COMPUTER VISION BASED ASSISTANCE SYSTEM WITH VOICE FEEDBACK FOR THE VISUALLY IMPAIRED PEOPLE

PHASE I REPORT

Submitted by

DEVIP
(2020246027)

in partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY IN INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING, GUINDY
ANNA UNIVERSITY CHENNAI 600 025
FEBRUARY 2022
Certified that this project report titled “COMPUTER VISION BASED ASSISTANCE SYSTEM WITH VOICE FEEDBACK FOR THE VISUALLY IMPAIRED PEOPLE” is the bonafide work of DEVI P (2020246027) who carried out project work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

PLACE: CHENNAI
DATE:

Dr. M. VIJAYALAKSHMI
ASSOCIATE PROFESSOR
PROJECT GUIDE
DEPARTMENT OF IST, CEG
ANNA UNIVERSITY
CHENNAI 600025

COUNTERSIGNED

Dr. S. SRIDHAR
HEAD OF THE DEPARTMENT
DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING, GUINDY
ANNA UNIVERSITY
CHENNAI 600025
ABSTRACT

Visually impaired people face the great challenge of moving independently. To assist the visually impaired people for safe movement in the environment, a model is developed to identify the obstacles and categorize the objects and notify them using audio feedback system. The proposed prototype is developed using Raspberry Pi 3 Model B+ in this project. The system consists of a sensors and Pi camera for obstacle detection and object recognition respectively. The distance between the user and the obstacle is measured using ultrasonic sensors and if the distance is below the given threshold, the Raspberry Pi Camera is used for real time object detection and the object detected is notified to the user through an audio feedback. In our system, the YOLO (You Only Look Once) algorithm is used for the object detection to categorize the objects into their appropriate classes. The proposed system when compared with the other aids like white cane, is very comfortable and provides ease of navigation for the visually impaired.
திட்டப்பணிசுருக்கம்

பார்வைக் குவைபாடுள்ளைர்கள் சுதந்திரமாகச் சசல்ைதில் சபரும். பார்வைக் குவைபாடுள்ளைர்கள் சுை்றுச்சூழலில் பாதுகாப்பான நடமாட்டத்தில் உள்ள, தவடகவள் கண்டறிந்து பின்னூட்ட உதைதவடகவள் கண்டறிந்து சபாருட்கவள். உதைதவடகவள் லிவைப்போது (ப்பார்வைக் குவைபாடுள்ளைர்கள்) 3 மாத வருமானம் பார்வைக் குவைபாடுள்ளைர்கள் இருந்த முன்மாதிரி உருைாக்கப்பட்டுள்ளது. இந்த திட்டத்தில் முன்சமாழி பாதுகாப்பான முன்மாதிரியான ராஸ்சபர்ரி 3மாடல் B+ பண்படுத்தப்பட்டுள்ளது. அவமப்பு முவைதவடகவள் கண்டறிந்து மை்றும் பார்வைக் குவைபாடுள்ளைர்கள் மை்றும் வபி யகமரா ஆகி ைை்வைக் சகாண்டுள்ளது. பார்வைக்கு வருமானம் இந்திய சுகை செயலானது பாதுகாப்பான உள்ளூர் பார்வைக்கு வருமானம் பிரபார்வைக்கு வருமானம் இருந்த முன்மாதிரி, Raspberry Pi Camera பிரிவில் பார்வைக் குவைபாடுள்ளைர்கள் பண்படுத்தப்பட்டுள்ளது. பார்வைக் குவைபாடுள்ளைர்கள் மை்றும் பார்வைக் குவைபாடுள்ளைர்கள் மை்றும் முச்சமாழி பாதுகாப்பான முன்மாதிரி மை்றும் பார்வைக் குவைபாடுள்ளைர்கள் மை்றும். நாய்கள் அவமப்பு, YOLO அவமப்புவில் அவமப்பு வருமானம், பார்வைக் குவைபாடுள் சுதந்திரமாகச் சசல்ைதில் பண்படுத்தப்பட்டுள்ளது. முன்சமாழியின் அவமப்பு மூலம் பார்வைக் குவைபாடுள்ளைர்கள் மை்றும் பார்வை குவைபாடுள்ளைர்கள் மை்றும் சசல்ைதில்.
ACKNOWLEDGEMENT

I thank the Lord Almighty, whose showers of blessings have made this project a reality.

