Object Detection for Assisting Blind Individuals Using ESP32-CAM and GPS

1st Anupam Kumar
BE CSE
Chandigarh University
Mohali, Chandigarh
anupambhumi@gmail.com

2nd Shashi Ranjan Mehta
BE CSE
Chandigarh University
Mohali, Chandigarh
itsmeshashi29@gmail.com

3rdth Jasneet Kaur
BE CSE
Chandigarh University
Mohali, Punjab
er.jasneetchawla@gmail.com

4th Satyam Bharti
BE CSE
Chandigarh University
Mohali, Punjab
tsatyams449@gmail.com

***Abstract*—The purpose of this study is to develop a conceptual framework of assistive technologies that will help meet the objective of increasing independence and safety of visually impaired persons. This framework is based on recent innovations in object detection, depth sensing, and communication for real-time spatial awareness and response. To increase efficiency in gathering the environment data the system combines several hardware components such as high frame rate web camera, infrared depth sensor as well as ESP32-CAM microcontroller board that plays a crucial role in capturing visions and spatial data. The object detection feature, which forms the basis of the system, uses the YOLO algorithm; a prominent algorithm in real-time object recognition and localization that allows the framework to quickly identify and locate objects in the user's environment.**

From the software side, the complex sensor inclusion and data treatment are maintained with ordinary C++ of the Arduino-based hardware. The companion application is built with React Native and is best suited to Android platforms with additional nice features of its user-friendly interface. These are major features of this application which include notification or warning to the user of the obstacles, landmarks or areas of interest around the environment. Also, the system under development has incorporated real-time GPS tracking so that the attendant or family members can have an idea of where the user is at any given time, giving more security.

Featuring an innovative compact and wheelable design, this mobility solution aims to decrease the level of reliance of the visually impaired individual on mechanized aids or human guidance /assistance with activities such as walking and/or getting around during daily activities/ in unknown territories. Thus this work, brings the advanced sensory and notification technologies in one small pocket device which is actual and practical invention solely serving to enhance the lives of visually impaired citizens [1]. Not only does this innovative approach thus improve individual liberty but it also demonstrates how assistive technologies can deliver better options for mobility for people with vision impairments, from a safety perspective.

I. INTRODUCTION

The largest problem with students who are blind or visually impaired is the lack of knowledge about how to get around alone and safely. Traditional mobility instruments such as Guide dogs and canes, though they are somewhat helpful but only cover certain distances and they cannot detect complicated obstacles. There is, indeed, something that has been battering the 'launching' of new and better technologies to provide even better solutions for the visually impaired, and that is the accelerating pace of advancements in technology.

In paper [2], the author stated that the purpose of this study is to create an intelligent object detection system for guiding the partially sighted individuals. The system gives immediate feedback about the user's. Some of their monetary resources come from applicable funding agency here. If none, delete this. environments by making use of integrated appliances and interfaces of computer technologies [2]. The ESP32 Cam microcontroller, the depth sensor, and camera are the key components that coordinate responsibilities regarding the acquisition and management of information about the environment.

In paper [3], the author develop a software of the system is in React Native, and it is friendly to users while using Android smartphones. C++ is used to write instructions for the Arduino micro controller that analyses data from the sensors and performs instructions in cases of the discovered barriers. The YOLO (You Only Look Once) algorithm, which is an identification of objects that are exceptional for speed and effectiveness recognized all around the world, plays a vital role within the framework of the system [3]. YOLO allows the system to identify an object and, therefore, categorize it in a very short span of time, ensuring that the user is offered relevant information.

In the paper[4], the author also introduce a addition features in object recognition the device that the system have embedded GPS tracking and notification [4]. The user gets alerts concerning potential threats, and the relatives can monitor the user's location, thanks the GPS navigation tool Moreover, apart from the object recognizing function, the device is equipped with GPS navigation and alert systems [5]. The alert informs the user of a possible threat, while relatives may monitor the user's location using the GPS navigation function – all this makes the application even more reliable and safe.

