A Project Proposal Submission
on

Real-Time Object Detection for Blind Assistance

For the course of

Project Based Learning
in
Department of Robotics and Automation

Submitted by,
Gunjay Suhalka (22070127022)
Mayank Ranade (22070127038)
Raj Shah (22070127053)
Sahran Altaf (22070127059)

Under the Guidance of,

Dr. Arunkumar Bongale
HOD,
Department of Robotics and Automation

Dr. Sameer Sayyad
Assistant Professor,
Department of Robotics and Automation

SYMBIOSIS INSTITUTE OF TECHNOLOGY,
A Constituent of Symbiosis International (Deemed University),
Pune – 412115

2024-25

SYMBIOSIS INSTITUTE OF TECHNOLOGY (SIT)

A Constituent of Symbiosis International (Deemed University), Pune

Annexure 2

Undertaking for Literature review and Project proposal Submission

Department	**Robotics and Automation**
Batch	**2022-26**
Name of the Students (PRN)	1. **Gunjay Suhalka (22070127022)** 2. **Mayank Ranade (22070127038)** 3. **Raj Shah (22070127053)** 4. **Sahran Altaf (22070127059)**
Name of the guide(s):	**1.** Dr. Arunkumar Bongale, HOD, Department of Robotics and Automation **2.** Dr. Sameer Sayyad, Assistant Professor, Department of Robotics and Automation
Project Title	**Real-Time Object Detection for Blind Assistance**
Similarity Index (should be less than 10%)	1. Similarity Index: 2. AI Report:

Date: 29th January 2025

Signature of Student(s):

1.

2.

3.

Signature of Project Guide

Signature of Project Co-guide

4.

SYMBIOSIS INSTITUTE OF TECHNOLOGY (SIT)

A Constituent of Symbiosis International (Deemed University), Pune

Annexure 3

Estimated Project Expenditure

Title of the Project: Real-Time Object Detection for Blind Assistance

Sr. No	**Particulars with Justification**	**Amount (INR)**
1	Material	
2	Electronics Hardware Components 1. Raspeberry Pi 4 Model B 2. Camera Module (2x) 3. Audio Output Module	Rs. 9500
3	Software	
4	Fabrication	
5	Miscellaneous Components 1. Jumper Wires 2. Bread Board 3. Cable ties 4. Power Bank	Rs. 1000
	Total	Rs. 10,500

Name and Signature of the Student(s) **Name and Signature of Guide(s)**

1.
2.
3.
4.

LIST OF ABBREVIATIONS/ SYMBOLS

Abbreviations		Description
RA	:	Robotics and Automation
AI	:	Artificial Intelligence
CNN	:	Convolution Neural Network
Faster R-CNN	:	Faster Regional- Convolution Neural Network
YOLO	:	You Only Look Once
SSD	:	Single Shot object Detection

Abstract

Vision is one of the most important senses, which plays a very important role in perceiving and controlling the environment. Unfortunately, there are 285 million people in the world who are visually impaired, and 39 million of them are completely blind according to the World Health Organization (WHO). Currently, India is a home for an estimated 4.95 million blind persons and 70 million vision impaired persons. A real-time object detection system is proposed for this purpose to help the visually impaired people to move around safely and independently in their surroundings. The system is worn around the chest and uses cameras for object detection with AI algorithms and audio feedback. This solution is expected to remove the restrictions of conventional tools like walking sticks by providing real time environmental information to the visually impaired user. This system increases the range of motion, decreases the risks while being ergonomic and lightweight to enhances the quality of life of the visually impaired.

1. Introduction

Vision impairment has emerged as a significant worldwide problem in the sense that millions are severely restricted in their mobility and freedom. The World Health Organization states that 39 million people are completely blind while the overall visually impaired population is estimated around 285 million. People with such needs require special care and empathy to explore and interact in and around their environment. The use of primitive assistive devices such as walking stick or assistance dogs fail to provide a way to help the visually impaired person detect or recognise the object around them.

The developing regions, where 90% of the visually impaired population lives, are most affected. Most of these people cannot afford expensive, advanced devices that can help them move around. Physical walking sticks are able to recognize physical barriers, however, only type of barrier and special training has to be acquired which is not available in all areas.

