditional assistance methods are inadequate in delivering timely, dependable information, limiting their capacity to interact confidently with their environment. With the advancements in embedded systems and machine learning, many solutions are proposed in recent times. The readers can refer to the work of [2, 3, 4, 5] for more details. However, most of these solutions are not user-friendly or very high in cost and cannot be afforded by most of the people.

There are few commercial solutions available in the market. For example, OrCam MyEye Smart [6] is an AI device that helps people with visual impairments read text aloud from any surface in real-time. Their cost starts at approximately INR 3.5 Lakhs. Envision Glasses [7] is another such product that

provides functionalities like reading text, describing scenes, recognise cash, detecting colours, identifying people, finding objects etc. Their cost starts from approximately INR 1.49 Lakhs.

As a part of the Trineta project, we aim to provide a low-cost, user-friendly solution that can be made available to the masses to ease their day-to-day lives with the integration of the IoT and AI. We plan to provide different capabilities like reading, person identification, navigating in indoor environments, navigating in outdoor environments, road crossing assistance, etc. We aim to offer a basic version starting at INR 10,000, with the full-featured version available for up to INR 25,000. Trineta v1.0 provides proof of concept for the indoor navigation capabilities. Visually impaired people can move around independently in the indoor environment, reducing their reliance on others for assistance with Trineta v1.0.

II. PREVIOUS WORK

There have been many attempts in literature to provide aid to visually impaired persons. The methods vary based on the type of camera used, either monocular or stereo vision, the depth sensing mechanism, the type of processing system used, the type of object detection algorithm, and so on. This section provides a few glimpses of these methods.

The mechanism will greatly depend on the camera system used. The cameras can be monocular, stereo vision, or a combination capturing visible-infrared spectrum. The depth-sensing mechanism will depend on the type of camera used. A few variations of these works are presented below:

- *Marzullo et al., "Vision-based Assistive Navigation Algorithm for Blind and Visually Impaired People Using Monocular Camera"* [8]: Marzullo et al. employed a three-stage image processing method using a monocular camera. They used the first-order Hough transform to detect regions like walls, doors, windows, etc. Feedback using a physical medium is provided to the user for necessary corrections of trajectory.
- *Liu and Aggrawal, "Local and Global Stereo Methods"* [9]: Liu and Aggrawal have presented a detailed overview of the concept of stereo vision. They reviewed different local and global algorithms used for depth estimation. Local methods are faster compared to global methods, which are based on optimization. However, global methods give more accurate disparity map as opposed to local methods.

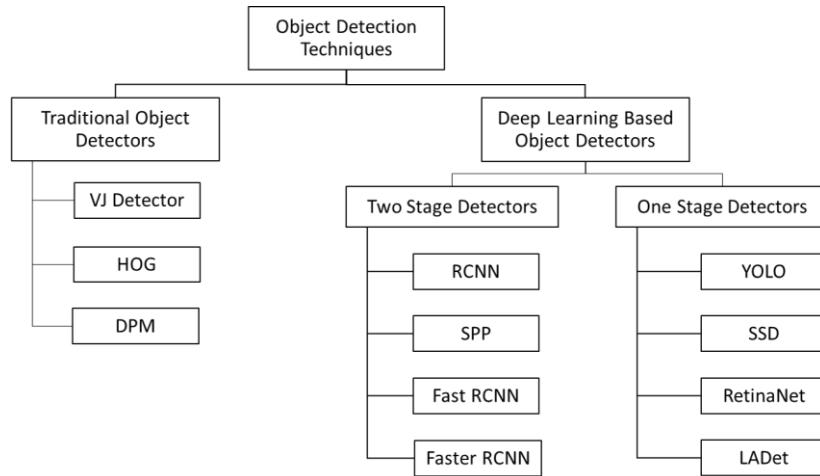


Fig. 1. Object Detection Methods (Reproduced from [14])