The goal of the work is to create a robust and reliable assistive aid that significantly improves the security and mobility of functionally blind individuals using those innovative technologies.

II. Literature review

In the paper [6], The author stated that the need to assist the blind and the visually impaired navigate has been sorted even with various and advanced technologies such as the modern gadgets even the basic traditional aids such as the guide dogs and white sticks. While in the past such methods have proven partially effective, they tend to fall short in areas previously uncharted or those which require the subject to not only recognize but understand obstacles before making safe passage [6].

In the paper [7], The author implemented the the region based convolution neural network which aim to assist blind people for doing daily life activities. for The recent technological innovation has led to enhanced development of more sophisticated assistive devices. Cameras, infrared sensors and ultrasonic sensors for object detection have been examined by researchers [7]. For instance, due to its cost-effectiveness and simplicity, ultrasonic sensors are used as part of the solution. However, their function in discerning small or low-lying obstructions is restricted also they cannot provide detailed information concerning the qualities of such things that they sense. As already noted, infrared sensors are accurate in terms of ambient sensing; however, its reliability may be affected by light and temperature in the surrounding environment.

In paper [8], the author uses the assistive thechnologies such as camera-based vision systems are used more often than others because they provide more refined information about the environment. These systems often employ machine learning algorithm for identification of objects, and further categorization. CNNs are one of such and because of the high accuracy and less sensitivity to data misrepresentation, they have gained the reputation as the go to net for any visual identification tasks [8]. And the author in [9] The YOLO algorithm which is very important as it is ideal for assistive technology because it detects things in real time using only one neural network [9].

To provide detail information about the environment and support object recognition and distance estimation, the combinations of depth sensors with cameras have been also researched. This makes it possible to yield a better understanding of the environment which is essential in the development of believable assistive technology.

In paper [10], the author additionally integration of sound as feedback has remained another region of current study focus. Auditory feedback, as you may already know, is helpful for the user to mentally identify the space around him by responding to certain signals [10]. This method enhances the general user experience and is suitable for large datasets containing visual information.

In paper [4], the author uses GPS technology that has been implemented into the assistive systems all in a bid to enhance safety of the users by way of navigation. Besides adding to security and protection it proved to be advantageous to the caregivers and or members of the family who may need to keep track of the user's movements and engagements. Since the gadget provides information regarding position in real time, it also provides caregivers with the information they need to invert in case of an emergency making the user safe as well as creating increased satisfaction between the user and their caregivers. Further, GPS integration may provide more specific future applications such as point of interest broadcasting where the system provides direction to points of interest, or hazard broadcasting alerting the user to hazards in certain areas.

Unfortunately, while there is progress with the individual technologies that can help with learning disabilities, putting together a coherent, correct, and smooth working first level of integration is not a small task. Most current approaches are restricted in one way or another: it becomes challenging to detect specific obstacles; the user interface may be interrupted or delayed. This work proposes to improve the quality of life to the physically challenged and especially the blind through; Combined with the conventional acoustic signals, a new form of detection interfaces will give the user more clear and simple prompts that can be understood in everyday situations, so the control of the space with numerous obstacles or risks becomes less problematic. Finally, this work aims at developing an assistive technology that enhances independence and improves the quality of the lives of the visually impaired by providing a satisfying solution to the safety and usability problems when using a computer.

III. Proposed Methodology

In developing this object recognition system, several important aspects are integrated into the design such as depth estimates, vanishing point estimation, procedural and architectural design, as shown in Fig 1. All of these parts offer a comprehensive plan of the most important steps, which will allow people with vision impairment to become conscious and achieve security in unfamiliar conditions.