I express my sincere thanks and deep sense of gratitude to my guide, Dr.M.Vijayalakshmi, Associate Professor, Department of Information Science and Technology, for her valuable support, suggestions, guidance and encouragement. And also I thank her for providing me with the counsel and direction to help me complete this project.

My sincere thanks to Dr.S.Sridhar, Professor and Head, Department of Information Science and Technology for his kind support and for providing necessary facilities to carry out the project work and prepare the thesis.

I wish to record my sincere thanks to the members of review panel, Dr.S.Sridhar, Professor and Head, Dr.N.Thangaraj, Assistant Professor, Dr.L.Sairamesh, Teaching Fellow, and Dr.Tina Esther Trueman, Teaching Fellow, Department of Information Science and Technology for their valuable suggestions and critical reviews throughout the project.

DEVI P
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER NO</th>
<th>TITLE</th>
<th>PAGE NO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT (TAMIL)</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td></td>
<td>LIST OF ABBREVIATIONS</td>
<td>xi</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1 OVERVIEW</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2 PROBLEM STATEMENT</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1.3 OBJECTIVE</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1.4 ORGANISATION OF THE REPORT</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>LITERATURE SURVEY</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2.1 SUMMARY OF THE SURVEY</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>SYSTEM DESIGN</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3.1 PROPOSED SYSTEM</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3.2 SYSTEM ARCHITECTURE</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3.3 DATA ACQUISITION</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>3.4 MODULE DESCRIPTION</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3.4.1 Detect Obstacle and Its Distance</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3.4.2 Object Detection</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>3.4.3 Generate Audio Feedback</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>3.5 HARDWARE AND SOFTWARE REQUIREMENTS</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>IMPLEMENTATION AND RESULT</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>4.1 DETECT OBSTACLE AND ITS DISTANCE</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>4.2 OBJECT DETECTION</td>
<td>18</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>CONCLUSION AND FUTURE WORK</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>5.1 CONCLUSION</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>5.2 FUTURE WORK</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>TABLE NO</td>
<td>NAME OF THE TABLE</td>
<td>PAGE NO</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>4.1</td>
<td>Data Collected Using Ultrasonic Sensor</td>
<td>16</td>
</tr>
<tr>
<td>4.2</td>
<td>Computation of Error, Standard Deviation and Variance</td>
<td>17</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO</th>
<th>NAME OF THE FIGURE</th>
<th>PAGE NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>System Architecture</td>
<td>8</td>
</tr>
<tr>
<td>3.2</td>
<td>Flow Diagram to Detect Obstacle and Its Distance</td>
<td>10</td>
</tr>
<tr>
<td>3.3</td>
<td>Architecture of YOLO Algorithm</td>
<td>11</td>
</tr>
<tr>
<td>3.4</td>
<td>Object Detection in Raspberry Pi</td>
<td>12</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison Between Actual and Observed Distance</td>
<td>18</td>
</tr>
<tr>
<td>4.2</td>
<td>Real Time Object Detection Using Cell Phone as Object</td>
<td>19</td>
</tr>
<tr>
<td>4.3</td>
<td>Output of Distance and Object Detected as Cell Phone</td>
<td>19</td>
</tr>
<tr>
<td>4.4</td>
<td>Object Detection Using Cell Phone as Object at Different Distance</td>
<td>20</td>
</tr>
<tr>
<td>4.5</td>
<td>Output of Different Distance Measure for Cell Phone</td>
<td>20</td>
</tr>
<tr>
<td>4.6</td>
<td>Real Time Object Detection Using Orange as Object</td>
<td>21</td>
</tr>
</tbody>
</table>
4.7 Output of Distance and Object Detected as Orange