- Žbontar and LeCun, "Computing the Stereo Matching Cost with a Convolutional Neural Network" [10]: Žbontar and LeCun used Convolutional Neural Networks (CNN) for computing the stereo matching cost. The supervised learning reduced the error rate drastically. However, the network is slow and not suitable for real-time implementation.
- Adi et al., "Blind People Guidance System using Stereo Camera" [11]: Adi et al. introduces a supportive solution that employed stereo cameras to help those who are visually impaired identify items and barriers in their immediate surroundings. It makes use of a ZED stereo camera and a computer to precisely compute the difference between nearby obstacles and conveys this information to the user through stereo sounds. The system is claimed to have an accuracy of 83.16%.
- Zhong et al., "Real-time depth map estimation from infrared stereo images of RGB-D cameras" [12]: In this paper, the authors used RGB-D cameras for 3D depth perception instead of using regular stereo cameras. They presented a robust and effective coordination formula based on semi-global synchronization principles to generate real-time, accurate, and comprehensive depth maps that consider the specific characteristics of infrared speckle images. It also used the idea of block matching to improve the system performance.
- Vanmali et al., "A Novel Approach For Image Dehazing Combining Visible-NIR Images" [13]: Vanmali et al. used a combination of visible and near infrared (NIR) images for depth map estimation followed by dehazing of images.

In the case of monocular vision, the identification of objects is easy. However, these systems struggle to provide accurate measurement of the depth of the object. The stereo vision systems can provide an accurate disparity map for depth measurement. However, they are computationally heavy and are difficult to implement on embedded systems with limited computing power. Systems with more computing power are required for real-time implementation of stereo vision systems, which increases the cost of the system. The same is applicable for systems involving RGB-D camera systems or visible-NIR camera systems.

Another crucial part of the project is object detection. There are a variety of methods available for object detection. Kaur and Singh have provided a detailed review of the object detection methods [14]. A quick glance at these methods is shown in Fig. 1 (reproduced from [14]). Kaur and Singh have divided object detection techniques into two broad classes, viz., traditional detectors and deep learning-based detectors. The early days of computer vision used traditional detectors in view of limited computational resources using manually crafted features. VJ (Viola-Jones) detector, Histograms of Oriented Gradients (HOG), and Deformable Part-based Model (DPM) are examples of popular traditional detectors.

With the advancement of technology, deep neural network-based object detectors have become popular due to their accuracy and flexibility. Deep learning-based object detection techniques are generally categorized into two types: two-stage and one-stage detectors. Two-stage detectors operate in two steps to identify objects within an image and are known for delivering state-of-the-art accuracy on standard datasets. However, this comes at the cost of slower inference speed. In contrast, one-stage detectors perform detection in a single step, making them significantly faster and more suitable for real-time applications, though they may sacrifice some accuracy compared to two-stage models.

The flexibility, speed, and accuracy of You Only Look Once (YOLO) makes it most suitable for real-time applications. The details of YOLO are explained below:

- Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection." [17]: YOLO offers a unique perspective on object tracking by diverging from conventional methods and avoiding the need to adapt prior approaches. Instead of separate steps, YOLO treats object tracking as a regression problem while simultaneously creating classifiers for identification tasks. This approach involves making predictions of bounding boxes with associated class probabilities within a single neural network. YOLO excels at accurately estimating both the bounding boxes and class probabilities for entire images in a single pass. The advantage of this approach lies in its unified network, which streamlines the entire detection process.

