CNN-Based Object Recognition And Tracking System To Assist Visually Impaired People

Mrs. S.Santhana Prabha[1], Mohammed ismail mydeen A[2], Pradeepkumar C[3], Prasanna R[4], Prasannavenkatesh R[5]
[1]Assistantprofessor, Dept of Computer Science and Engineering
[2, 3, 4, 5]Dept of Computer Science and Engineering
[1, 2, 3, 4, 5]PSNA College of Engineering and Technology
Dindigul, TamilNādu, India.

Abstract- *Having good vision is a wonderful gift, yet it is disheartening that more and more people are experiencing vision loss these days. To assist visually impaired individuals, the visual world can be transformed into an audio world that can provide them with information about objects. By detecting objects within an image and converting their names into speech, visually impaired individuals can live independently. Technology has made significant advancements in the past few decades that enable individuals with disabilities to have more control over their lives. This project proposes an assistive system for the visually impaired that utilizes YOLO, a deep neural network-based technique, to quickly detect objects within images with high accuracy. The system also employs eSpeak, which is given by Open CV under Python, to provide auditory details. Our proposed model successfully enables visually impaired users to navigate unfamiliar indoor and outdoor environments with the assistance of a user interface. This project utilizes artificial intelligence to detect and analyze objects, which are then conveyed via speaker to the user. To enhance recognition time and achieve better time complexities, the proposed system uses dark net YOLO techniques to recognize multiple objects in a shorter period of time.*

Keywords- CNN(Convolutional Neural Network), YOLO(You only look once), OpenCV, DARKNET, Gray- Level Co-Occurrence Matrix (GLCM), Wiener filter.

I. INTRODUCTION

The visually impaired face a significant challenge with safe mobility on a daily basis, often unable to detect and avoid obstacles in their path. This not only causes emotional distress and undermines their independence but also exposes them to the risk of injuries. According to recent statistics by the World Health Organization (WHO), around 253 million people worldwide suffer from visual impairment, out of which 36 million are blind, and the number is expected to double by 2020. The need for assistance in performing daily tasks, such as navigating and exploring unfamiliar environments, is essential for those with vision impairments and eye ailments. However, despite the alternative approaches developed by them, they face certain difficulties in navigation and social awkwardness. For instance, they struggle to locate a specific room in an unfamiliar environment, and during a conversation, it is challenging to determine if someone is addressing them or someone else. In this paper, we explore the possibility of using the hearing sense to understand visual objects since both the sense of sight and hearing share a similarity in spatial localization. Our objective is to guide the visually impaired through a processor output for safe navigation. Our methodology involves Object Extraction, Feature Extraction, and Object Comparison to achieve this aim.

The development of supporting devices or assistants for individuals with visual impairment is an active area of research, primarily focused on the navigation problem. Currently, some navigation devices and systems are available, such as seeing-eye dogs and white canes. However, their performance is limited in terms of speed, coverage, and ability, which are generally available to individuals with normal vision. For instance, a cane can only detect obstacles at knee level and has a short range of detection. Moreover, seeing-eye dogs require intensive training and are costly, making them inaccessible to most individuals. Our project is motivated by the need for navigation assistance among the blind and visually impaired individuals and the availability of advanced technology to solve this problem. Our goal is to use technology to ease tasks for the visually impaired and help them navigate more safely and independently. Our engineering profession drives us to utilize available technology to improve the lives of individuals with disabilities.

II. LITERATURESURVEY

Rabia Jafri, et al. [2] have presented a novel system that utilizes visual and infrared sensor data to assist visually impaired individuals in detecting obstacles while navigating indoors. The system is developed for the Google Project Tango Tablet Development Kit, which is equipped with powerful graphics processing capabilities and multiple sensors that allow for real-time tracking of motion and orientation in 3D space. It leverages the built-in functionalities of the Unity engine in the Tango SDK to create a 3D reconstruction of the surrounding environment and associates a Unity collider component with the user to detect obstacles based on their interaction with the reconstructed mesh. The system provides audio feedback in the form of obstacle warnings, and empirical evaluations have shown promising results, highlighting its potential for future development.

R.R. Bourne et al. [3] proposed a systematic review and meta- analysis to estimate the global prevalence, temporal trends, and projections of blindness and vision impairment. Using population- based datasets published between 1980 and 2015, hierarchical models were fitted to estimate the prevalence of mild visual impairment, moderate to severe visual impairment, blindness, and functional presbyopia by age, country, and sex for the year 2015. This study provides important global estimates that can inform public health policies related to vision impairment and blindness.