4.8 Object Detection Using Orange as Object at Different Distance

4.9 Output of Different Distance Measure for Orange
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>ABBREVIATION</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLO</td>
<td>You Only Look Once</td>
</tr>
<tr>
<td>SSD</td>
<td>Single Shot Detector</td>
</tr>
<tr>
<td>COCO</td>
<td>Common Objects in Context</td>
</tr>
<tr>
<td>RPi</td>
<td>Raspberry Pi</td>
</tr>
<tr>
<td>GPIO</td>
<td>General Purpose Input/Output Pin</td>
</tr>
<tr>
<td>RGB</td>
<td>Red Green Blue</td>
</tr>
<tr>
<td>RCNN</td>
<td>Recurrent Convolutional Neural Network</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

A statistics from World Health Organization reveal that there are around 253 million individuals with vision impairment in the world. This number tends to increase by 2 million per decade and this number is estimated to double in the following years. Major problem that visually impaired people encounter is unsafe mobility, they usually cannot detect and avoid obstacles in their way.

Visually impaired people need help in performing their daily tasks, like walking or moving in an unfamiliar environment. They need to identify the obstacles in their way and avoid them, so they need assistance for this and especially for secure navigation. Now the research is focusing on this problem to develop a supporting assistant for the visually impaired.

Few navigation assistances like white canes and seeing-eye dogs are available, however their performance is limited. A cane can detect obstacles at lower level that is below the knee but cannot detect the obstacles at a higher level and also their range is very short. Seeing-eye dogs also helps in navigation but they need intensive training which is difficult for the visually impaired.

Nowadays, walking assistants are developed with the embedded sensors to enhance the accuracy of vision devices. Travelling aids with computer vision are used to detect obstacles more accurately than sensor-based devices. Computer vision helps in object detection, which is used to detect and categorize the obstacles can be extended to help visually impaired people to recognise the objects. The computer vision-based object detection is used in categorizing the objects. YOLO - You Only Look Once, an object recognition deep learning framework, is used to detect objects in real time using a camera. The YOLO algorithm runs through a variation of an extremely complex Convolutional Neural Network architecture called the Darknet. Here, YOLO which is
a regression-based model is used so that predicting classes for the whole image is done quickly in one run of the algorithm.

The Voice guidance technologies are also used so the visually impaired can know the objects with voice feedback from system. This technique helps visually impaired get to know what they cannot see. The proposed system is useful to detect the objects with its distance and notify the visually impaired with the voice feedback system so it will help them move independently.

### 1.2 PROBLEM STATEMENT

The motivation of this project is to develop a visual aid system for the visually impaired person so that they can gain knowledge of the surrounding environment, avoid the obstacles in front of them, recognize the objects and navigate independently.

### 1.3 OBJECTIVE

- To develop a vision and audio based assistant system for visually impaired people to help them navigate independently in their environment.
- To detect the obstacles and find its distance to notify the visually impaired person.
- To detect and categorize the objects in the environment and notify the visually impaired person.

### 1.4 ORGANISATION OF THE REPORT

The rest of the report is organized as follows. Chapter 2 describes the related works in the development of a visual aid system. Chapter 3 explains the system design of this project. Chapter 4 explains the implementation and results obtained in this project. Chapter 5 explains the conclusion and future work of this project.
K. Sundar Srinivas et al., [1] proposed an electronic device for obstacle detection and face recognition to assist visually challenged users in social interactions. The proposed device was in the form of smart glasses that has ultrasonic sensor, pi camera and raspberry pi installed on it. The ultrasonic sensors are connected to the raspberry pi that receives data signals from the sensors for further data processing and detects the obstacles around the user up to 1000 cm. A pi camera is used to recognize the person in front of the visually challenged according to the database list. This device supports to recognize people easily. It is of low cost and also low power consumption device.

