iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68713

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Research on Real -Time Object Detection with Speech Feedback for Visually Impaired

Mr. M. S. Khan[1], Ritul Ghumare[2], Prerna Malode[3], Rutuja Murkute[4], Sakshi Pagare[5]
[1]*Assistant Professor, Department of Information Technology, Matoshri College of Engineering & Research Centre, Eklahare, Maharashtra, India*
[2,3,4,5]*Students, Department of Information Technology, Matoshri College of Engineering & Research Centre, Eklahare, Maharashtra, India*

Abstract: ***It is widely recognized that approximately 285 million people worldwide are visually impaired, which accounts for nearly 20% of India's population. Many of these individuals rely heavily on others to fulfill even their most basic daily needs. In our project, we utilized TensorFlow, a cutting-edge library developed by Google, to power our neural network models. The TensorFlow Object Detection API was employed to detect various objects in real-time. Additionally, we introduced the YOLO algorithm, which plays a pivotal role in object detection. During training, YOLO matches anchor boxes with the bounding boxes of ground truth objects in an image. The anchor box with the highest overlap with an object is responsible for predicting that object's class and location.***

Our system also integrates a microcontroller equipped with a Wi-Fi module for seamless communication. This user-friendly interface allows visually impaired users to navigate both indoor and outdoor environments with ease. To further enhance the user experience and reduce navigation challenges, we incorporated an obstacle detection system using ultrasonic sensors. This system identifies nearby obstacles and alerts the user, helping them avoid potential hazards and navigate with greater confidence.

Keywords:Python, Machine Learning, YOLO lib. Datasets, Opencv.

I. INTRODUCTION

The technology available for the navigation of visually impaired individuals is still not sufficiently accessible, as many devices depend heavily on infrastructure. Without vision, navigating through rooms or different paths can be particularly challenging for the blind. The primary goal of this project is to assist visually impaired people by detecting obstacles and road traffic signs, ultimately making their lives easier. With this system, blind individuals will be able to navigate freely without the need for assistance. They will be able to walk through streets, parks, and unfamiliar environments independently.

To prevent users from entering dangerous areas, the system collects data from various environmental sources (e.g., cameras, sensors, scanners) and transmits it to the user in an audio format. When the sensors detect objects, the data is sent to the audio module, which converts it into an audio signal. This system is capable of detecting obstacles as close as 4 cm. The goal of this project is to provide an affordable solution for blind people, enabling them to plan their paths safely and efficiently.

By combining object detection and navigation systems, visually impaired individuals can easily identify obstacles through cameras or sensors. For example, it can be difficult for blind individuals to find a specific room in an unfamiliar environment, or even recognize if someone is speaking to them. This project aims to address these challenges using computer vision technologies, particularly neural networks, which have made significant advancements in recent years. The system guides blind people by processing the information through a controller and providing real-time navigation assistance.

The purpose of this project is to develop a system that helps blind people lead a more independent and effective life. With this system, blind individuals no longer need to rely on others to perform daily activities. They can independently navigate public spaces, traffic areas, and parks, enjoying the world around them. Many blind individuals are curious about their surroundings, and with this system, they can experience the beauty of the world more fully.

This portable device is specifically designed to assist visually impaired individuals in navigating their environments. Unlike many commercially available devices, this system provides real-time directions and alerts users to obstacles in their path. One of the most crucial features is the ability to warn users of any obstacles and guide them through different areas via voice assistance. The system also detects road traffic signs, making it easier for blind people to navigate urban environments. With all these features, the lives of blind individuals can be greatly improved, and they can enjoy a higher degree of freedom and independence.