E. Cardillo et al. [4] demonstrated the feasibility of an electromagnetic sensor prototype to assist visually impaired and blind individuals in autonomous walking. The proposed system utilizes a microwave radar integrated into a traditional white cane to provide awareness of obstacles in a wider and safer range. Compared to existing Electronic Travel Aids devices, the system exhibits better performance, noise tolerance, and reduced dimensions. The authors also discuss the latest developments in the miniaturization of circuit boards and antennas, and present the results of obstacle detection tests to showcase the effectiveness of the system

A.Riazi et al.,[5] conducted a qualitative study using semi- structured individual interviews to identify common outdoor difficulties faced by visually impaired individuals in Iran. The study included 20 legally-blind participants with different etiologies. The results revealed various challenges such as installation of tactile ground surface indicators, unsafe sidewalks, obstacles on sidewalks, difficulty reading bus numbers, disorientation, fear of falling, recognition of faces, inability to read street names, and other barriers in outdoor environments.

S.Mahmud et al.,[7] conducted a review of different types of interfacing techniques between humans and machines in the field of human machine interaction. The review explored contemporary technologies developed for this purpose, their advantages, and limitations. The authors outlined the current status and future development perspectives of interfaces for assistant devices in the field of human machine interaction, human computer interaction, and human robot interaction.

N.Sahoo et al.,[9] proposed a design and implementation of a walking stick aid for visually challenged people (VCPs). The system includes a walking stick embedded with programmable interface controller (PIC) and Raspberry Pi, a global position system (GPS) module, alert-providing components and sensors. The sensors detect obstacles and the VCP is alerted through vibrations or a buzzer. The GPS module allows location tracking by parents using an app. An emergency app also allows the VCP to communicate with parents or friends in panic situations. The system is designed to be simple, lightweight, and user-friendly with useful features for visually challenged individuals.

H.Zhang et al. [10] proposed an innovative Indoor Wayfinding System that utilizes Geometric Features Aided Graph SLAM to assist the visually impaired with navigation. The system incorporates a 6-DOF pose estimation (PE) method, which involves two graph SLAM processes to reduce the cumulative pose error of the device. In the first step, the floor plane is extracted from the 3D camera's point cloud and added as a landmark node into the graph for 6-DOF SLAM, resulting in reduced roll, pitch, and Z errors. In the second step, the wall lines are extracted and incorporated into the graph for 3-DOF SLAM, leading to reduced X, Y, and yaw errors. This approach results in more accurate pose estimation with lower computational time compared to existing planar SLAM methods. Based on the PE method, the researchers developed a wayfinding system that assists visually impaired individuals in navigating indoor environments. The system utilizes the estimated pose and floor plan to locate the user in a building and provides guidance through a speech interface, announcing points of interest and navigational commands. Experimental results have validated the effectiveness of the PE method and demonstrated the potential of the system to greatly facilitate indoor navigation tasks for the visually impaired.

In a related study, A.J. Ramadhan [12] presented a wearable smart system designed to assist visually impaired individuals with independent navigation in public places. The system comprises a microcontroller board, various sensors, cellular communication and GPS modules, and a solar panel. The sensors track the user's path and provide alerts through a buzzer and wrist vibrations to notify the user of obstacles ahead, which is particularly helpful for those with hearing loss or in noisy environments. Additionally, the system alerts people in the surroundings when the user stumbles or requires assistance, and sends alerts along with the system's location to registered mobile phones of family members and caregivers.

The registered phones can also be used to retrieve the system's location and activate real-time tracking of the visually impaired individual. The functionality and effectiveness of the system prototype were verified through testing, and it is expected to be a valuable tool in improving the quality of life for visually impaired individuals due to its advanced features compared to similar systems.

In conclusion, both the Indoor Wayfinding System based on Geometric Features Aided Graph SLAM and the Wearable Smart System for Visually Impaired People are innovative technologies that have the potential to greatly assist visually impaired individuals with navigation in indoor and outdoor environments, respectively. These systems utilize advanced sensors, algorithms, and communication modules to provide real-time alerts and guidance, ultimately improving the independence and quality of life for individuals with visual impairments.

III. EXISTING SYSTEM

Limited accuracy: The system's output accuracy is low, which means that it may not be able to detect obstacles or potholes with sufficient reliability to keep blind people safe.

Simulation-based: The system is modeled in simulation, which means that it has not been implemented in real-time. This limitation makes it challenging to test the system's effectiveness in real-world scenarios.

Lack of integration: The system seems to use multiple technologies, including Ultrasonic Sensors, IR Sensors, and OpenCV based object detection. However, there is no indication of how well these technologies integrate with each other, which could lead to lessreliable results.