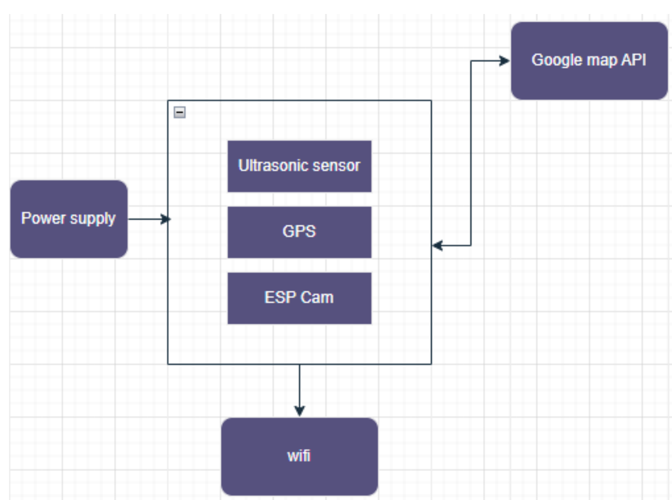

Fig. 1. Flowchart of the system

A. Procedural design

Development is quite formalized and is one of the components of the systems procedural setting. The first step is knowledge gathering, and in this the external camera and depth sensor for admitting visual and depths in formation. Depth sensor gives data in the three planes and the camera perceives the surroundings. Data collection is succeeded by data processing where the object detection is performed using the YOLO approach [9]. The ESP32 CM micro controller enables objects recognition and measurement of distances, as well as incorporation of such details [11] , as shown in Fig 4.

The next stage is the feedback generation whereby the data has to be converted from audible signals through speakers. These auditory signals augment the user's SA by using different sound or tones to stuffs or tones to tell the existence and closeness of the risks. Finally,React Native are used for a wider elaborate interface and easier operation of the user interface [12]. This interface creates a workable and accessible system to make those changes and explore the system to get real-time GPS tracking.

B. Architectural design

The construct of the system identifies a broad layout of the system to ensure that it has hardware and software components that run continually. The ESP32 CM Micro controller that can handle, system control and act as a middleman between the different systems while handling different inputs from different sensors [11], as shown in Fig 2 and Fig 3 To offering three dimensional spatial data which enables the system to obtain the relationship distance of the objects the Depth Sensor is mandatory. Moreover, it captures the nearby environment images which is necessary for the object recognition, and the Speakers gives the input audio with several navigational hints and alarm signals in accordance with the processed data in real time.

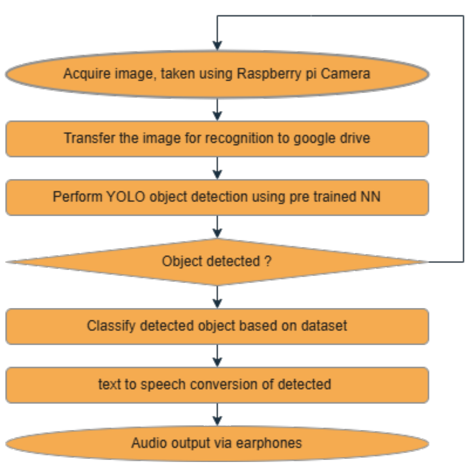

Fig. 2. Architectural design

Component of the system is the software that includes mobile application, which was developed with the use of React Native as the application interface. Accordingly, this application allows users to change settings, interact with the devise and track real-time GPS. The firmware in C++ application layer executes on ESP32 CM and manages the sensor inputs, executes object detection algorithm (YOLO), and controls the reactions of the system. Real-time object recognition and classification are essential for implementing the integration of the YOLO algorithm as it helps the system spot barriers and provide the acknowledging response to the user [13].