II. LITERATURE SURVEY

Abdul Muhsin M, Farah F. Alkhalid, Bashra Kadhim Oleiwi have proposed their work on "Online Blind Assistive System Using Object Detection" in 2020. In this work, the function of computer vision is to detect indoor objects accurately. The visually impaired people can be assisted by navigating the purposes of the CNN framework.4,5,14 To identify the specific objects first, we need to detect the pixels available in the images. If the lighting conditions are wrong, then it is challenging to capture and identify the objects with high accuracy. To detect the indoor objects, the algorithm needs to extract the image features with a particular class, and it can be done by RetinaNet.25 To enable the network for small object detection by a Region Proposal Int J Cur Res Rev

The object detection system in [3] is used to detect objects in the traffic scenes. Here they have used the combination of optimized you only look once (OYOLO), which is 1.18 times faster than YOLO and R-FCN (Regression based Full Convolution Network). It is used to detect and classify the images such as cars, cyclist and pedestrian. Use of YOLO makes location errors, to avoid that we use OYOLO. Paper [4] presents a prototype that extracts the text from image and is converted to speech. Extraction of text is done by using the Tesseract Optical Character Recognition (OCR). This method is carried out by using Raspberry Pi. Text recognition is done by using Open Computer Vision (Open CV), considering the large library of functions when compared to MATLAB. Capturing of image is done by using a portable camera and the image is converted to gray scale and filtered by Gaussian filtering. Then it is binarized and cropped. The cropped image is given to Tesseract OCR. The e-Speak creates an analog signal of the text and is given to the headset.

Object detection using machine learning for visually impaired people Networks (RPN), which involves subsampling to obtain the image information. The Resort with 152 samples achieved an average precision with 83.1%, and Dense Net with 121 samples achieved an average precision with 79.8%.

Dr. K. Sreenivasulu, P. kiran Rao have proposed their work on "A Comparative Review on Object Detection System for Visually Impaired" in 2016. This model is used for detecting the patterns in urban areas such as public streets, raining, restaurants, etc.13 This method characterizes the audio clips, which yields the patterns. The main limitation of this model is to require a trained data set. 6

III. METHODOLOGY

The aim of our project is to design an affordable system that enables individuals with impaired vision to lead a normal life without depending on others in environments such as their home, workplace, or other selected areas. The system uses a webcam to detect objects involved in day-to-day activities. The process flow of the system is depicted in Figure 1.

Object detection, recognition, and communication speed are crucial factors in the system's design. Since this system is specifically intended for visually impaired users, the response time plays a pivotal role. Any delay in communicating information about obstacles or surroundings would defeat the purpose of the system. To address the challenges of processing speed and minimize delays, the YOLO V5 algorithm is utilized in this system. YOLO V5 offers improved performance and is faster than many other real-time object detection algorithms, ensuring quick and accurate communication to the user.