Limited functionality: While the system can detect obstacles, it is unable to detect potholes on the road surface. This limitation could beparticularly hazardous for blind people who are relying on the systemto navigate.

Overall, it appears that the existing system has several significant drawbacks that limit its effectiveness in aiding blind people to navigate safely. To improve the system, developers may need to address these issues by improving accuracy, integrating technologies more effectively, and expanding the system's functionality to include the detection of potholes on the road surface.

PROPOSED SYSTEM:

Our proposed system aims to achieve real- time object detection using advanced technologies such as AI, OpenCV, and YOLO (You Only Look Once). Upon acquiring an image, it undergoes pre-processing and compression. In order to train the model, we utilize a diverse set of images featuring common objects that we encounter in our daily lives. The model is trained through feature extraction, which is performed on the image to obtain the necessary patterns. Feature fusion and dimension reduction are then carried out to compress the image, ensuring efficient and reliable performance in real-time. We use the YOLO dataset to train the classifier, selecting the best performing classifier after comparing various options. Through this process, we develop a highly accurate object recognition model. Any test image can be fed into this model, and it will classify the image into one of the categories it has been trained to identify.

Image processor

An image processing system performs a variety of functions including image acquisition, storage, preprocessing, segmentation, representation, recognition, and interpretation before displaying or recording the final image. The system begins with image acquisition by an imaging sensor and digitizer, which converts the image into a digital format. Preprocessing is then applied to improve the quality of the image and prepare it for further processing. Preprocessing typically involves tasks such as noise reduction and region isolation. Segmentation is the process of dividing the image into constituent parts or objects, producing either boundary or pixel data. Representation transforms raw pixel data into a form suitable for subsequent processing. Description extracts features that differentiate object classes, while recognition assigns a label to an object based on its descriptors. Finally, interpretation assigns meaning to recognized objects using a knowledge base that guides the operation of processing modules and controls interaction between them. The image processing system's composition varies based on its intended application, and its frame rate is typically around 25 frames per second.

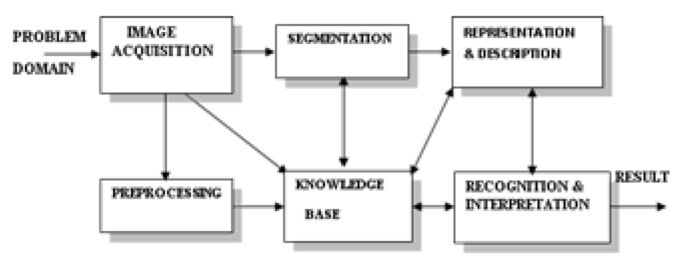

Figure 1 Block Diagram Of Fundamental Sequence Involved In An Image Processing System

Image Preprocessing:

Image preprocessing is a crucial step in image processing, as it deals with modifying the input image to match the requirements of the process. The input image may have varying sizes, noise, and color combinations that need to be adjusted accordingly. Noise in images can be quite apparent, especially in regions with low signal levels, such as shadow regions or under-exposed images. Filtering algorithms, such as the Wiener filter, are commonly used to remove noise. In the preprocessing module, the acquired image is processed to ensure correct output. This involves using various algorithms to improve the image's quality before further processing.

Feature Extraction:

Feature extraction is the process of identifying unique features from an image. These features are then used to classify the image into different categories or classes. Statistical features, such as mean, variance, skewness, and standard deviation, are commonly used in image feature extraction. Another approach to texture analysis is using the Gray-Level Co-Occurrence Matrix (GLCM). This statistical method examines texture by considering the spatial relationship of pixels.

Classification

Classification is the process of assigning an image to a specific class or category. This process requires a well-understood relationship between the data and the classes. In order to achieve this, computers must be trained to understand the relationship. The success of classification techniques heavily depends on training. The image classifier discriminates one class against others, where the discriminant value is highest for one class and lower for other classes in multiclass. In a two-class scenario, the discriminant value is positive for one class and negative for the other. By using the right techniques for image preprocessing, feature extraction, and classification, high-accuracy image recognition can be achieved.

ADVANTAGE

One of the key advantages of using YOLO for object detection is its speed. YOLO is known for being extremely fast compared to other object detection algorithms. Additionally, YOLO's accuracy is high, making it an attractive choice for various applications that require real-time object detection. Another advantage of YOLO is its ease of implementation on various processors with its dependencies. This means that it can be integrated into different systems and platforms without too much difficulty. Finally, YOLO is a cost-effective solution compared to other object detection techniques, making it accessible to a wider range of users and applications.