Pavan Hegde et al., [2] developed a low cost, reliable, portable, user friendly, low power and robust Smart Glasses for visually disabled people for smooth navigation. The system is cost effective and can be worn easily as a glass. The device has an in-built sensor in it which spreads ultrasonic waves in the direction the person is going and it can scan at most 5-6 meters of range. The earlier version of this systems used basic image processing and computer vision techniques, and further developments were made to find a safe path for user navigation. When the obstacle is detected, the sensor detects it and sends it to the device which generates an automatic voice through the earphone connected to the user.

Wu Tang et al., [3] proposed a system for detection and recognition of outdoor obstacles for blind people. A new obstacle detection dataset OD was created for their system which contains 15 common objects, which was considered as a new benchmark for outdoor obstacle detection. To find the best model for object detection, three object detection algorithm were used in their work such as YOLO, SSD and Faster RCNN and it was found that the YOLO algorithm performed best among the three algorithms from the evaluation results.
Muhammad Sheikh Sadi et al., [4] implemented a spectacle prototype to assist the visually impaired people. The walking guide used ultrasonic sensors to identify the obstacles in different directions. The system can also detect potholes on the road surface using another ultrasonic sensor and Deep Learning algorithm, Convolutional Neural Network. The CNN runs on an embedded controller to identify obstacles on the surface of the road. Images were trained initially using a CNN on a host computer and then used as classifier on the embedded controller in real-time. The experimental analysis reveals that the system has 98.73% accuracy for the distances detected using ultrasonic sensor when the obstacle is at 50 cm distance.

A. Annapoorani et al., [5] proposed a Blind Sight-Object Detection system that is based on computer vision and state of the art object detection techniques. The system tries to automate tasks that the human visual system can do. Image classification techniques are used to identify the features of the image and categorize them into their appropriate class. The COCO dataset used in this project consists of around 123,287 hand labelled images classified into 80 categories. This wide set of data is used to describe spatial relationships between objects and their location in the environment. You Only Look Once is a Object Detection framework that is used to detect object in this system. An Indian currency recognition module was also developed to identify the denominations of the currency.

Dola Das et al., [6] developed a system that includes Ultrasonic Sonar sensors, three Passive Infrared sensors (PIR sensor), Raspberry Pi, SD card, vibrator and headphones and all equipment was built into a mobile cane. The PIR sensors allow the system to detect movable obstacles like walking people, vehicles etc., around the user in all the directions. The Ultrasonic Sonar sensors measures the distance and velocity of the movable object from the cane. Then the system notifies the user using vibration and audio message depending on the direction of the coming objects, distance and velocity. The author suggests that it can be used in real life also.
Saumya Yadav et al., [7] proposed an assistive device for visually impaired people to provide them guidance in automatic navigation. The whole setup for the visually impaired people consists of the set of sensors and computation components such as ultrasonic sensors, camera, single-board DSP processor, a wet floor sensor, and a battery. The machine learning model was used for object recognition to make the user familiar with the environments. To provide the best navigation guidance, proposed device was used to detect obstacles, staircases, potholes, speed breakers, wet floors, and narrow passages. The output was provided in the form of the vibration, to alert the user about the obstacle.

Mansi Mahendru et al., [8] proposed a system that will detect every possible day to day objects and prompt an alert to the person about the near as well as farthest objects around them. The system was developed using two different versions of the algorithms that is YOLO and YOLOv3. Both the versions were analysed in various situations to measure accuracy of the algorithm in every possibility. Both the YOLO and YOLOv3 were tested under same criterion to measure the accuracy and performance of each of them. After the experimental analysis the results revealed that the YOLOv3 is much more powerful than YOLO in detecting small objects and distant objects.