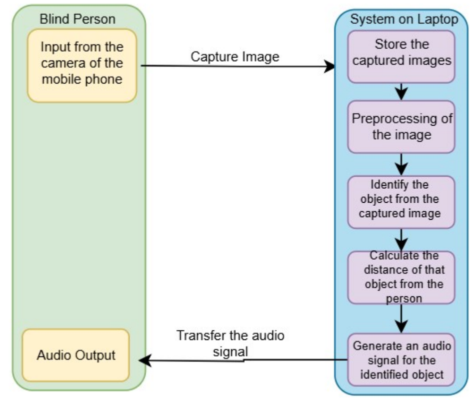

Fig. 3. Architectural design

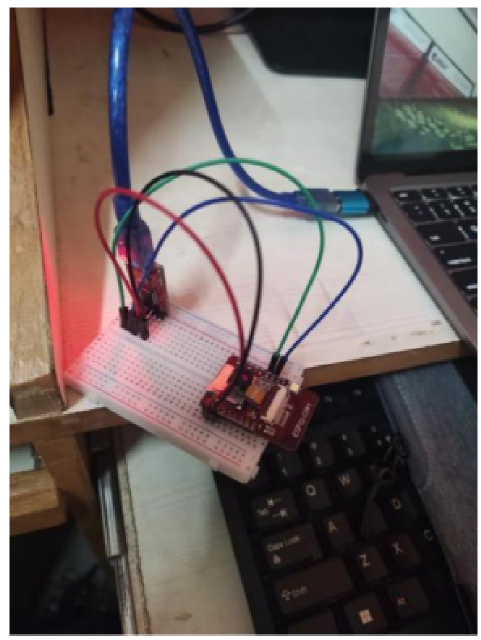

Fig. 4. Design Used

C. Vanishing point estimation

The perspective of the environment is of paramount importance; hence, vanishing point estimation is a primary approach to the functioning of the system. This approach identifies the location on the visual display where the parallel lines intersect, a piece of information that is useful in many ways. If it guesses the position of the vanishing points, the system can differentiate distances between the objects. The look large and are near the vanishing point if large and look small when they are far from the vanishing point. For the system to provide these navigation cues with the desired degree of accuracy at the requisite cadence, this perspective information is essential for determining object distances. Also, the vanishing point provides information concerning the surroundings, which assist the viewer in learning about perspective and positioning. This change of view point.This improvement in perspective assists to give out better latitude and longitude direction and hinders identification making the system more reliable and effective. The integration of vanishing point estimation may help the system to offer more relevant and natural comments making communication across virtually any of the relevant domains more enjoyable in the broadest sense.

D. Depth estimation

Depth sensing is a critical process that employs the depth camera and the RGB to gauge an object's distance from the user. The depth sensor is critical because it has direct distance measurement, as it sends out signals based on the reflections of objects and analyzes how far they are. The use of raw depth sensor data to supplement data that is captured by the camera is referred to as Camera data integration which enhances the measure of accuracy [14]. This integration helps in comprehending various object positions and their relationships; an entire three dimensional overview of the surrounding environment is constructed.Analyzing the combined data using sophisticated algorithms, the depth information is further analyzed with a view to accurately determine distances and, therefore, give the user real time information about the impediments around them. In addition to making program improvements, this method refines the users' ability of receiving accurate or timely help in navigation and then to travel across their environment thus reducing on the probabilities of experiencing mishaps.

IV. Evaluation parameters

In this experiment we have used performance metrics for the evaluation of predictive models:

For a model accuracy is one of the simplest performance measure that relates to portion of instances that are correctly either classified as positive or negative as context to fig 5.

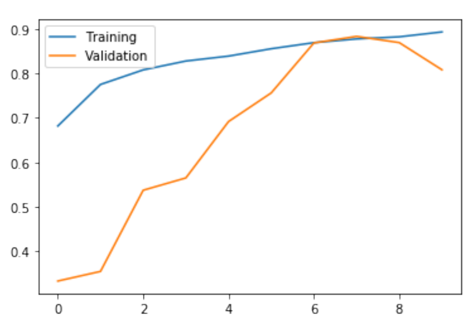

Fig. 5. Accuracy representation

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{1}$$

Precision is applied to evaluate the performance models. In the context of binary classification, precision defined as the proportion, or fraction of cases where the test concluded that the patient had the condition, out of all cases where the test overall said the patient had the condition. It embodies the extent of validity of the positive aspects of a model.