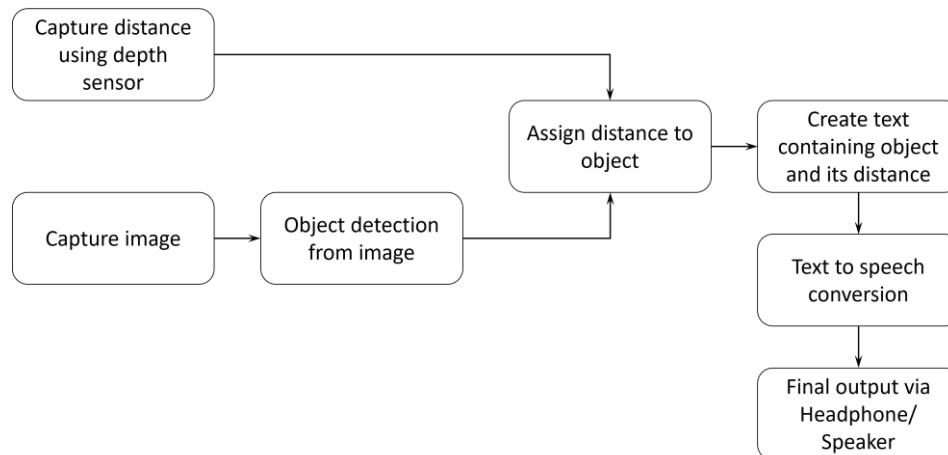


Fig. 2. Block Diagram of the Proposed Solution

- *Terven et al., "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS"* [16]: Terven et al. conducted an in-depth analysis of the YOLO architecture's evolution. They highlighted the key innovations and contributions introduced in each version from the original YOLO through to YOLOv8.
- *Atitallah et al., "An effective obstacle detection system using deep learning advantages to aid blind and visually impaired navigation"* [17]: Atitallah et al. proposed an obstacle detection system based on a modified YOLO v5 neural network architecture. They tested the system on IODR datasets and MS COCO datasets for indoor and outdoor operations.
- *M. Yaseen, "What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector"* [18]: The newest of the YOLO family is YOLO v8. YOLO v8 employs the CSPNet backbone, and the FPN+PAN neck makes it more versatile compared to its predecessors.

III. PROPOSED SOLUTION

The proposed solution combines the strengths of IoT and AI to give a real-time solution that will aid visually impaired persons. Based on the literature surveyed, we decided to use YOLO as an object detection technique along with a separate depth sensing mechanism to keep the system simple and implementable in real-time. The block diagram of the proposed solution is given in Fig. 2. The corresponding interfacing of IoT system is shown in Fig 3.

The key hardware components of our IoT system are:

- *Ultrasonic Sensor*: Various types of distance sensors are available for IoT systems, which include ultrasonic, IR proximity, laser distance, etc. The ultrasonic sensor emits high-frequency sound waves towards the target object and records the round-trip time from the target for the reflected waves to calculate the distance. We have used an ultrasonic sensor HC-SR04 for the proposed system, which has a range of 2cm-450cm.

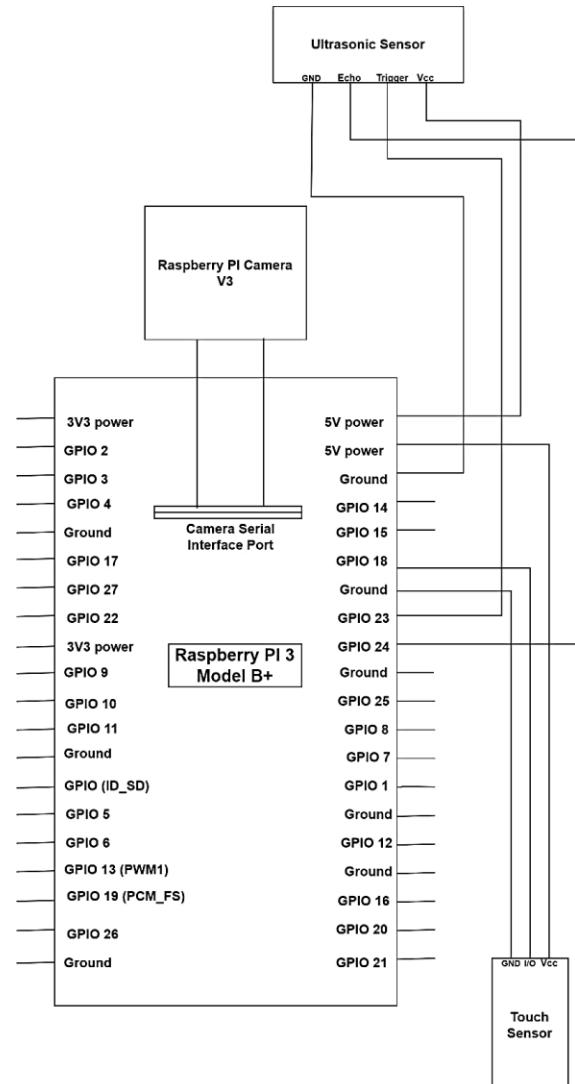


Fig. 3. Interfacing of IoT System

Step 1: Initialize the system.
 Step 2: Check the status of touch sensor.
 Step 3: If touch is recognized on the touch sensor
 (a) Capture a frame using Raspberry Pi camera.
 (b) Perform object detection on the frame using YOLO v3.
 (c) Measure the distance of the object using ultrasonic sensor.
 (d) Create a text containing object name and distance.
 (e) Convert the text to audio output using gTTS.
 (f) Feed the audio output to speaker/headphone.
 Step 4: Go back to Step 2.