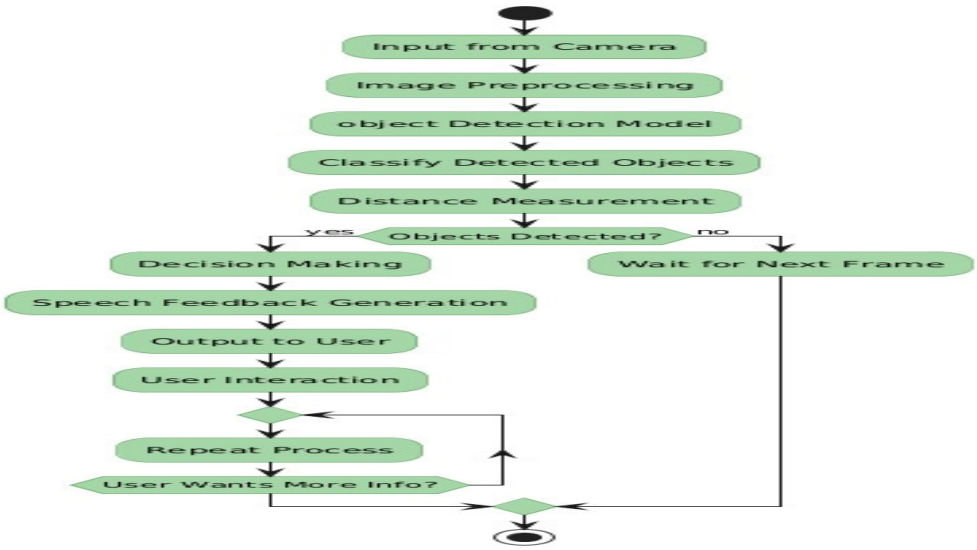


Fig no. 01: Flow chart

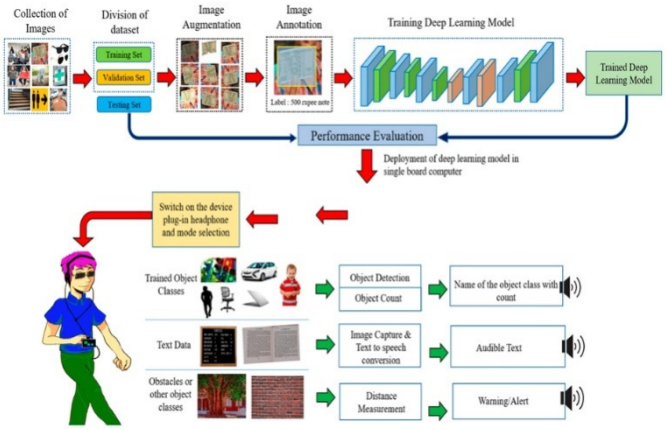


Fig no. 02: System Architecture

IV. YOLO ALGORITHM

YOLO algorithm [6],[8] is primarily used for the prediction of bounding boxes accurately from an image. Images are divided into N x N grids and for each grid the prediction of the bounding boxes are done as well as the class probabilities. After performing object localization and image classification for each grid of the input image, each grid is given a different label.

The following project is designed as a YOLO algorithm applied separately on each grid and the objects in it are marked with their particular label and corresponding bounding boxes are also highlighted. The grids having no object are labeled as 0. Initially, YOLO algorithm is applied to the received image. In our project, the real time image is divided into grids of matrices. As the image complexity varies the image can be split into any number of grids. After division of the images, both classification and the localization process are done on each grid containing the object. The confidence score is computed for all the grids. The confidence score and also the bounding box for each of the grid will differ based on whether the object is detected or not. For no object it displays the value as 0 and displays the value 1 if object exist. Bounding box value will show how confident the network is, that is, how much the detected object matches to the object under observation.

V. PROPOSED SYSTEM

Visually impaired persons face many challenges in identifying objects and performing the day to day activities. Most of the time they are dependent on others in moving from one place to other. They face a lot of difficulty especially when moving in outdoor environment, where objects are continuously moving. This object detection system would assist the visually impaired persons by providing the perception of the surrounding environment or position of the objects. This device can assist visually impaired persons in avoiding the obstacles in both indoor and outdoor environment. The proposed system also helps the user in identifying the objects around them. It would minimize the visually impaired person's difficulties and help them lead a quality life.

The proposed system aims to be simple, user-friendly, handy, economical and efficient solution. Most of the existing assistive systems are highly sophisticated and expensive. Smartphones are widely used these days and the usage of smartphones by all persons has become very usual in the recent past. Therefore this system uses all the advanced built-in features of a smartphone. Smartphone's integrated camera is used to capture the real time video to detect the objects present and headphones or smartphone's speaker is used to communicate to the user through audio instructions. The proposed application can easily be accessed by the visually impaired persons.

The Yolo V3 (You only look once) algorithm is used to detect the real time objects captured from the continuous streaming by the smartphone camera. It is considered as one of the most powerful real time object detector algorithms. Unlike other algorithms like R-CNN and Faster R-CNN which examine several regions of the image to identify objects, Yolo passes image or video only one time through its network and uses a unique neural network using the characteristics of the complete image to predict multiple boxes containing an object. This is a significant feature in Yolo, which reduces the processing speed when compared with other algorithms. Processing speed plays a key role particularly while detecting objects in real time video stream. OpenCV provides a function that facilitates image pre-processing for deep-learning classification. Group of connected pixels in an image that share common properties (Blob ex: Grayscale value) of each frame captured is identified. Bounding boxes are created for each object identified and based on the Yolo V3 pre-trained weights, confidence score and coco dataset each object is processed and labelled.