$$precision = \frac{(TP)}{(TP + FP)} \tag{2}$$

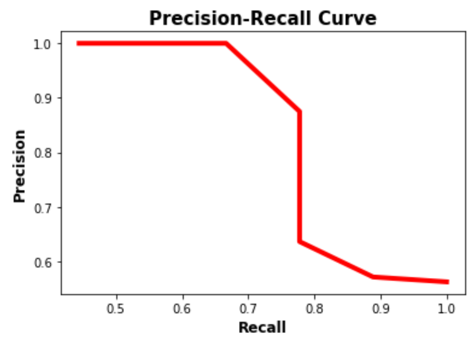

Fig. 6. Accuracy representation

Recall quantifies how many of the actual positive cases were correctly predicted by the model.Recall is crucial when the cost of a false negative is high. It shows the model's ability to detect all positive instances as shown in fig 7.

$$Recall = \frac{(TP)}{(TP + FN)} \tag{3}$$

The F1 score is the harmonic mean of precision and recall. Harmonic mean is also affected in the same way as the nomal mean calculation, where if either precision or recall is low the F1 score will be low because the harmonic mean weights lower values than higher ones.

$$F1score = \frac{2.(Precision.Recall)}{(Precision + Recall)} \tag{4}$$

V. Result

The experiment was helpful to create a complex object detection system as a combination of different hardware and software solutions that may assist the visually impaired people. Validity of the system was evaluated by a battery of tests that estimated its accuracy, reliability and practical utility.

Class	True Positives (TP)	False Positives (FP)	False Negatives (FN)	True Negatives (TN)	Precision	Recall	F1-Score	Accuracy
Object A	85	10	15	589	0.891	0.850	0.872	0.910
Object B	71	18	9	583	0.778	0.875	0.824	0.897
Object C	90	15	20	568	0.857	0.818	0.837	0.905
Object D	60	12	18	580	0.833	0.769	0.800	0.880
Object E	75	8	28	591	0.903	0.751	0.824	0.895
Object F	94	6	9	503	0.950	0.905	0.927	0.939
Overall	475	69	109	3414	0.871	0.813	0.841	0.906

Fig. 7. calculation for accuracy,precision, recall, F1 score

In fig 8 (graph) the model has been tested in six classes of objects ranging from A to F to a reasonable precision. It reached 85 TP; a Precision of 0.891 and Recall of 0.850; hence an F1-Score of 0.872 accurately ascertained an Overall Accuracy of 0.910 based on the confusion matrix. As concerning Object B, it obtained 71 TPs, Precision of 0.778 means a high recognition of actual objects, Recall: 0.875 means high ability detecting actual objects, F1-Score: 0.824, and Accuracy: 0.897. Object C had highest TP with 90; Precision was computed as 0.857, Recall as 0.818; F1-score was 0.837 and accuracy as 0.905. For Object D, We had 60 TPs, Whereby the Precision is 0.833 Recall = 0.769, F1-Score = 0.800, Accuracy = 0.880. In the case of Object E, there were 75 TPs, 0.903 Precision, 0.751 Recall reaching F1-Score of 0.824 and Accuracy of 0.895. Lastly, Object F performed extra efficiently with 94 TPs and gained a high Precision of 0.950 and, Recall of 0.905 Hence, realized a good F1-Score of 0.927 and Accuracy of 0.939. The model identified 475 TPs, 69 FPs, 109 FNs, and 3,414 TNs in the general case. The compound in question shows an Overall Precision of 0.871, Overall Recall of 0.813, Overall F1-Score of 0.841, Overall Accuracy of 0.906. These results once again demonstrate the strong and stable performance of the model in recognizing and categorizing objects, which is critical for applications aimed at improving the navigation aids for the blind.