V.N. Honmane et al., [9] proposed an object detection system for the blind using Deep Learning technologies. Also, a voice guidance technique is used to inform visually impaired persons about the location of objects. The object recognition deep learning model used the You Only Look Once algorithm and a text-to-speech synthesizer is used to make it easier for the person to get information about the objects. The object-detection system can be used to find the objects in a specific space without help from others, and the system was also verified through experiments.

Raghad Raied Mahmood et al., [10] proposed a real-time object detection system to help visually impaired people in their daily life. The system used an object detection framework to detect objects more reliably and independently and also identify objects with the fixed location of an object in the x-axis, y-axis of an image. Then the label of
the object is identified and converted into audio by using Google Text to Speech, which was used to make the system more reliable and user friendly.

2.1 SUMMARY OF THE SURVEY

The above cited papers were useful in visualizing and understanding the flow of the entire project and also in designing its main architecture, the input and its format to be given for each module along with its expected output is identified. The use of ultrasonic sensors for obstacle detection and also for finding the distance between the user and the obstacle is adopted in this project. The YOLO algorithm which is an object detection framework is also used in this project for object detection. The above cited paper does not have a system or device that combines all the properties of the proposed system that is to detect the obstacle and find its distance and also detect the object category and then notify the user through a audio feedback. A visual aid can be developed using these techniques and also be used as an effective supporting aid for visually impaired people.
CHAPTER 3
SYSTEM DESIGN

3.1 PROPOSED SYSTEM

The proposed system aims to build a visual aid for visually impaired individuals with a voice feedback system in order to help them in their safe mobility. The system captures the real time environment and provide auditory feedback to the user through a headphone.

Raspberry Pi 3 Model B+ was chosen as the functional device because of its low cost and high portability. It also offers a multiprocessing capability. To detect the obstacles and to categorize the objects, an ultrasonic sensor and Pi camera module is used. YOLO algorithm is used for detecting the objects and eSpeak, which is a compact open-source speech synthesizer gives the user an auditory feedback of the detected object.

3.2 SYSTEM ARCHITECTURE

The real-world data is collected using ultrasonic sensors and a pi camera. Then the data from sensors is processed in Raspberry Pi for obstacle detection and if the distance of the obstacle is less than the given threshold then the object is detected using the YOLO algorithm and then the detected object and its distance is given as audio feedback to the user else, the distance between the obstacle and the user is measured continuously until it goes below the threshold value.
3.3 DATA ACQUISITION

The Raspberry Pi 3 Model B+ has 40 general purpose input/output (GPIO) pins, to which the ultrasonic sensor is connected to find the obstacle and the distance between the user and the obstacle. There are four pins in the ultrasonic sensor that were connected to the Raspberry Pi’s GPIO pins. VCC was connected to pin 2 that is VCC in Raspberry Pi, GND to pin 6, the GND in Raspberry Pi, TRIG to pin 12 that is GPIO18, and the ECHO to pin 18 that is GPIO24. The data from ultrasonic sensor is retrieved by our program, in real time and used to detect the obstacle.

The data for object detection is collected through the Pi camera as a video stream. The Pi camera is attached to the Raspberry Pi module through the dedicated camera serial interface. The Pi camera, which has a fixed focus lens, has been designed to fit onboard into Raspberry Pi. It has the capability to capture 640 pixels × 480 pixels high quality video. The RGB data is retrieved by our program, in real time, and recognize objects from every video frame that is already known to the system.
3.4 MODULE DESCRIPTION

The system can be divided into three modules

➢ Module 1: Detect Obstacle and Its Distance
➢ Module 2: Object Detection
➢ Module 3: Generate Audio Feedback

3.4.1 Detect Obstacle and Its Distance

The data from the ultrasonic sensor (HC-SR04) is used to find the distance between the user and the obstacle. The ultrasonic sensor output (ECHO) will always give output as LOW that is 0 Volt, unless it is triggered, in which it will give output as HIGH that is 5 Volt. Therefore, one GPIO pin is set as an output to trigger the sensor and one as an input to detect the ECHO voltage change.