The coco dataset contains all the objects or class names on which the model is trained. Non-maximum suppression (NMS) uses function called “Intersection over Union (IoU)” which is used to determine the best boxes. In order to select the best box, NMS follows three steps. 1. It selects the box with highest score. 2. Computes the overlap with other boxes, removes the overlap which has more than a certain

VI. EXPERIMENTAL PROCEDURE

A. *Collection of Data Base*

The first step in developing any object detection system is to gather a dataset. For this system, a database was compiled containing images of the objects to be detected. In this case, the system needs to recognize five classes of objects: bottles, watches, keys, pens, and glasses. Each class consists of a large set of labeled images, each showing the object from different angles, lighting conditions, and environments.

Process:

1) Data Acquisition: The dataset could be sourced from existing image datasets or manually captured using a camera. Images for each class (bottles, watches, keys, pens, and glasses) should be diverse enough to train the model to handle variations in real-world settings.
2) Data Labeling: Images need to be labeled with annotations that define the objects' bounding boxes. This means drawing rectangles around each object in the image and assigning the correct class label (e.g., bottle, watch, etc.).
3) Data Augmentation: To increase the model's robustness, data augmentation techniques like rotating, flipping, scaling, or adjusting brightness may be used to artificially increase the size of the dataset.

B. *Network Output Analysis*

Once the YOLO network is trained, the next step is analyzing its output. The YOLO (You Only Look Once) algorithm provides predictions in the form of bounding boxes around detected objects, along with the corresponding class labels and confidence scores.

Key Steps:

1) Bounding Boxes: For each detected object, YOLO provides a rectangular box around it, with coordinates indicating the location of the object in the image.
2) Class Labels: YOLO outputs the class labels for the detected objects, such as "bottle," "watch," etc.
3) Confidence Scores: For each detection, a confidence score (ranging from 0 to 1) indicates how confident the model is in its prediction.

The output analysis involves reviewing how well the network is detecting the objects, checking for any missed detections, false positives (incorrectly detected objects), and evaluating its performance in various scenarios (e.g., cluttered backgrounds or poor lighting).

C. *Bounding Box Dimensions*

Bounding boxes are a crucial part of object detection systems. The bounding box provides the coordinates of the detected object within an image. These coordinates are typically represented by the following:

1) x, y coordinates: These define the top-left corner of the bounding box.
2) Width and Height: These define the size of the box.

YOLO uses these bounding box dimensions to locate and classify objects. The network's accuracy in predicting these bounding box dimensions is critical for the system’s performance. An incorrect bounding box could result in improper detection of the object, causing the system to provide inaccurate information to the user. The goal is to have tight bounding boxes that accurately cover the object while minimizing any background or unnecessary area.

D. *Output Filtering*

After YOLO detects objects, there may be multiple predictions for the same object or incorrect detections in areas without objects (false positives). Output filtering is a technique to refine the results by removing these inaccuracies.

Key methods for output filtering include:

1) Non-Maximum Suppression (NMS): This technique helps remove duplicate detections of the same object. When multiple bounding boxes overlap, NMS keeps only the box with the highest confidence score and removes others.

2) Confidence Threshold: Only detections with a confidence score above a specified threshold are kept. For example, if the confidence score is lower than 0.5, the detection is discarded.
3) Class Filtering: In some cases, the system might be trained to detect a wide range of classes, but filtering could be applied to keep only the relevant classes (e.g., bottles, watches, etc.).

This output filtering ensures that the system only provides valid and accurate detections for further processing.

E. Training

The training process involves teaching the YOLO network to detect the objects in the dataset. This is done by feeding the labeled data into the model and adjusting its parameters (weights) over several iterations.