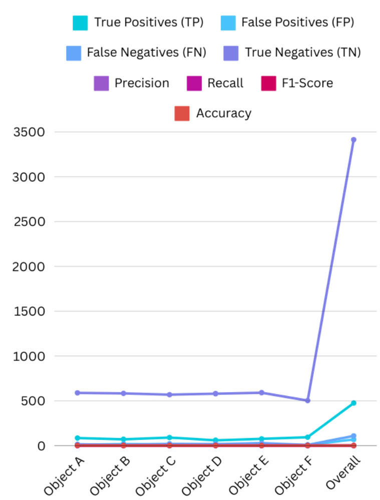

Fig. 8. Graph of evaluation

Object Detection and Distance Estimation: The system was proved to show high efficiency and accuracy on object detection based on the YOLO method [3]. In synergy with the depth sensor, the camera was able to accurately reconstruct the environment in 3 dimensional space, based on which the features and layout of objects are easily distinguishable due to the well-defined mesh. The depth aspect gave accurate distance information of objects and the captured video added to identifying object locations in the user's trajectory. As illustrated in Figure 5, the process for living begin detection is demonstrated while Fig 6 and Fig 7 illustrate the detection of non-living objects. It enabled the system to address the issues with obstacles' recognition and provide accurate distance information that is essential for real-time navigation. Moreover, flexibility together with stability is another factor of the system that permits a fast response to change different environments and give stable feedback following the result which support improvement of it as an assistant in the mobility help. Auditory Feedback: The aural feedback part greatly enhanced the efficiency of the notification concerning the presence of obstacles which were to be in the way soon, thus providing users with accurate and timely information. The system's speakers produced different pitches depending on type, size, and distance of the identified objects to offer the users an easy orientation in the proximity of threats [15]. Such differentiation of tones enabled the users to immediately understand the nature of surrounding environment and respond adequately to threats. The auditory signals that were adopted incurred approval from the user testing as being clearly discernible and easily understandable The signals also received positive input concerning their ability to boost the confidence of users while perusing through the vista. This real-time auditory guidance improved the overall mobility experience adding to its efficiency, safety and accessibility for the visually impaired. Application Interface: The created mobile application that is developed using React Native was received well among the users. Those were a touch panel, enabling users easily set, monitor and command different options of the gadget, a GPS tracking display for full monitoring of the device's location. The sample argued that the ability to change preferences on the spot and monitor them enhanced their sense of security

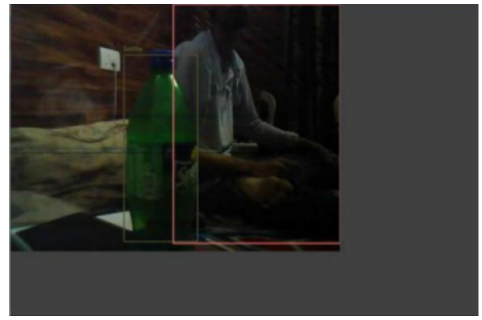

Fig. 9. Non-living being as an output

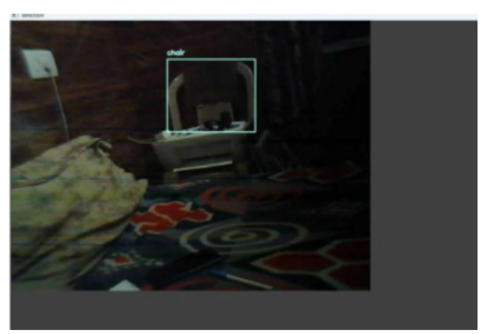

Fig. 10. Non-living being as an output

among the users [5]. Overall Performance: The depth and vanishing point estimation were among the most important results which contributed to the general enhancement of spatial information acquisition and precision of the system. By fusing depth perception with real-time object detection, the system was able to offer the user a holistic sense of the environment and thus allow for accurate navigation through the complex unfamiliar environment. People said the features improved overall perception of space and gave them more confidence while commenting that they would be more mobile when they did not need help from carers. The technology was found to be helpful in providing information about obstacles and ways in the indoor and outdoor environment, which was clearer and consistent than observations. This well measured and controlled effort showed that such sophisticated technologies could indeed evolve into useful and change agent assistive tools in the daily existence of a blind population. The latter strengthened the belief in the ability of the developed assistive system to navigate the visually impaired individual successfully, thus returning a large degree of freedom and safety to their lives. Fig 8 shows the visual coordination process.