However, this HC-SR04 sensor requires a short pulse to trigger the module. This causes the sensor to start generating ultrasound bursts, at 4 kHz, to obtain an echo response. So, the trigger pulse is created that is, the trigger pin is set HIGH for 10 seconds and then set to LOW again. The sensor sets ECHO to HIGH for the time it takes for the pulse to travel the distance and the reflected signal to travel back. Once a signal is received, the value changes from 0 to 1 and remains in 1 for the entire duration of the echo pulse. The difference between the two recorded time stamps is calculated and from that the distance between the ultrasound source and the reflecting object can be calculated. The speed of sound depends on the medium and the temperature. In our proposed system, 34300 cm/s, which is the speed of sound at sea level, is used to measure the time stamps.
Figure 3.2 Flow Diagram to Detect Obstacle and Its Distance

So, the first step is to trigger the sensor, if any obstacle is detected then calculate its distance else send trigger again. The calculated distance is displayed to the user. For any obstacle within the given threshold value, the system identifies the object type and sets off voice feedback. The audio feedback informs the user about his or her distance from the obstacle, thereby, alerting the blind person and avoiding any potential accident.
3.4.2 Object Detection

In this project, YOLO algorithm is used for object detection. You only look once (YOLO) is one of the fastest object detection algorithm used in real time detection. YOLO algorithm initially has 53 layers of Darknet but for the detection purpose 53 layers are added, that is total 106 layers. The most interesting feature of YOLO algorithm is that it makes detections at three different scales at three different sizes and at three different places in the network by down sampling the input image dimensions by 32, 16 and 8 respectively. So, the shape of detection kernel is 1 x 1 x 255.

![Figure 3.3 Architecture of YOLO Algorithm](image)

The 82nd layer makes the first detection. The image is down sampled by the network for the first 81 layers, such that the layer has a stride of 32. If an image of dimension 416 x 416 is given as input it would produce a feature map of size 13 x 13. Using the 1 x 1 detection kernel one detection is made, that gives us a detection feature map of 13 x 13 x 255.
Then, the feature map from 79th layer, before being up sampled by 2 times to dimensions of 26 x 26 it is subjected to convolutional layers. This feature map is then concatenated with the feature map from 61st layer. Then to fuse the features from the earlier layers the combined feature maps is subjected to 1 x 1 convolutional layers. Then, the 94th layer makes the second detection, producing a feature map of 26 x 26 x 255.

Similarly, the feature map from 91st layer is subjected to convolutional layers before being concatenated with a feature map from 36th layer. Then it is followed by 1 x 1 convolutional layers to combine the information from the previous layers. Then the final detection is made at the 106th layer, producing a feature map of size 52 x 52 x 255. The 13 x 13 layer is responsible for detecting large objects, the 52 x 52 layer is responsible for detecting the smaller objects, and the 26 x 26 layer detects the medium objects. Each class score is predicted using logistic regression and prediction of multiple labels for an object is done using a threshold value. The Classes which has the highest scores than the threshold is assigned to the object.

Figure 3.4 Object Detection in Raspberry Pi
To detect objects in real time Pi camera is used. For object detection, every object must be localized in each frame of a video input. The paths to the model and label maps are set, the model is loaded into the memory, the Pi camera is initialized, and then the object detection is performed on each video frame from the Pi camera. The system detects the common objects inside the view of the user.

Algorithm 3.1 Object Detection using YOLO

**Input**: Real time video stream from Raspberry Pi  
**Output**: Label of the detected object

1. Load the YOLO model.  
2. Read the input stream frame by frame.  
3. Initialize the class_id and confidence score.  
4. Find the dimensions of current frame.  
5. Detect blobs within the frame.  
6. Set the blob as input to the YOLO model.  
7. Class_id is set to index of maximum value in the list of objects.  
8. Confidence score is set to the highest value from the list.  
9. If the confidence score $> 0.6$, then all the object’s class_id and their respective confidences of blobs are stored as a list.  
10. Print the label of object detected with its confidence score.