Key steps in the training process:

1) Data Preprocessing: Before training, images are resized, normalized, and converted into a format suitable for YOLO. The labels for each image are also processed into the YOLO format.
2) Training the Model: During training, the model tries to minimize the loss function, which measures how far the predicted bounding boxes, class labels, and confidence scores are from the actual values. YOLO uses a combination of regression (for box coordinates) and classification (for object categories) to improve its predictions.
3) Validation and Testing: After training, the model is validated using a separate dataset to ensure it generalizes well to new, unseen data. Hyperparameters such as learning rate and batch size are tuned to optimize performance.

F. Text to Speech

Once the objects are detected, the next step is to convert the detection results into a form that the visually impaired user can understand. This is where the text-to-speech (TTS) module comes into play.

Process:

1) Detection Data to Text: After an object is detected and classified by the YOLO model, the class label and confidence score are converted into text. For example, "Object: Bottle, Confidence: 95%."
2) Text to Speech Conversion: This text is then passed to a text-to-speech engine (like Google TTS or other similar systems), which converts the text into an audio message. This audio message is then played back to the user, informing them about the detected object and its confidence level.

For instance, if a bottle is detected, the system would output an audio message like "Bottle detected, 95% confidence." This provides real-time feedback to the visually impaired user, helping them navigate their environment.

VII. EXPECTED OUTCOMES

1) *Accurate Object Detection*

- The YOLO V5 algorithm should successfully detect and classify the five predefined objects (bottles, watches, keys, pens, and glasses) with high accuracy.
- The bounding boxes around the detected objects should align closely with the actual objects in the images, minimizing false positives and false negatives.
- The system should be capable of distinguishing between objects from different classes (e.g., a bottle from a pen) in a variety of environments (indoor, outdoor, varied lighting conditions, and cluttered backgrounds).

2) *Low Response Time*

- The object detection and recognition process should happen quickly, with minimal delay between object detection and the audio output.
- The system should be responsive enough to provide real-time feedback to the user. The time from detecting an object to providing audio feedback should be low enough to ensure that users receive timely information.

3) *Robustness in Object Recognition*

- The system should handle various scenarios, such as different object orientations, occlusions (objects partially blocked), and environmental noise, without significantly degrading performance.
- It should also work well in varying lighting conditions, such as bright sunlight or low-light environments, by maintaining the accuracy of object detection.

4) *Accurate Audio Feedback*

- The text-to-speech (TTS) system should deliver clear and understandable audio feedback to the user, providing information such as the object detected ("Bottle detected," "Pen detected," etc.) along with the confidence level.
- The audio output should be immediate and appropriately timed, allowing the visually impaired user to react to the detected obstacles or objects.
- The TTS should also be able to handle multiple objects and provide feedback for each detected object in real-time.

5) *Obstacle Detection*

- The system should successfully detect obstacles (e.g., objects in the user's path) and alert the user with an appropriate audio warning. This could include alerting users about obstacles that are close (within a set range, like 4 cm) and providing directions on how to avoid them.

6) *User-Friendliness*

- The device should be easy to use for visually impaired individuals. The interface should be intuitive, and the device should provide clear audio instructions to help users navigate their environment.
- The system should provide feedback on objects in the user's surroundings, including any potential hazards or obstacles, to ensure safe navigation.

7) *Cost-Effective and Portable*

- The system should be affordable, utilizing commercially available hardware (e.g., webcams, microcontrollers, sensors) and software libraries (e.g., TensorFlow, YOLO V5) to keep costs low.
- The device should be portable, compact, and lightweight, enabling visually impaired users to carry it with ease while moving around.

8) *Reliability*

- The system should be reliable and able to function in real-world environments without frequent crashes or performance degradation.
- The object detection and speech feedback should be consistent over time and in various environmental conditions.

9) *Enhanced Independence for Visually Impaired Users*

- Ultimately, the expected outcome is that the visually impaired users will be able to navigate their environments more independently, safely, and confidently, without needing assistance from others for routine tasks like walking down the street or finding objects.
- The system should reduce dependency on others, allowing users to make informed decisions about their surroundings based on the real-time feedback provided by the system.

VIII. RESULTS

A. *Home Screen*

The home screen is the main screen of an app or system that users typically see when they first open the app. It serves as a hub or starting point, providing easy access to all key features and functionalities.