Technological progress has provided a new opportunity to tackle problems in novel ways. This work develops a real-time object based detection and recognition system to facilitate independent mobility and movement for a visually impaired person. This system will use cameras, sensors, and artificial intelligence-based object recognition in the process for real-time audio feedback delivery to the listener. This way, the users can learn information about obstacles and objects located in the environment, that allows them to move with more safety and autonomy.

The objective of the project consists in building a system based on state-of-the-art technologies, including cameras, sensors, and AI algorithms, in order to give real-time feedback. This device, being both easy and convenient to use, allows for the mounting of on the user's chest strap to recognise objects present in the immediate environment of the user, estimate their distance, and present this information as audio feedback. This device will work to overcome the limitations of the currently existing tools to provide an inexpensive, easily available solution for the patients of the developing world- promoting independence and safety.

The intent of the present project is to design a real-time object detection system that gives greater independence and mobility to visually impaired individuals. It uses current technologies like cameras and sensors along with AI-based object recognition algorithms to provide the user with real-time audio feedback.. In this way, users will be provided with information about objects/barriers present in their environment to navigate more safely and autonomously. Consideration is given to the fact that such usages within that context, attaining its functions for a chest strap type of a device, will, among other factors, specifically include prediction of the possible distance to nearby objects on the basis of the systems using that very device. The information used shall be feedback through audio means. This solution holds the potential to outgrow the limitations of tools that are being designed with the development of respective community based on freedom of independence and safety for the individual developers in its area at a current stage.

2. Review of Literature

In the course of years, these alternative technologies have given birth to a totally new device environment that has modified greatly the ways that vision is augmented. Several researchers have incorporated the visual scene perception algorithms for sensory feedback systems into the navigational aids for the blind. This paper, Kumar et al. (2021): Object Detection System for Visually Impaired Persons by Smartphone, expresses a projection whereby such smartphone-based object detection systems can be enhanced for real-time spoken feedback. The system employed machine learning in identifying everyday objects surrounding a visually impaired person. Albeit portable and off-the-shelf, the team discovered that its speed and accuracy of detection were limited while operating in generally low-light conditions. Another article, "Object Detection in Visual Impairments Using Yolov3" by Deshmukh et al. (2020) pointed out the application of deep learning algorithms for real-time object recognition. This was done using the YOLOv3, a very accurate object detection model, that provides more accuracy on detecting various objects in the environment of the user. The paper demonstrated significant advance over the previous models in terms of both speed and accuracy, but also claimed that it lacks the detection capability for such small objects or objects embedded in a dense scene.

Mahesh et al. (2021) describe "CICERONE-a Real-time object Detection system for the Visually impaired, which is easily programmed by applying static artificial intelligence (AI) algorithm to the real-time video feeds. This system designed for the urban environment had functionalities like obstacle detection, path guidance, and even real-time feedback from haptic and audio signals. It has great potential to make the visually impaired users more independent in an urban busy environment."

In this paper, Monika et al. (2021) described a more advanced object detection system called "Object Detection for Visually Impaired People Using YOLOv4." This system built on prior work utilized YOLOv4 attempt real-time object detection with enhanced accuracy. Moreover, it introduced depth prediction features so the users could estimate the distance between themselves and detected objects, thus greatly enhancing navigation in complex environments. This study concludes that, although YOLOv4 managed to get high accuracy in successfully detecting objects, it already needed a high computation power to make the portable electronic (e.g., smartphones, low-power wearables) implementations questionable.

In terms of assistive technologies for the visually impaired, development has been focused on the sensors and cameras and especially on the application of artificial intelligence. Different studies described the possibility of real-time object recognition systems to improve the mobility of the visually impaired to a great extent. For example, some systems employ ultrasonic transducers to detect the obstacle and then report its position as vibration or sound. But they can't discriminate what the object is not its distance from the robot, and as a result they are not useful at all.

Camera-based systems provide with richer information, namely since they not only recognise but also track objects in real-time. The present-day system utilizes AI-driven algorithms, such as Convolutional Neural Networks (CNNs), for detection of objects, a combination that improves upon object detection's performance. Besides, it has certain audio feedback that enables the user to hear characteristics of the object in real-time. However, for all of these reasons, the market has openings where more reliable, accurate, cost-effective, and accessible solutions are required.