**Pseudocode for Object Detection**

1. Load the YOLO model  
2. Read the input frame  
3. while True  
4. frame_no+=1  
5. class_id = []  
6. confidence = []
height, width = frame.shape[]

blob = cv2.dnn.blobFromImage()

setInput(blob)

for i in indexes
  for each i th value having class label class_id[i] do
    class_id = max (scores)
    confidence = score(class_id)
    if confidence > 0.6
      class_ids.append(class_id)
      confidences.append(float(confidence))
  end if
end for
end for

for i in class_ids
  return label, confidence
end for
end while

---

### 3.4.3 Generate Audio Feedback

The eSpeak, which is a compact open-source speech synthesizer (text-to-speech) for the Raspberry Pi, is used for auditory feedback for object type and distance between the object and the user. Along with eSpeak, pyttsx3 a cross-platform text to speech library, is also used in our system. It is compatible with both Python version 2 and version 3. The major advantage of using this library for text-to-speech conversion is that it works even in offline.

The first step is to initialize the text-to-speech engine using pyttsx3.init() function. Then to convert the object detected and its distance into audio form use the say() function. The runAndWait() function is used to make the speech audible in the system otherwise the speech will not be audible to the user. The audio feedback
generated is used to inform the user about his or her distance from the obstacle, thereby, alerting the blind person and avoiding any potential accident.

### 3.5 HARDWARE AND SOFTWARE REQUIREMENTS

**HARDWARE**

- Raspberry Pi – Model 3 or above
- Ultrasonic sensor
- Raspberry Pi camera
- Jumper wires
- Headphones
- Battery

**SOFTWARE**

- Operating system - Raspian OS
- Programming language – Python
- Python IDE - Thonny IDE
CHAPTER 4

IMPLEMENTATION AND RESULT

The system is implemented in the Raspberry Pi 3 Model B+ with ultrasonic sensor, an RPi camera, headphone and power supply for the Pi. Each part of the system was evaluated individually and the overall system is tested after assembling all the parts. The developed system is evaluated both for obstacle and object detection. The system is tested in real time to evaluate the performance of the system.

4.1 DETECT OBSTACLE AND ITS DISTANCE

The data is collected from ultrasonic sensor by positioning obstacles in different orientations. For each interval, we have taken data for five times and calculated the average value of these data. We have also estimated the accuracy, error rate, standard deviation and variance from the observed data. The comparison between actual distance and observed distance for the sensor is done and the results represents the distortion of the observed distance to the real distance.

<table>
<thead>
<tr>
<th>ACTUAL DISTANCE</th>
<th>OBSERVED DISTANCE</th>
<th>AVERAGE</th>
<th>ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>24.8</td>
<td>24.5</td>
<td>24.9</td>
</tr>
<tr>
<td>50</td>
<td>49.0</td>
<td>49.2</td>
<td>49.9</td>
</tr>
<tr>
<td>100</td>
<td>99.0</td>
<td>98.7</td>
<td>98.0</td>
</tr>
<tr>
<td>150</td>
<td>148.6</td>
<td>147.8</td>
<td>148.2</td>
</tr>
<tr>
<td>200</td>
<td>198.8</td>
<td>198.6</td>
<td>199.0</td>
</tr>
</tbody>
</table>
Table 4.2 Computation of Error, Standard Deviation and Variance

<table>
<thead>
<tr>
<th>ACTUAL DISTANCE</th>
<th>AVERAGE OF OBSERVED DISTANCE</th>
<th>ERROR</th>
<th>STANDARD DEVIATION</th>
<th>VARIANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>24.58</td>
<td>1.68</td>
<td>0.21</td>
<td>0.07</td>
</tr>
<tr>
<td>50</td>
<td>49.30</td>
<td>1.40</td>
<td>0.36</td>
<td>0.13</td>
</tr>
<tr>
<td>100</td>
<td>98.67</td>
<td>1.33</td>
<td>0.53</td>
<td>0.17</td>
</tr>
<tr>
<td>150</td>
<td>148.22</td>
<td>1.18</td>
<td>0.53</td>
<td>0.28</td>
</tr>
<tr>
<td>200</td>
<td>198.18</td>
<td>0.91</td>
<td>0.78</td>
<td>0.61</td>
</tr>
</tbody>
</table>