SYSTEM DESIGN:

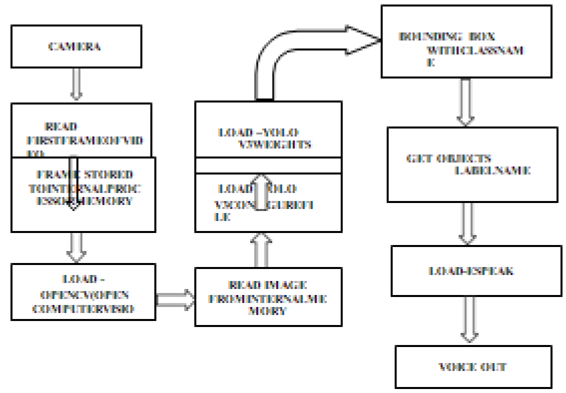

Figure 2 SYSTEM DESIGN

IV. RESULTS AND DISCUSSION

YOLO

You Only Look Once (YOLO) is a Deep Learning (DL) network used for object detection that employs an innovative approach to identify objects within an image. Instead of using multi-step pipelines, YOLO performs object detection by classifying the objects and their locations in one pass, generating a vector of bounding boxes for each object. This is different from previous methods such as Region-Convolution Neural Networks (R-CNN) and its variations, which required the training of individual components separately and are thus slower (47 seconds per individual test image) and harder to optimize. YOLO, on the other hand, is much faster (45 frames per second) and easier to optimize, thanks to its reliance on a single neural network that handles all components of the task. YOLO's architecture comprises three main parts: the algorithm (predictions vector), the network, and the loss functions. By using YOLO, object detection can be achieved accurately and efficiently, making it an ideal choice for real-time applications. YOLO, or You Only Look Once, is a popular object detection network that uses deep learning algorithms to detect objects in an image. Unlike previous object detection methods, such as Region-Convolution Neural Networks (R-CNN), which perform object detection in a pipeline of multi-step series, YOLO takes a different approach. Instead of focusing on a specific region within the image and training each individual component separately, YOLO classifies certain objects in the image and determines where they are located by generating a vector of bounding boxes for each object. This approach is much faster and easier to optimize than previous algorithms. For example, R-CNN requires the network to classify 2000 regions per image, making it very time-consuming and not suitable for real-time implementation. Additionally, R-CNN uses a fixed selective algorithm, which may generate inferior region proposals. YOLO, on the other hand, uses a single neural network to run all components of the task, making it much faster and easier to optimize. Its architecture includes an algorithm or predictions vector, a network, and loss function

The YOLO Algorithm

The YOLO Algorithm is a crucial part of object detection using deep learning algorithms. It uses a grid of size SxS to predict the object's bounding box and its class within an image. To understand how the algorithm functions, we need to comprehend how it specifies and builds each bounding box. The YOLO algorithm uses four components:
the center of a bounding box (bx by)

- width (bw)
- height (bh)
- class of the object (c)

Additionally, it predicts confidence (pc), which represents the probability of the existence of an object within the bounding box. The (x,y) coordinates denote the center of the bounding box. However, most of the bounding boxes will not contain an object. Hence, we need to use the pc prediction and a process called non-max suppression to remove unnecessary boxes with low probability and those that share a large area with other boxes

The Network

The YOLO network follows a typical CNN structure with convolution and max-pooling layers, as well as two fully connected CNN layers. However, since the YOLO algorithm predicts multiple bounding boxes for each grid cell, we only want one bounding box to be responsible for the object in the image. To achieve this, a loss function is used to compute the loss for each true positive. By selecting the bounding box with the highest Intersection over Union (IoU) with the ground truth, the loss function becomes more efficient and specialized bounding boxes are created.

YOLO V3

YOLO V3 is an improved version of YOLO V2, which uses another variant of Dark net. The YOLO V3 architecture comprises 106 layers, with 53 layers trained on ImageNet and another 53 tasked with object detection. While this has significantly increased the network's accuracy, it has also reduced the speed from 45 fps to 30 fps

V. CONCLUSION

Our work is mainly to address the day-to-day challenges of visually impaired individuals by developing a solution through object detection methods. After exploring various object detection techniques, we opted the OpenCV Library and YOLO due to their effectiveness. Our proposed system for visually impaired individuals relies on object recognition in frames using YOLO's configuration, weights, and feature matching for accurate object identification. We conducted various experiments to evaluate the application's performance in detecting objects under different conditions. Overall, the proposed system provides a novel and useful approach to address the challenges faced by visually impaired individuals.