Fig. 4. Main Steps of Operation

- *Raspberry Pi*: Raspberry Pi is a series of small single-board computers (SBCs) developed by the Raspberry Pi Foundation and forms the heart of our IoT system. Raspberry Pi boards are small, compact, and easy to install and maintain. They are highly energy-efficient and provide support for various types of sensors and actuators. They also have an interface for the camera, which is the main requirement of the proposed system. We have used the Raspberry Pi 3 Model B+ with 4GB RAM with laptop integration for power supply. For more features, one needs to use a Raspberry Pi 5 with 8GB RAM or higher.
- *Raspberry Pi Camera*: Raspberry Pi has readymade support for the Raspberry Pi camera modules, which can be connected with flex cable. We have used the Raspberry Pi Camera v3 module which can record full HD video at 50fps with autofocus functionality.
- *Touch Sensor*: Touch sensors or tactile sensors can detect touch and operate as a switch. We have used TTP223 1-Channel Capacitive Touch Sensor Module,

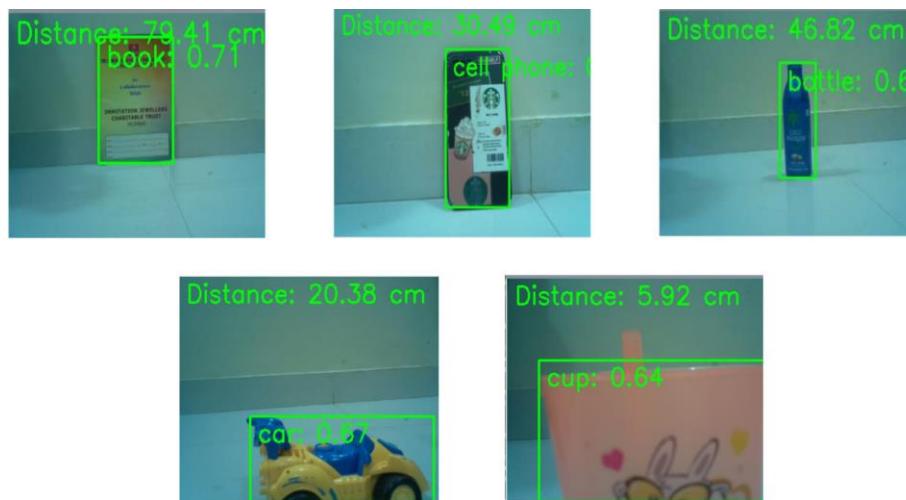
**Fig. 5. Real-time Implementation**

which works as a trigger for the system.

The key software components of the proposed system are:

- *OpenCV*: OpenCV is an open-source computer vision and machine learning software library. It provides a comprehensive set of tools for image and video processing. It performs tasks such as image and video analysis, object detection, face recognition etc.
- *YOLO v3*: Since we have used the Raspberry Pi 3 Model B+ with 4GB RAM, we have used the pretrained model of YOLO v3 for object detection. The model is trained for 80 different objects and is available at [19]. For higher versions of YOLO, one needs to go for Raspberry Pi 5 with 8GB RAM or higher.
- *Text To Speech*: Text to speech (TTS) conversion provides real-time auditory feedback by converting information that is text-based, like the name of the detected object and its distance, into spoken words. We used gTTS (Google Text-to-Speech) library [20] in our project.

The main steps of operation of the Trinetra system are depicted in Fig. 4.

**Fig. 6. Outputs for Different Objects**

IV. RESULTS

The photograph of real-time implementation of our Trinetra v1.0 system is shown in Fig. 5. The corresponding results of various objects detected with the confidence level and their distances are shown in Fig. 6.

It is observed that the system gives precise distance for the objects that are in line of sight of the ultrasonic sensor. We have used 50% as a threshold level for the confidence level of object detection using YOLO v3. Hence, only these objects above the threshold will be detected by the proposed system.

V. CONCLUSION AND FUTURE SCOPE

The Trinetra v1.0 marks a significant milestone in addressing the needs of visually impaired individuals. By integrating object detection, depth estimation, and voice feedback technologies, it has created a comprehensive solution for enhancing the safety and independence of visually impaired persons. The system's performance in identifying objects, providing real-time distance measurements, and simulating object interactions has yielded promising results.

The current system with laptop integration costs approximately INR 7,500 without battery. With the onboard rechargeable battery, the cost will be approximately INR 10,000. With the inclusion of Raspberry Pi 5 with 8GB RAM or higher, we can have a system with more features ranging from INR 15,000 to 25,000 as opposed to OrCam (INR 3.5 Lakhs) and Envision Glasses (INR 1.49 Lakhs) systems.