The deformity is shown to be not severe, and the observed distance is acceptable. The value of the distortion from the actual distance rises in a positive approach with the increment of actual distance. The accuracy is increased with the increase of distance.

From the table data it can be observed that with the increase of distance, the error rate becomes low. Standard deviation and variance are two closely associated measures of deviation. The variance is the measure of how much each value varies from the mean. The value of the variance represents the greater data range in the overall system. From the table data it can be observed that the lower values of standard deviation and variance are achieved when the hindrances are very near to the users. These values are increased with the increase of obstacles’ distance. The lowest standard deviation and variance that are obtained by the sensor of the system are 0.21 and 0.07 respectively.
4.2 OBJECT DETECTION

The Pi camera will capture real-time video stream, then the video is split into frames and each input frame is read to get its width and height. Object detection is done using YOLO Object Detection Algorithm in this system. To get the correct prediction the OpenCV function blobFromImage() is used to get the blob of the input frame, then it is sent to the YOLO pre-trained model. For each detection from each output layers, we will get the class label and confidence score. Then ignore the object whose confidence score is less than 0.6, and apply non-max suppression to determine the final detection. The class probability and confidence score together will help in determining the object accurately.
Figure 4.2 Real Time Object Detection Using Cell Phone as Object

Figure 4.3 Output of Distance and Object Detected as Cell Phone

```python
if __name__ == '__main__':
    try:
        dist = distance()
        print(f"measured distance = {dist} cm")
        if dist < 50.0:
            with open('/home/pi/objectdetect/coco80.names', 'r') as f:
                classes = [line.strip() for line in f]
            od = objectdetect()
            if (od not in classes):
                engine = pyttsx3.init()
                engine.say(f'object at distance {dist} centimeter')
                engine.runAndWait()
                time.sleep(1)
            else:
                engine = pyttsx3.init()
                engine.say(f'object at distance {dist} cm (od, dist)')
                engine.runAndWait()
                time.sleep(1)
        except KeyboardInterrupt:
            print('measurement stopped by user')
```
Figure 4.4 Object Detection Using Cell Phone as Object at Different Distance

Figure 4.5 Output of Different Distance Measure for Cell Phone
Figure 4.6 Real Time Object Detection Using Orange as Object

Figure 4.7 Output of Distance and Object Detected as Orange
The audio feedback is generated to inform the user about his or her distance from the obstacle, and the object type, which will help the visually impaired to move safely in their environment without depending on others.
CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

The main aim of this project is to develop a computer vision-based system to help visually impaired people to navigate independently in their environment. The obstacle detection system is designed to indicate the presence of obstacles in the front of a user in their surroundings. The ultrasonic sensor is used to find the distance between the user and the obstacle. The system detects the common objects captured by Pi camera using the YOLO algorithm. The output is provided in the form of the audio feedback instead of the vibration, to make it user-friendly. Thus, the proposed system provides proper guidance about the object and obstacles in front of the visually challenged user, which is not possible in conventional guidance devices and canes.

5.2 FUTURE WORK

In future, the system can be further developed to detect the presence of potholes, staircase, wet-floor and speed breakers. Provide the visually impaired person with the detailed description of the scene in front of them so that they will know about their surroundings, any potential danger approaching them, also to help them to take next decision based on their current situation. Also, to provide guidance based on the environment for the visually impaired person to help them in their mobility that is specify the direction in which the user has to take the next step and move around safely in their indoor and outdoor environment.
REFERENCES