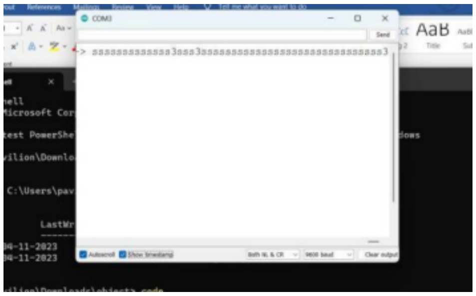

Fig. 11. Visual Coordinates

VI. Conclusion and Future scope

It also demonstrates the need to incorporate the best of today's technology for better help with navigation as the work involves establishing and using of a very advanced detection object recognition system for the purpose of assisting the visually impaired. Real time object detection involves the use of specific parts which includes the ESP32 CM micro controller, a depth sensor, a camera and speakers and complicated software such as YOLO [11].

From our evaluation, we have observed that the proposed system has a very smooth and high rate of object detection and distance measurement, while offering users enhanced and timely directions in the navigation system. This is achieved by the use of camera views in synchrony with a depth view that renders the full three dimensional view of the environment and improve on the detection of hurdles as well as operational space of the car. It has been clearly considered that the capacity of producing quite discrete signals is available in the auditory feedback system and it is quite helpful for the navigation depending on the context [6].

Through an interface developed in React Native, the mobile application lets the user directly make changes to settings in addition to providing real-time GPS tracking. This function allows family members and caregivers to track the user's location which increases user's confidence and adds important layer of security.

The vanishing point estimate added to the system complements comprehension of spatial layouts and distances to objects by providing precise and comprehensible navigation support. From all the foregoing, it is evident that the work have served its purpose of increasing the mobility as well as safety of the visually impaired people thereby showing that technology has a positive impact on the lives of the physically challenged people.

The prospect of further development in the field of assistive technology is revealed by the smooth compatibility of the different appliances into a single system. Thus, to make the system still more effective and to increase customers' satisfaction, the following steps might be taken in the future: More attention can be paid to the aspects of the user interface and interaction, including design and interface adaptations; More data related to the environment can be collected to be used in the application; Object detection algorithms can also be improved in the future [16].

The potential of expanding the proposed assistive technology framework for the future scope has been discussed in a number of promising directions. First, object recognition capabilities could be extended to include additional distinctive features and shapes of the obstacles which includes: road signs, traffic signals and moving objects. Including better algorithms like a new YOLOv5 or YOLOv7 the system could be optimized in both, precision and reaction time, which would make navigation less dangerous and more user-friendly in harsh conditions [17]. Furthermore, LiDAR with the current camera and infrared system, will also be useful by providing depth detection which will prove useful in different lightning environment and improves the perception of the surroundings. Later releases might even provide feedback in multiple modalities, including ensuring tactile features for those who are using software with noisy or poor visibility environments [18]. This would improve usability by adding touch feedback to obstacle or directions which alongside the audial notifications would serve to greatly improve the use of the device and an aid to the visually impaired persons.

References

[1] K. Swathi, B. Vamsi, and N. T. Rao, "A deep learning-based object detection system for blind people," in *Smart Technologies in Data Science and Communication: Proceedings of SMART-DSC 2021*, pp. 223–231, Springer, 2021.