3. Research Gaps

Despite the progress in blind assistance technology, there are still significant shortcomings that limit the effectiveness of current solutions.

i. Limited Contextual Awareness: Traditional assistive devices, like walking sticks, lack contextual information of the objects that are present in the environment. The systems studied by Kumar et al. (2021) and Deshmukh et al. (2020) deliver object detection, but the systems are challenged to provide detailed, real-time contextually rich information to a visually impaired person which informs what the meaning of an object is in different situations, including low light or environments with occlusions.
ii. Real-Time System Accuracy and Affordability: While YOLOv3 and YOLOv4 have proven to be very accurate in detecting objects, research by Monika et al. (2021) showed that these systems need high computational power, which is costly and not readily available to many in developing regions. There is a gap in the creation of low-cost, low-power systems maintaining high accuracy in real-time object detection.
iii. Understanding of depth: Although in recent years, studies like Mahesh et al. (2021) have developed systems with object detection and feedback capability, limited cases alone seem to join with reliable depth estimation. Depth estimation is crucial and will help develop better navigation and spatial awareness for the blind. It is very important to consider for the enhancement of user safety and mobility within which the visually

impaired users measure distances correctly. Despite recent developments such as Mahesh et al. (2021), which have produced approaches to develop systems capable of object detection and delivering feedback, there has not been much synergy on reliable depth prediction, only a few instances. Such prediction of depth will help better navigation and spatial awareness for blind individuals, again an important point to consider for the further enhancement of user safety and mobility, with blind users measuring distances accurately.

iv. Integration of Multiple Sensors: Existing models are typically based on camera- or ultrasonic-sensor-only single sensor systems, thus limiting the overall performance in complex situations. An empty space, however, is provided, referring to how to combine stereo vision camera system, and so to build a robust real-time system that is able to capture high efficiency input under all kinds of lighting and environment conditions.

Closing these gaps with the project will result in an affordable real-time object-detection system providing accurate context and depth information, thereby facilitating better mobility and safety for a visually-impaired individual.

4. Research Questions

i. To what extent has the new object detection system been effective considering visually impaired individuals could use it instead of standard aids such as walking sticks?

ii. Is the YOLO algorithm capable of performing object detection and classification in conditions that are either at low light or where the objects are occluded? If yes, how does this performance compare with other deep learning techniques such as SSD and Faster R-CNN?

iii. How will the capability of harboring depth prediction with stereo vision or depth sensors enhance the object detection systems for the blind? Will it be more accurate and effective?

iv. Considering the developing parts of the world where resources might be limited, how cost effective and affordable in terms of space does a real time object detection system strive to be?

v. How much does real-time audio feedback increase the independence and mobility of the Visually Impaired user in urban and rural environments?

5. Statement of the Problem

The proposed project involves the development of a method for recognizing objects of regular occurrence in a specified setting via a functional prototype that provides real-time audio feedback. The project ultimately aims to promote the movement and independence of the visually impaired by overcoming limitations of primary tools like a walking stick that do not possess object recognition properties.

6. Objectives of the Study

i. Develop a vision-based system based on the fusion of current state-of-the-art algorithms to detect and track objects in real-time. The emphasis of the system will be on providing direct feedback to the visually impaired so that they are aware of, and do not collide with, the environment in which they find themselves.

ii. Depth Perception for Improved Safety: Introduce depth formation to the system in order that distance between the recognized objects and the user is estimated. This will enhance spatial orientation, in consequence, which means that it will be possible to navigate more safety in complex environments, through the possibility of knowing how far away from obstacles, the user is.

iii. Audio Feedback: Audio feedback will be pleasant enough to mediate an object information presence, object type, and object position to the user in real time. Sound Output will be noiseless and not bulky, but clear, crisp, and unmistakable, for rapid decision making for navigation.

7. Scope and Limitations

7.1 Scope

The mentioned project aims to design and develop an innovative system for real-time object detection that acts as a support tool in a way that the visually impaired could navigate through it. It offers existing technologies by merging AI object detection and depth prediction to give feedback to ensure mobility. Detection of normal objects in real time along with obstacles can be done through advanced AI algorithms such as YOLO:

•Supports accurate depth prediction to tell a user that he/she is some distance away from an object.

•Clear audio feedback that lets the visually impaired user know what is going on in the surroundings around them.

•Be cheap and readily available in an effort to make it relatable to developing regions, where technology for such purpose is not easily accessible.