OUTPUTS SCREENSHOT:

Figure 3 Bottle Image

Figure4 Mouse Image

Figure 5 Person Image

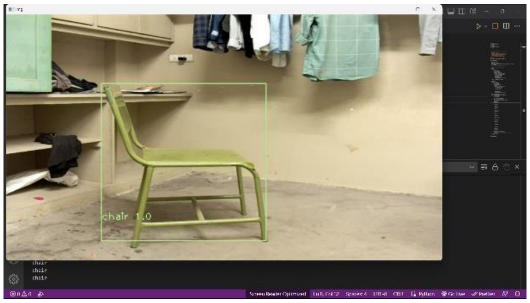

Figure6 Chair Image

Figure 7 Knife Image

Figure 8 Laptop Image

Figure 9 Cellphone Image

Figure 10 Toothbrush Image

Figure 11 Cup Image

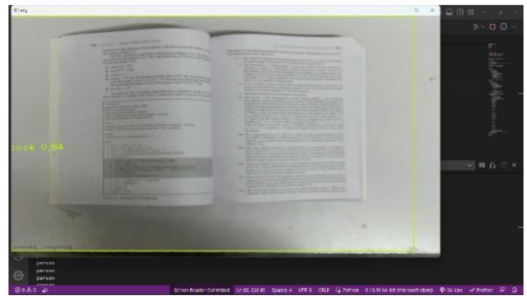

Figure12 Book Image

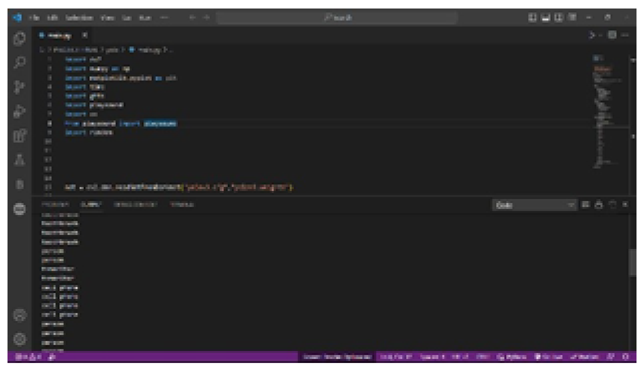

Figure 13 Python shell shows images name

REFERENCES

[1] V. V. Meshram, K. Patil, V. A. Meshram and F. C. Shu, "An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People," in IEEE Transactions on Human- Machine Systems, vol. 49, no. 5, pp. 449-460, 2019.

[2] R. Jafri, R. L. Campos, S. A. Ali, and H. R. Arabnia, "Visual and Infrared Sensor Data-Based Obstacle Detection for the Visually Impaired Using the Google Project Tango Tablet Development Kit and the Unity Engine," IEEE Access, vol. 6, pp. 443-454, 2018.

[3] R. R. Bourne et al., "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta- analysis," The Lancet Global Health, vol. 5, no. 9, pp. e888- e897, 2017.

[4] E. Cardillo et al., "An Electromagnetic Sensor Prototype to Assist Visually Impaired and Blind People in Autonomous Walking," in IEEE Sensors Journal, vol. 18, no. 6, pp. 2568- 2576, 2018.

[5] A. Riazi, F. Riazi, R. Yoosfi, and F. Bahmeei, "Outdoor difficulties experienced by a group of visually impaired Iranian people," Journal of current ophthalmology, vol. 28, no. 2, pp. 85-90, 2016.

[6] M. M. Islam, M. S. Sadi, K. Z. Zamli and M. M. Ahmed, IEEE Sensors Journal, vol. 18, no. 6, pp. 2568-2576, 2018."Developing Walking Assistants for Visually Impaired People: A Review," in IEEE Sensors Journal, vol. 19, no. 8, pp. 2814-2828, 2019.

[7] S. Mahmud, X. Lin and J. Kim, "Interface for Human Machine Interaction for assistant devices: A Review," in 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 2020, pp. 0768-0773.

[8] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," Pattern Recognition Letters, 2018.

[9] N. Sahoo, H.-W. Lin, and Y.-H. Chang, "Design and Implementation of a Walking Stick Aid for Visually Challenged People," Sensors, vol. 19, no. 1, 2019.

[10] H. Zhang and C. Ye, "An Indoor Wayfinding System Based on Geometric Features Aided Graph SLAM for the Visually Impaired," IE

[11] Rabia Jafri, Rodrigo Louzada Campos, " Visual and Infrared Sensor Data-Based Obstacle Detection for the Visually Impaired Using the Google Project Tango Tablet Development Kit and the Unity Engine "in year of 2019