[2] M. Konaite, P. A. Owolawi, T. Mapayi, V. Malele, K. Odeyemi, G. Aiyetoro, and J. S. Ojo, "Smart hat for the blind with real-time object detection using raspberry pi and tensorflow lite," in *Proceedings of the International Conference on Artificial Intelligence and its Applications*, pp. 1–6, 2021.

[3] R. Tu, Z. Zhu, Y. Bai, G. Jiang, and Q. Zhang, "Improved yolo v3 network-based object detection for blind zones of heavy trucks," *Journal of Electronic Imaging*, vol. 29, no. 5, pp. 053002–053002, 2020.

[4] A. Rashid and H. S. Dhillon, "Gps-enabled blind stick using ultrasonic technology," *International Journal of Innovative Research in Engineering & Management*, vol. 11, no. 1, pp. 6–9, 2024.

[5] D. R. Bolla, S. Trisheela, P. Nagarathana, and H. Sarojadevi, "Object detection in computer vision using machine learning algorithm for visually impaired people," in *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, pp. 1–6, IEEE, 2022.

[6] M. Darshan A, P. Dipam, and S. Rina, "Object detection and recognition using tensorflow for blind people," *International Journal of Trend in Scientific Research and Development*, vol. 8, no. 5, pp. 747–752, 2024.

[7] D. Devahema, S. Shyaam, M. Karthikeyan, V. Vishal, and G. Pushpak, "Object detection for blind people using faster region-based convolutional neural networks," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 11, pp. 4915–4919, 2020.

[8] A. Bhandari, P. Prasad, A. Alsadoon, and A. Maag, "Object detection and recognition: using deep learning to assist the visually impaired," *Disability and Rehabilitation: Assistive Technology*, vol. 16, no. 3, pp. 280–288, 2021.

[9] M. Mahendru and S. K. Dubey, "Real time object detection with audio feedback using yolo vs. yolo_v3," in *2021 11th international conference on cloud computing, data science & engineering (confluence)*, pp. 734–740, IEEE, 2021.

[10] R. R. Karmarkar and V. Hommane, "Object detection system for the blind with voice guidance," *Int. J. Eng. Appl. Sciences Technol*, vol. 6, no. 2, pp. 67–70, 2021.

[11] R. K. Kaushal, K. Tamilarasi, P. Babu, T. Mohanaprakash, S. Murthy, and M. J. Kumar, "Smart blind stick for visually impaired people using iot," in *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, pp. 296–300, IEEE, 2022.

[12] X. Huang, *Comparative Evaluation of Java and Kotlin for Android App Development*. PhD thesis, University of Applied Sciences Technikum Wien, 2018.

[13] A. Aralikatti, J. Appalla, S. Kushal, G. Naveen, S. Lokesh, and B. Jayasri, "Real-time object detection and face recognition system to assist the visually impaired," in *Journal of Physics: Conference Series*, vol. 1706, p. 012149, IOP Publishing, 2020.

[14] S. Malkan, S. Dsouza, S. Dasmohapatra, M. Jain, and A. Joshi, "Navigation using object detection and depth sensing for blind people," in *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 1–7, IEEE, 2021.

[15] H. Nguyen, M. Nguyen, Q. Nguyen, S. Yang, and H. Le, "Web-based object detection and sound feedback system for visually impaired people," in *2020 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, pp. 1–6, IEEE, 2020.

[16] M. Banne, R. Vhatkar, and R. Tatkare, "Object detection and translation for blind people using deep learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 03, 2020.

[17] O. K. T. Alsultan and M. T. Mohammad, "A deep learning-based assistive system for the visually impaired using yolo-v7," *Revue d'Intelligence Artificielle*, vol. 37, no. 4, pp. 901–906, 2023.

[18] C. Dragne, I. Todiriţe, M. Iliescu, and M. Pandelea, "Distance assessment by object detection—for visually impaired assistive mechatronic system," *Applied Sciences*, vol. 12, no. 13, p. 6342, 2022.