The Trinetra system can focus on addressing existing challenges, improving system functionality, and expanding its capabilities to further benefit visually impaired individuals. Here are some potential areas for future development:

- The current system can give the distance of only one object, which can be extended for multiple objects.
- The current system works well for static objects. One can improvise it for moving objects as well.
- Integration with wearable devices like smart glasses, which can provide a hands-free and intuitive means for visually impaired individuals to access Trinetra's information and feedback.
- Providing a reader mode to read signboards, product labels, forms, books, etc.
- Person identification mode to identify known persons in the visibility.
- Cloud integration to make the product scalable and to accumulate data to enhance model capabilities.

As the project continues to evolve and address its challenges, it holds promise as a transformative tool for the visually impaired community.

REFERENCES

[1] World Health Organization, World report of vision. Available at: <https://www.who.int/publications/item/9789241516570>, Last accessed on - Dec. 2024.

[2] E. D'Atri, C. M. Medaglia, A. Serbanati, U. B. Ceipidor, E. Panizzi and A. D'Atri, A system to aid blind people in the mobility: A usability test and its results, in *Second International Conference on Systems (ICONS'07)*, 2007, pp. 35–35.

[3] L. Dunai, G. P. Fajarnes, V. S. Praderas, B. D. Garcia and I. L. Lengua, Real-time assistance prototype - a new navigation aid for blind people, in *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, 2010, pp. 1173–1178.

[4] N. Tyagi, D. Sharma, J. Singh, B. Sharma and S. Narang, Assistive navigation system for visually impaired and blind people: A review, in *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*, 2021, pp. 1–5.

[5] M. Bousbia-Salah and M. Bettayeb, "A navigation aid for blind people," *Journal of Intelligent & Robotic Systems*, vol. 64, pp. 387–400, 2011.

[6] OrCam, Available at: <https://www.orcam.com/en-us/home>, Last accessed on – May 2025.

[7] Envision Glasses, Available at: <https://shop.letsenvision.com/en-in/collections/envision-glasses>, Last accessed on – May 2025.

[8] G. D. Marzullo, K.-H. Jo, and D. Caceres, "Vision-based assistive navigation algorithm for blind and visually impaired people using monocular camera," in *2021 IEEE/SICE International Symposium on System Integration (SII)*, 2021, pp. 640–645.

[9] Y. Liu and J. Aggarwal, "3.12 - Local and global stereo methods," in *Handbook of Image and Video Processing*, Second Edition ed., A. Bovik, Ed. Burlington: Academic Press, 2005, pp. 297–308.

[10] J. Žbontar, and Yann LeCun, "Computing the Stereo Matching Cost with a Convolutional Neural Network, " in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1592–1599.

[11] I. P. Adi, H. Kusuma, and M. Attamimi, "Blind people guidance system using stereo camera," in *2019 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2019, pp. 298–303.

[12] J. Zhong, M. Li, X. Liao, J. Qin, H. Zhang, and Q. Guo, "Real-time depth map estimation from infrared stereo images of rgb-d cameras," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-2-2021, pp. 107–112, 2021.

[13] A. V. Vanmali, S. G. Kelkar, and V. M. Gadre, "A novel approach for image dehazing combining visible-nir images," in *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, 2015, pp. 1–4.

[14] J. Kaur, and W. Singh, "Tools, techniques, datasets and application areas for object detection in an image: a review," *Multimed Tools and Applications*, vol. 81, pp. 38297–38351, 2022.

[15] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779–788.

[16] J. Terven, D. Córdova-Esparza, and J. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023.

[17] A. Ben Atitallah, Y. Said, M. A. Ben Atitallah, M. Albekairi, K. Kaaniche, and S. Boubaker, "An effective obstacle detection system using deep learning advantages to aid blind and visually impaired navigation," *Ain Shams Engineering Journal*, vol. 15, no. 2, p. 102387, 2024.

[18] M. Yaseen, "What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," in *2024 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[19] YOLO v3. Available at: <https://github.com/ultralytics/yolov3>, Last accessed on - Dec. 2024.

[20] gTTS. Available at: <https://gtts.readthedocs.io/en/latest/>, Last accessed on - Dec. 2024.