In the context of your app or system, the home screen might include:

- Navigation menus (to access different sections like Object Detection or Location Pages)
- Search bar (to quickly find specific objects or locations)
- Quick links or shortcuts to frequently used features or tools.

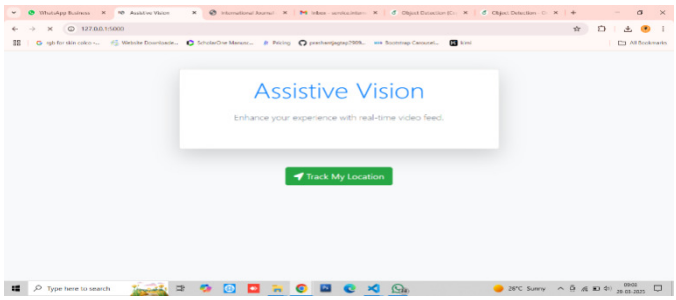


Figure 1: Home Screen

B. Object Detection Page

The Object Detection Page allows users to not only upload images or videos but also use a live camera feed to detect and identify objects in real-time. Here's how this page could be structured:

- Live Camera Feed Interface: Real-time object detection that activates the device's camera, immediately processing the live feed.
- Overlay of Detected Objects: Detected objects are highlighted with bounding boxes, labels, and different colors (e.g., "car," "dog," "person").
- Dynamic Updates: The detection model continuously updates as new objects appear in the frame.

Object Detection with Bounding Box 1:

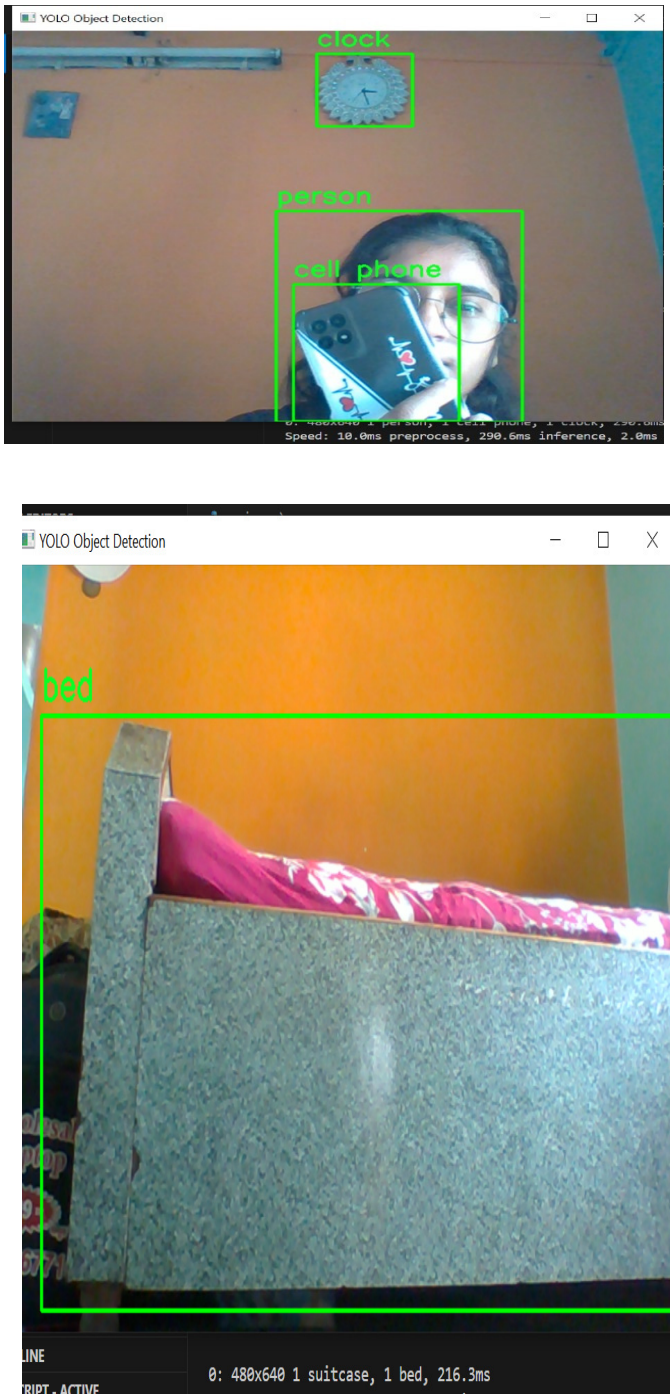


Figure 2: Object Detection with Bounding Box 1

C. Location Page

The Location Page displays geographical or map-based information. It includes the following features:

- Maps: A map interface where users can view their location or a specific address.
- Location Search Bar: A search bar for users to input a location by name or coordinates.
- Pins/Markers on the Map: Markers to highlight key places, objects, or events on the map.
- Details: Information about the selected location, such as coordinates, directions, descriptions, or related data.

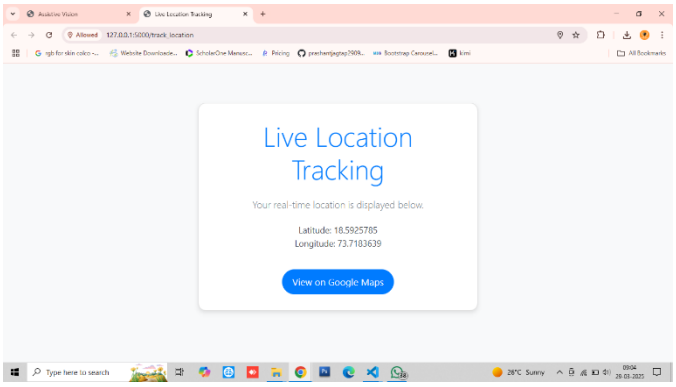


Figure 3: Location Page

IX. CONCLUSIONS

In recent years, there were many assistive solutions developed for the visually impaired persons to provide assistance in detecting objects present in their surrounding environment and in moving from one place to another. But most of the existing solutions are expensive, highly sophisticated, difficult to handle, designed as a dedicated aid, require training etc. Our primary goal is to provide an assistive solution to the visually impaired persons which is simple, user-friendly, affordable and handy and assist the visually impaired persons in understanding the surrounding environment and help them in moving from one place to another place independently. The real time video stream is captured by the smartphone camera and then the label of the object detected using object detection algorithm is communicated to the user in speech through the speakers or headphones. The audio output communicated to the user would help them in performing their day to day activities and lead a quality life

REFERENCES

[1] Chen X, Yuille AL. A time-efficient cascade for real-time object detection: With applications for the visually impaired. In2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops 2005 Sep 21:28-28.

[2] Chi-Sheng, Hsieh. "Electronic talking stick for the blind." U.S. Patent No. 5,097,856, 24 Mar. 1992. [3] WafaMElmannai, KhaledM.Elleithy. "A Highly Accurate and Reliable Data Fusion Framework for Guiding the Visually Impaired". IEEE Access 6 (2018) :33029-33054. [1]