•Be portable and user-friendly, easy to use on a daily basis without complicated setup and maintenance.

7.2 Limitations

While this system is quite advantageous, it has a lot of limitations:

•Processing Power: The system becomes complex because it will incorporate mechanisms like image processing and interpretation based on rather intricate algorithms. Therefore, execution could have considerably high computational costs for this, making it less portable and expensive in resource-limited areas.

•Conditions: Relatively unfriendly ambient conditions, poor lighting sensitivity, glare, and clutter constitute an impediment to object detection accuracy.

•Real-Time Processing: It could be operational in real-time, even from distant regions as its working is ensured by further designs in additional hardware that may be sold.

•Limitations of Audio Feedback: Too many auditory channels may confuse the user, especially when multiple objects and/or barrier-like structures are cluttering the environment.

•Durability: The equipment is to be worn by a person, and concerns arise as to the ruggedness of certain components under certain extreme environmental conditions or rigorous physical activities.

8. Methodology, Tools and Techniques

The developed assembly incorporates a variety of critical components which make a decisive contribution to both its operational parameters and usability. One other key component of the system is the camera (which is on the subject's chest strap). The setup allows the video camera to film the video feed in the line of sight of the user and the displays of live video information related to the surrounding environment close to the user, so as to allow accurate perception of whatever surroundings they are passing through.

The processor will run the video stream from the camera in a real-time scenario with the help of some object detection algorithms. The recognition and classification of different objects based on Convolutional Neural Networks(CNNs) will be pipelined into the video pipeline. Run such a YOLO object detection algorithm that will be fast and accrue accuracy in real time. YOLO can apply to multi object detection in a single frame, and it can also localize an object; therefore, it can be applied in dynamic environments meant to assist visually impaired people.

The system will calculate after the objects are detected, object depth or distance from the user. For this the stereo vision method or depth sensors are employed as it is depended on the system that is employed. Depth information is crucial to its ability to give to the user not only information contained in the context in and around a subject but also the relative distance of objects from each other.

The processed information will be delivered as audio feedback. They will be genetically propagated using a single pair of headphones or mini speaker system on the chest skin of a subject attached to the chest box (whole system). Based on audio information, the user will be warned of the presence and the distance to obstacles in the environment as well as be led more accurately through that environment. Technology will ensure that the auditory system presents the information with a high level of clarity and unequivocalness without unnecessary complication, but in no way overwhelm the user with the uncontrollable data flood. At specific objects a vigilance will happen, depending on the distance and the relative motion to the object, when the human, arbitrarily "alarms" the obstacles at varying degrees of importance. It has to be fastened to a subject's chest harness. The entire system consists of a camera and a microcontroller to obtain real time images. Object detection algorithms using Convolutional Neural Networks (CNNs) will be used to work on video streams and perform the classification and detection of the objects. After recognizing the object target, the system can then provide the estimated distance/depth of the object using the solutions from the stereo vision or the distance sensors. This data will be further elaborated via auditory feedback, i.e., the data can always be provided in real time also relative to the used context in which the user resides.

It will train the CNN for a group of everyday objects to the point where recognition will be more reliable in a range of contexts. The hardware setup would be to pick up a camera of high resolution, a microcontroller with the ability to perform considerable processing, and speakers or headphones for effective audio feedback.

10. References

i. D. Ravi Kumar, H. K. Thakkar, S. Merugu, V. K. Gunjan, and S. K. Gupta, “Object Detection System for Visually Impaired Persons Using Smartphone,” *ResearchGate*, 2021. [Online].

ii. P. Deshmukh, A. Khedkar, S. Kulkarni, and S. Morkhandikar, “Object Detection for Blind People Using Yolov3,” *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 8, no. 4, 2020.

iii. T. Y. Mahesh, P. S. S., S. Thomas, S. R. Thomas, and T. Sebastian, “CICERONE-A Real Time Object Detection for Visually Impaired People,” in *IOP Conference Series: Materials Science and Engineering*, vol. 1012, no. 1, 2021, pp. 012-024.

iv. K. Monika, N. N. Nivetha, and R. A. Rohini, “Object Detection for Visually Impaired People Using YOLOv4,” *International Journal of Research in Technology and Innovation (IJRTI)*, vol. 5, no. 6, 2021, pp. 501–512.