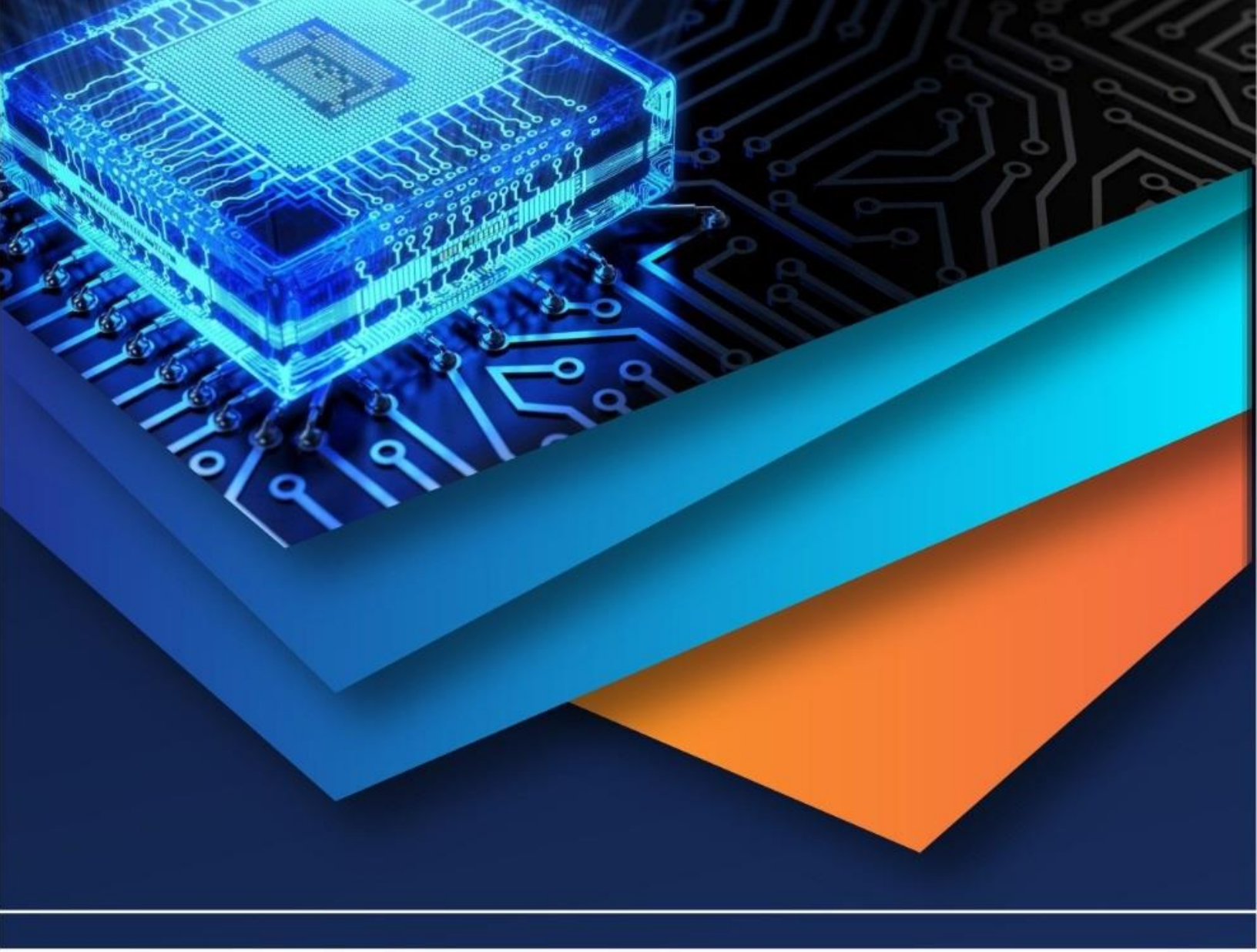
[3] Ifukube, T., Sasaki, T., Peng, C., 1991. A blind mobility aid modelled after echolocation of bats, IEEE Transactions on Biomedical Engineering 38, pp. 461 - 465.

[4] Cantoni, V., Lombardi, L., Porta, M., Sicard, N., 2001. Vanishing Point Detection: Representation Analysis and New Approaches, 11th International Conference on Image Analysis and Processing.

[5] Balakrishnan, G. N. R. Y. S., Sainarayanan, G., 2006. A Stereo Image Processing System for Visually Impaired, International Journal of Information and Communication Engineering 2, pp. 136 145.

[6] C.S. Kher, Y.A. Dabhade, S. sK Kadam., S.D.Dhamdhere and A.V. Deshpande "An Intelligent Walking Stick for the Blind." International Journal of Engineering Research and General Science, vol. 3, number 1, pp. 1057-1062

[7] G. Prasanthi and P. Tejaswitha "Sensor Assisted Stick for the Blind People." Transactions on Engineering and Sciences, vol. 3, number 1, pp. 12-16, 2015



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY