VISVESVARAYA TECHNOLOGICAL UNIVERSITY
BELGAUM-590014

A Project Report
On

Smart Assistance For Visually Impaired And Blind People

A Project report submitted in partial fulfillment of the requirements for the VIII Semester degree of
Bachelor of Engineering in Computer Science and Engineering
of Visvesvaraya Technological University, Belgaum.

Submitted by

NEHA N	**1DT18CS071**
RIA SATHYA	**1DT18CS079**
SKANDANA G	**1DT18CS097**
YASHASWINI R	**1DT18CS119**

Under the Guidance of

Dr. V SIVAKUMAR
Associate Professor
Computer Science and Engineering
Dayananda Sagar Academy of Technology & Management

Department of Computer Science and Engineering
DAYANANDA SAGAR ACADEMY OF TECHNOLOGY AND MANAGEMENT
Udayapura, Kanakapura Road, Bangalore-560082

2021-2022

DAYANANDA SAGAR ACADEMY OF TECHNOLOGY AND MANAGEMENT

Udayapura, Kanakapura Road, Bangalore-560082

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

Certified the project work entitled **"Smart Assistance for Visually Impaired and Blind People"** carried out by **Ms. Neha N** bearing USN **1DT18CS071**, **Ms. Ria Sathya** bearing USN **1DT18CS079, Ms. Skandana G** bearing USN **1DT18CS097** and **Ms. Yashaswini R** bearing USN **1DT18CS119** bonafide students of **Dayananda Sagar Academy of Technology and Management** in partial fulfillment for the award of Bachelor of Engineering in **Computer Science and Engineering** of the Visvesvaraya Technological University, Belgaum during the year 2021-22. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said Degree.

Dr. V Sivakumar	Dr. C. Nandini	Dr. M. Ravishankar
Associate Professor	Vice Principal & HOD,	Principal
Dept. of CSE, DSATM	Dept. of CSE, DSATM	DSATM

Sl. No.	Name	Signature with Date
1.		
2.		

DECLARATION

We, Neha N **[USN:1DT18CS071]**, Ria Sathya **[USN:1DT18CS079]**, Skandana G **[USN: 1DT18CS097]** and Yashaswini R **[1DT18CS119],** students of VIII Semester B.E., in Computer Science and Engineering, Dayananda Sagar Academy of Technology & Management hereby declare that the Project entitled "Smart Assistance for Visually Impaired and Blind People" has been carried out by us and submitted in partial fulfillment of the requirements for the *VIII Semester degree of* ***Bachelor of Engineering in Computer Science and Engineering*** *of Visvesvaraya Technological University, Belgaum* during academic year 2021-2022.

Sl. No.	Name	USN	Signature with date
1.	**NEHA N**	**1DT18CS071**	
2.	**RIA SATHYA**	**1DT18CS079**	
3.	**SKANDANA G**	**1DT18CS097**	
4.	**YASHASWINI R**	**1DT18CS119**	

ACKNOWLEDGEMENT

We are thankful to **Dr. M. Ravishankar**, **Principal**, DSATM, Bangalore, for providing the necessary facilities and for the encouragement to carry out the Project in DSATM.

We express our humble gratitude and thanks to **Dr. C. Nandini**, Vice Principal and HOD, Computer Science and Engineering, DSATM for guiding us and having facilitated us to complete the Project successfully.

We are very thankful to our guide **Dr. V Sivakumar**, Associate professor, Department of Computer Science and Engineering**,** DSATM for his constant encouragement, guidance and support throughout the course of the Project without whose efforts this would not have been possible.

We thank the Project Coordinators **Dr. Kavitha C & Dr. R Sarvana Kumar**, for their invaluable critical reviews, guidance and support in completing the Project on time.

We finally thank to all the Technical Staff for all the support and help while carrying out the project. Last but not the least we wish to thank our friends and parents who have been encouraging us constantly and inspiring throughout the execution of the project.

Name	**USN**
NEHA N	**1DT18CS071**
RIA SATHYA	**1DT18CS079**
SKANDANA G	**1DT18CS097**
YASHASWINI R	**1DT18CS119**

ABSTRACT

The idea is to develop a smart application which will be helpful for people with visual impairment. The aim of this research is to create a smart app that will assist persons with vision problems. People who are blind or visually handicapped are at a disadvantage in today's society since they are oblivious of their surroundings. The wider populace holds stereotypes about blind persons that they feel are accurate. Many people feel that blind persons are incapable of working or living a normal life. One of the most challenging obstacles is getting from one site to another without the help of others. Other difficulties include identifying people, detecting obstacles, and so on. Some commercially available products can help them overcome some of these challenges. There is constantly a slew of fresh research in the works, all with the goal of inventing solutions to help these visually impaired people. As a result, advanced new technology can be used to increase the amount of support supplied to them. The concept is achieved using a smartphone app that includes functions like as smart speaker, image processing, and more. The device is a voice-activated tool that aids people who are blind in doing daily tasks. The software incorporates a variety of opportunities for the developing into a management framework that is accessible to the visually impaired. The goal is to help blind or visually impaired persons navigate around obstacles.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Blindness is a misunderstood type of disability. The general public has preconceived views about blind persons that they sincerely believe are correct. The majority of non-blind individuals assume that blind persons are incapable of working or living a regular life. Blind people have their own way of doing things and live normal lives. They do, however, confront difficulties as a result of inaccessible infrastructure and social issues. They are affected not just by vision loss, but also by a decline in the general quality of their life, which has an impact on their level of independence. The country's population of visually impaired persons is so large that it has a considerable economic impact. In reality, in today's fast-paced society, the average person does not have time to notice these individuals with special needs. As a result, those who are blind or visually impaired require ongoing assistance in their daily activities. Getting from one area to another without the help of others is one of the most challenging undertakings. Being reliant on someone can lead to a sense of demotivation and, in certain situations, a loss of self-confidence. Other difficulties include recognizing people and sensing barriers in their path, among others. Some products on the market can assist them in overcoming some of these obstacles. There are always a bunch of new studies underway with the sole purpose of developing technologies to assist these visually impaired individuals. As a result, there is a need to create an application that can assist the visually impaired in detecting objects while on the move. The proposed technology, which is entirely dependent on voice commands, employs Artificial Intelligence to aid visually impaired persons. It can also recognize images from still images or video data.

CHAPTER 2

LITERATURE SURVEY

Paper TITLE & authors	Concept	Merits and limitation	Tool used/ specific algorithm / platform
"Smart Eye for Visually Impaired-An aid to help the blind people" Joe Louis Paul I, Sasirekha S, Mohanavalli S, Jayashree C, Moohana Priya P and Monika K. *ICCIDS-2019*	The device is a voice enabled system that would direct the visually challenged person in their day to day work.	• Two subsystems, route navigation and face detection and recognition. • This is helpful when the Internet connectivity is not available. • Not affordable for all. • It captures the object with a specific distance.	• Augmented Reality (AR) • Open Source Computer Vision (OpenCV) • voice enabled route navigation (GPS based) system with obstacle detection, image processing.
"Smart Assistance Navigational System for Visually Impaired Individuals" Mrs. S. Divya, Shubham Raj, M. Praveen Shai, A. Jawahar Akash and Ms. V. Nisha. *INCOS-2019*	This proposed methodology is used to determine the obstacle with help of ultrasonic waves and infrared waves produced by the respective sensors attached to a walking stick.	• It provides alert message and location during the time of emergency • Works on Battery which has a capacity around 12 hours. • Not affordable for all • Walking stick could get stuck in cracks and on uneven surfaces.	• Fine fiber Stick • Arduino Micro-Controller • Ultrasonic sensor • Buzzer and LED • GSM and GPS module
"Smart Device for Visually Impaired People" Kasthuri R, Nivetha B, Shabana S, Veluchamy M and Sivakumar S. *ICONSTEM -2017*	Uses Selendroid drives written in Selenium 2 API and use a speech recognizer.	• Converts spoken languages into machine understandable format very easily. • Google maps and Music system.	• Selendroid • WebDriver API • Speech Recognition Engine(SRE)

"Smart Cap for Visually Impaired Person using Raspberry Pi" Mrs. Subhasini Shukla, Shreya Pimple, Shraddha Shetke, Namrata Gaikwad and Krutika Patil. *ICSSIT 2020*	The Smart Cap proposed in this paper will help the visually impaired people in many ways, such as by describing the surroundings, recognizing familiar faces, reading out texts as well as providing the latest information via an online newspaper.	• It is a simple, cheap, and user-friendly smart assistant system is designed and implemented to improve the quality of life of both the blind and the visually impaired people • The facial landmark predictor fails to extract the necessary features.	• Text Recognition Module • Caption generation • Decoder with attentions • Image Captioning Module • Face Recognition Module
"Ultrasonic Stick for Blind" Agarwal A, D.Kumar, and A. Bhardwaj. *IEEE-2015*	This paper aims to design an artificial navigating system with adjustable sensitivity with the help of ultrasonic proximity sensor and a GPS module to assist these blind persons to walk fearlessly and independently in both indoor and outdoor environments.	• The detection covers most of the surroundings including of upcoming potholes • They use artificial vibrators as most of the time user may not able to differentiate between the sound levels in noisy environment. • Minimum obstacle thickness sensing capacity 3mm and minimum obstacle distance should be 2cm for detection. • Sensing accuracy can be affected by change of temperature of 10 degrees or more and can also be affected by soft materials. • The user has to habituated to the use of this system.	• Obstacle avoider • Ultrasonic sensor • GPS module

"Smart Cap for Visually Impaired Person using Raspberry Pi" Mrs. Subhasini Shukla, Shreya Pimple, Shraddha Shetke, Namrata Gaikwad and Krutika Patil. *ICSSIT 2020*	The Smart Cap proposed in this paper will help the visually impaired people in many ways, such as by describing the surroundings, recognizing familiar faces, reading out texts as well as providing the latest information via an online newspaper.	• It is a simple, cheap, and user-friendly smart assistant system is designed and implemented to improve the quality of life of both the blind and the visually impaired people • The facial landmark predictor fails to extract the necessary features.	• Text Recognition Module • Caption generation • Decoder with attentions • Image Captioning Module • Face Recognition Module
"Ultrasonic Stick for Blind" Agarwal A, D.Kumar, and A. Bhardwaj. *IEEE-2015*	This paper aims to design an artificial navigating system with adjustable sensitivity with the help of ultrasonic proximity sensor and a GPS module to assist these blind persons to walk fearlessly and independently in both indoor and outdoor environments.	• The detection covers most of the surroundings including of upcoming potholes • They use artificial vibrators as most of the time user may not able to differentiate between the sound levels in noisy environment. • Minimum obstacle thickness sensing capacity 3mm and minimum obstacle distance should be 2cm for detection. • Sensing accuracy can be affected by change of temperature of 10 degrees or more and can also be affected by soft materials. • The user has to habituated to the use of this system.	• Obstacle avoider • Ultrasonic sensor • GPS module

"Navigating Blind People with a Smart Walker" Andreas Wachaja, Pratik Agarwal Mathias Zink, Miguel Reyes Adame Knut Moller and Wolfram Burgard. *IROS -2015*	The contribution of this paper is a smart walker that enables blind users to safely navigate. It includes an innovative vibro-tactile user interface and a controller that takes into account human characteristics based on a user study.	• It calculates the ego motion of the walker based on the measurements of this sensor by laser scan matching • The algorithm used cannot guarantee that the path is not followed in the wrong direction in case that the robot is oriented in the wrong way	• Vibration motor • Height-length-density(HLD) classifier • Dijkstra algorithm

CHAPTER 3

PROBLEM STATEMENT & PROPOSED SYSTEM

3.1 PROBLEM STATEMENT

It's a known fact that the estimated number of visually impaired people in the world is about 285 million, approximately equal to 20% of the Indian Population. They suffer regular and constant challenges in Navigation especially when they are on their own. They are mostly dependent on someone for even accessing their basic day-to-day needs. So, it's quite a challenging task and the technological solution for them is of utmost importance and much needed. The problem statement is to come up with a system that will help them interact with the environment without anyone's aid and make them less dependent on others. The issues faced by blind persons were explored in this paper and attempted to give a satisfying answer for them in their daily lives. One such try from our side is that we came up with an Integrated Machine Learning System which acts as an Obstacle Detector and allows the visually impaired to identify and classify Real Time Based Common Day-to-day Objects and generate voice feedback and calculates approximate distance which produces warnings whether he/she is very close to the object. In short, our aim is to develop a system that can identify indoor and outdoor objects and notify the users.

3.2 SYSTEM ARCHITECTURE

The programme makes use of real-time images taken using a smartphone. It sends these images to a server, which then stores the images. These images are then analysed and broken into frames. In figure 3.1, the pre-trained model is fed with these frames in order to identify the object within them. The depth estimation algorithm is used to calculate the distance between the camera that acquired the image and the object within the frame. An audio message is generated and transmitted to the user after the object and distance are determined. This audio is played on the user's smartphone as shown in figure 3.2.

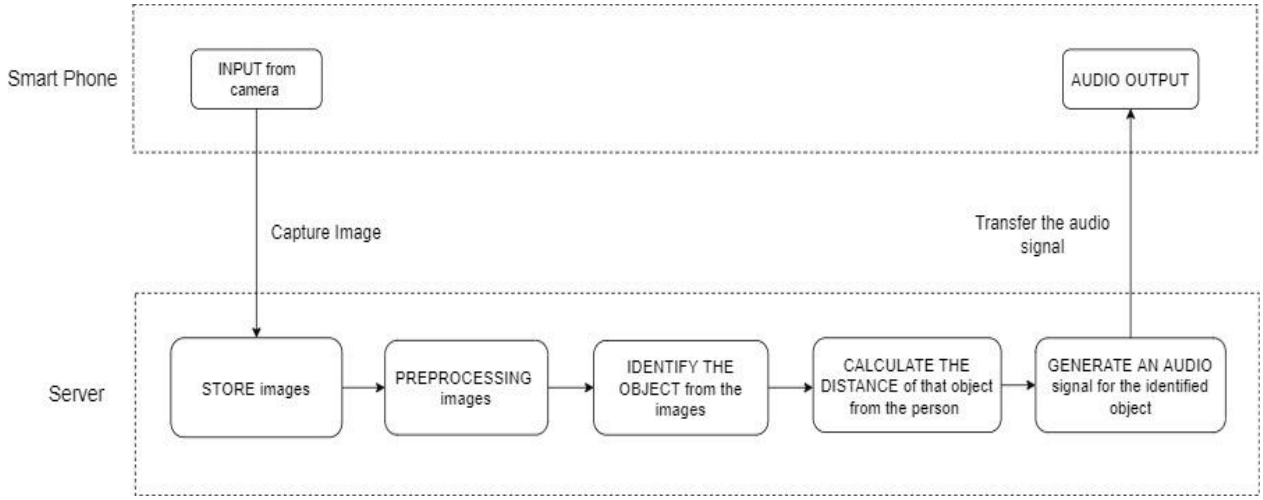

Figure 3.1 : System Architecture Flowchart

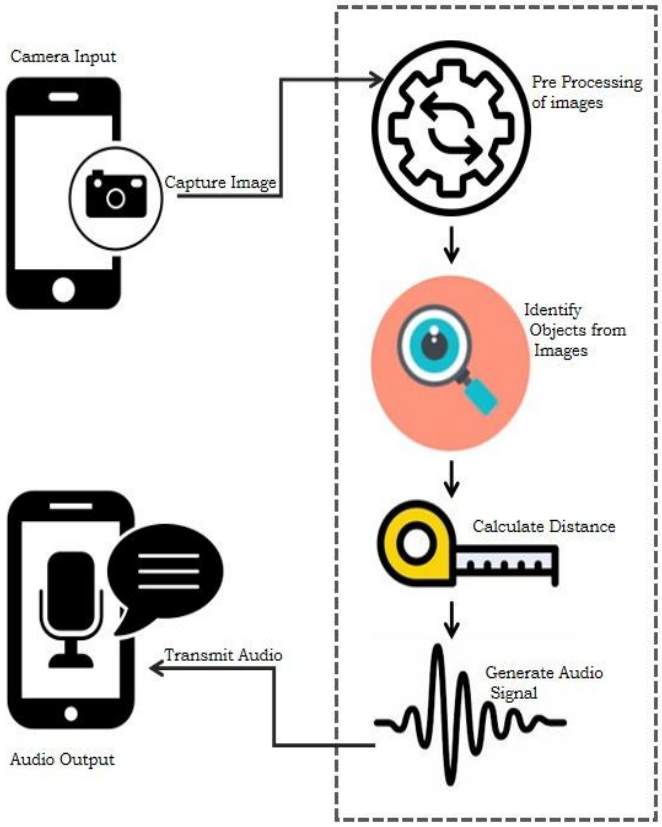

Figure 3.2 : System Architecture Diagram

3.3 PROPOSED SYSTEM

3.3.1 SCOPE OF THE PROJECT

In today's advanced hi-tech world, the need for independent living is recognized in the case of visually impaired people who are facing the main problem of social restrictiveness. They suffer in strange surroundings without any manual aid. Some of the major challenges include difficulty in moving from one place to another without the assistance of someone. Other challenges include difficulty in recognizing people, detecting obstacles, etc. Visual information is the basis for most tasks, so visually impaired people are at a disadvantage because necessary information about the surrounding environment is not available. With the recent advances in inclusive technology, it is possible to extend the support given to people with visual impairment. This project is proposed to help those people who are blind or visually impaired. The idea is implemented through an Android mobile application that focuses on object detection and voice assistant. This app provides voice assistance that would direct the visually challenged person in their day to day work. The device combines the various available technologies. The system detects the type of the object and provides the approximate distance of the object from the user. It also provides alert messages when the object is in close proximity. It will be an efficient way in which blind people can also interact with the environment with the help of technology and utilize the facilities of the technology

3.3.2 OBJECTIVE

- The system is set up such that the programme may capture actual pictures and transfer them to a pc that serves as a server for all of the computations.
- The primary objective of this project is to develop an Integrated Machine Learning System which acts as an Obstacle Detector.
- Object detection's main goal is to identify and find one or more effective targets in still imagery or video data.
- The basic goal is to identify indoor and outdoor objects and notify the users.

- It allows the visually impaired to identify and classify Real Time Based Common Day-to-day Objects and generate voice feedback and calculates approximate distance which produces warnings.

3.3.3 METHODOLOGY

- The system is structured so that the application may take real photographs and send them to a computer that acts as a server for all of the calculations.
- The laptop that serves as a server will employ a detection model that was trained utilising coco datasets. The model will test the output class and reliably detect it.
- The object's class will be turned into a collection of default voice notes after testing with speech modules, which will be delivered to the blind victims for assistance.
- We used an alert system that will calculate an estimate in addition to the object detection. It will generate voice-based outputs as well as distance units if the Blind Person is very close to the frame or in a safer position.

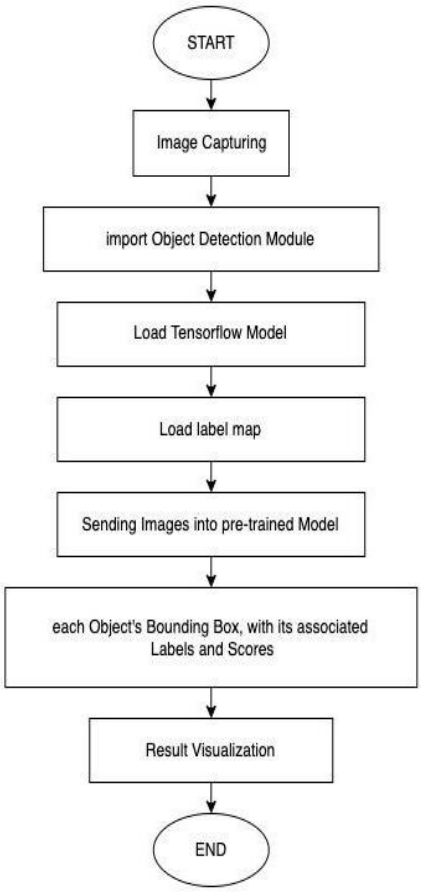

Figure 3.3 : Workflow of Object Detection Algorithm

3.4 SYSTEM REQUIREMENT SPECIFICATION

3.4.1 FUNCTIONAL REQUIREMENTS

A functional requirement is a requirement concerning a result of behaviour that shall be provided by a function of the system. A functional requirement describes a functionality to be made available to the users of the system, characterizing partially its behaviour as an answer to the stimulus that it is subject to.

The functional requirements include –

The model needs camera access permission to analyse the real-time images.

The system detects the various objects in the captured image and provides voice alerts.

3.4.2 NON FUNCTIONAL REQUIREMENTS

Non-Functional requirements are defined as something that describes the non-behavioural aspects of a system, capturing the properties and constraints under which a system must operate. A non-functional requirement corresponds to a set of restrictions imposed on the system to be developed, establishing, for instance, how attractive, useful, fast, or reliable it is. Non-Functional requirements define desired qualities of the system to be developed and often influence the system architecture more than functional requirements do.

The non-functional requirements include-

A. Availability Describes how likely the system is accessible for a user at a given point in time. Performance - Performance relative to the amount of resources used under stated conditions. The model is required to make prediction with minimum response time and minimum percentage error.

B. Scalability The system is required to be flexible and adaptive to growing and changing functionalities. This model should be scalable to adapt new features and functionalities.

C. Usability Degree to which a product or system can be used by visually impaired people to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. The model should be easily usable by the visually impaired.

D. Reliability The system is expected to make reliable prediction.

3.4.3 HARDWARE AND SOFTWARE REQUIREMENTS

HARDWARE:

The proposed system does not have a dedicated hardware interface. It is powered by an application server that's hosted on enterprise hardware in-house.

Device support: Desktop/Laptop systems with internet connectivity.

System Memory: 512 MB and above

SOFTWARE:

- Operating System:
    - Ubuntu 16.04 or higher (64-bit)
    - macOS 10.12.6 (Sierra) or higher (64-bit) *(no GPU support)*
    - Windows 7 or higher (64-bit)
- IDE: Anaconda
- Coding Language: Python
- Deep Learning APIs: TensorFlow

3.5 TOOLS AND LIBRARIES

A. Protobuf 3.0.0 l:

Protobuf 3.0 provides a code generator for C++, Java (including JavaNano, a dialect intended for low-resource environments), Python, Go, Ruby, Objective-C, C#. It also supports JavaScript since 3.0.0-beta-2.

B. Python-toolkit (python-tk)

Python-Toolkit (PTK) is an interactive environment for python. It was originally designed to provide a python based environment similar to Matlab for scientists and engineers when used together with the numpy, scipy and matplotlib python packages. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

C. Pillow 1.0

Python Pillow is built on the top of PIL (Python Image Library) and is considered as the fork for the same as PIL has been discontinued from 2011. Pillow supports many image file formats including BMP, PNG, JPEG, and TIFF. The library encourages adding support for newer formats in the library by creating new file decoders.

D. Spyder

Spyder is a very powerful and easy-to-use environment that is written in python, for python. What makes it special is the diverse functionalities that it offers. For example, easy debugging, advanced editing, aesthetically pleasing visualization techniques, interactive executions, and many more. In this article, you will get accustomed to some vital parts of the IDE and its key panes and functionality.

E. Matplotlib

Matplotlib is a low level graph plotting library in python that serves as a visualization utility. Matplotlib was created by John D. Hunter. Matplotlib is open source and we can

use it freely. Matplotlib is mostly written in python, a few segments are written in C, Objective-C and Javascript for Platform compatibility.

F. Tensorflow (1.15.0)

TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.

G. COCO API

COCO dataset is a large-scale object detection, segmentation, and captioning dataset published by Microsoft. Machine learning and Computer Vision engineers popularly use the COCO dataset for various computer vision projects.

CHAPTER 4

HIGH LEVEL DESIGN

4.1 DATA FLOW DIAGRAM

LEVEL-0 DFD:

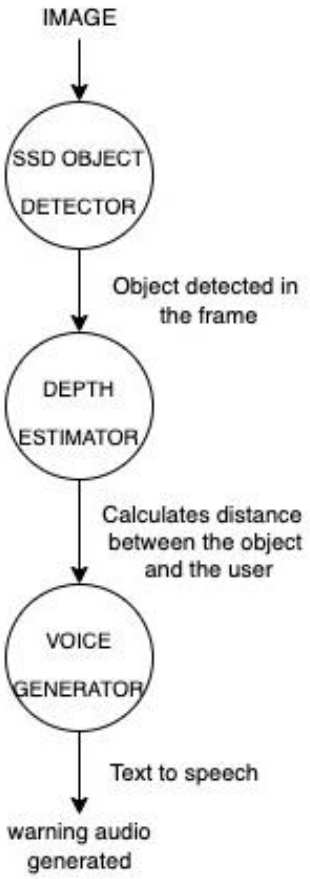

Figure 4.1: Level-0 DFD

Our project consists of 3 major modules:

1) Object detection module using SSD object detector

2) Depth estimator

3) Voice generation

LEVEL-1 DFD:

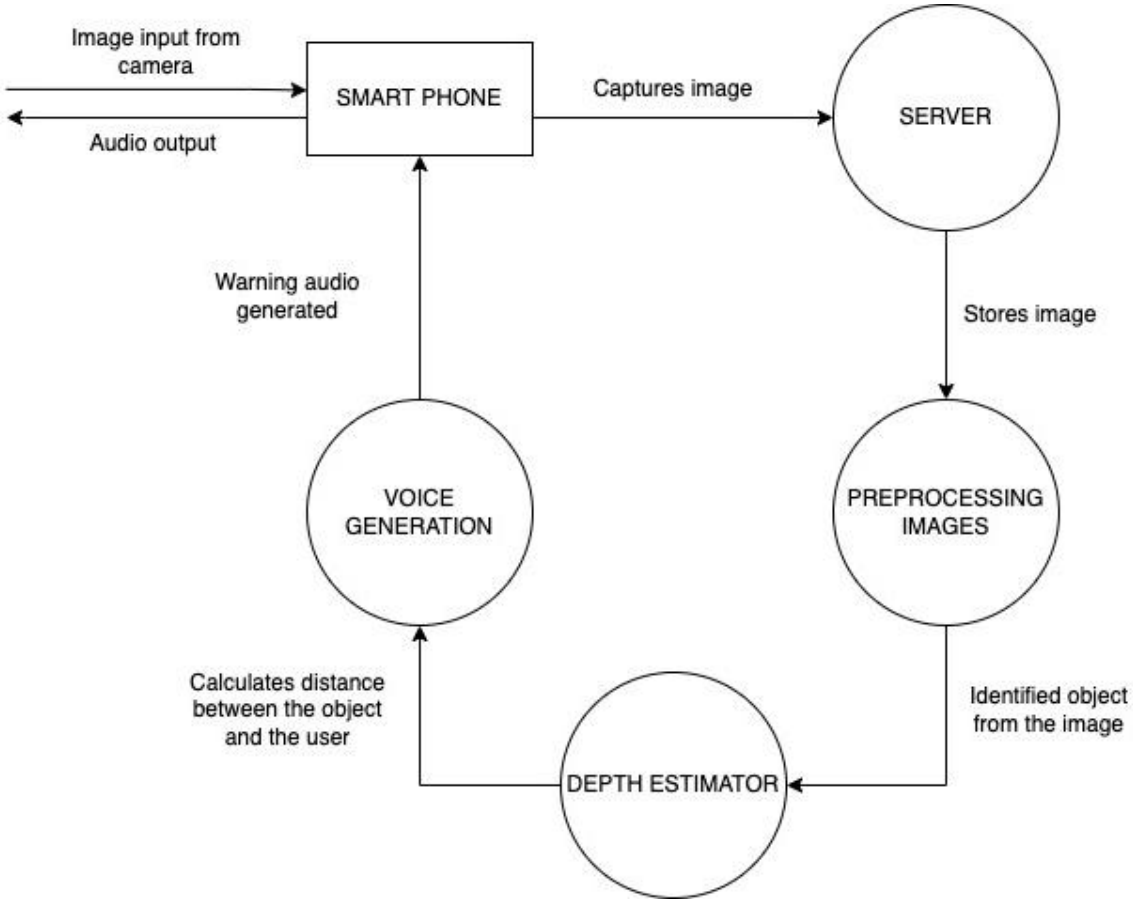

Figure 4.2 : Level-1 DFD

The smart phone captures the images from the camera, then the captured images are stored at the server, these stored images will be pre-processed, later they will be fed to a pre-trained model to identify objects from the images. Further, it calculates the distance between the user and the object to generate an audio signal for the identified object. The audio signal is sent back to the smart phone for the user to hear.

LEVEL-2 DFD:

Object Detection

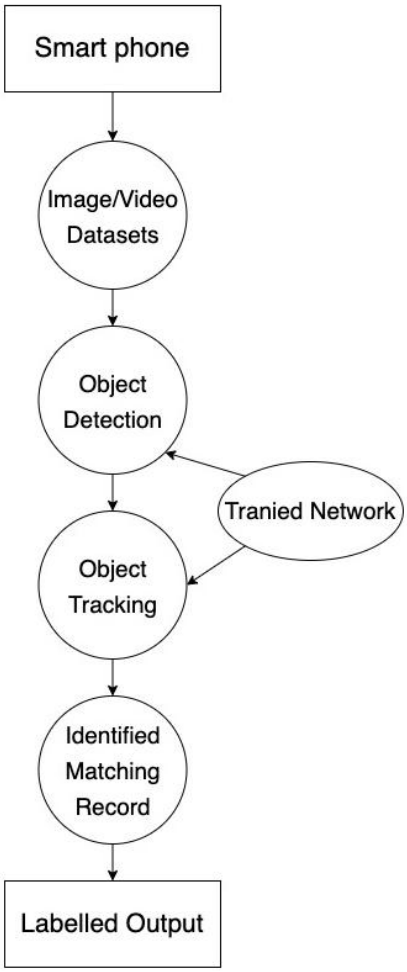

Figure 4.3 : Level-2 DFD: Object Detection

The smart phone captures the images which acts as the dataset for the already trained object detection module. Object detection mainly consists of object detection which detects the object within the image and object tracking which detects the position of the object within the image. After detection, it is compared with identified matched record for identifying the kind of object detected in the image. Once the object is

recognized it labels the object identified enclosed within a bounding box. The output of object recognition is further sent to the other modules.

DEPTH ESTIMATION

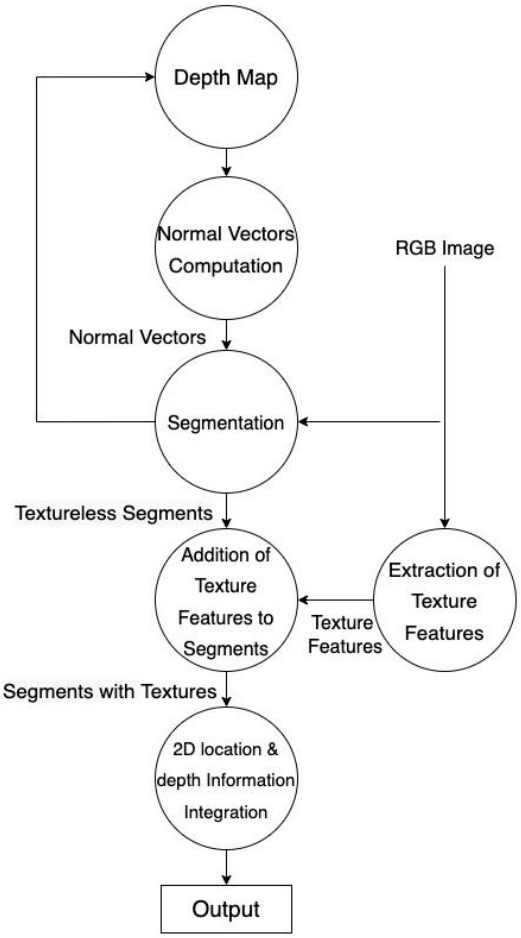

Figure 4.4 : Level-2 DFD: Depth Estimation

Depth is crucial when it comes to understanding the world around us such as navigation and mapping. To interact with the physical world around, depth is the deciding factor. The RGB image undergoes texture extraction feature and segmentation. It is very necessary to extract meaningful features from raw images that describes

texture properties. It can be defined as function of spatial variation of brightness intensity of pixels. Segmentation is a process of partitioning of a digital image into multiple image segments. Depth map creates a distance representation of an image from a reference point, it provides details of the depth based on how far or near the object is in terms of perspective. Texture segments and texture features are added together to give segments with texture which provides integration of 2D location and depth information. Using this relative distance of the object from the person is calculated and determined and finally produces an output.

Text to Speech

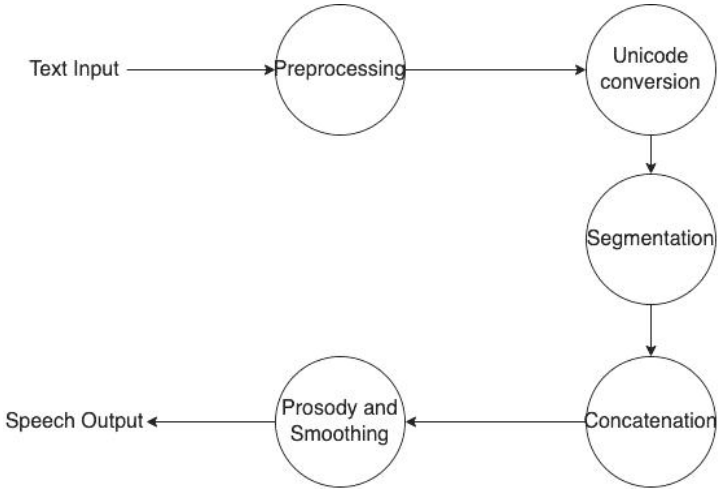

Figure 4.5 : Level-2 DFD: Text to Speech

PYTTSX3 is a text to speech conversion library in python. It is accomplished by starting with the method of pre-processing of an input text. Text, abbreviators, acronyms and numbers and expanded. Pre-processed text will then be converted into Unicode. Unicode has the explicit aim of transcending limitations of traditional character encodings. Now a text is segmented into syllables and duplicates are removed. The syllables are then concatenated and it further undergoes prosody and smoothening process. This finally gives a smooth human speech output. Variations can be applied to resulting output.

4.2 MODULES

TensorFlow is a free and open-source machine learning platform. It consists of a variety of tools, modules, and community activities that allow academics to gain machine learning competence while also allowing developers to swiftly build and deploy ML applications.

4.2.1 SSD ARCHITECTURE

Tensorflow comes with a number of models that have already been trained. SSD detection is utilised to deliver faster and more accurate results, depending on the system specs. A Single Shot Detector (SSD) is made up of two parts: one with an SSD head and another with a backbone. As a feature extractor, the backbone model is basically a learnt image categorization network. One or more convolutional layers are added to the backbone and read as bounding boxes and object classifications in the spatial position of the final layer activations in the SSD head. An image is divided into grid cells, each of which is responsible for identifying objects in its own zone. Object detection includes anticipating an object's class and placement inside a given area.

4.2.2 ANCHOR BOX

In SSD, each grid cell can have many anchor boxes linked to it. These predefined anchor boxes are individually accountable for a specific size and shape within a grid cell. SSD uses the matching phase during training to ensure an appropriate match between the anchor box and the bounding boxes of each actual data object within an image. The anchor box with the most resemblance to an object can predict the class and position of that object. This characteristic is utilized for training the system as well as predicting the identified objects and their locations once it is trained. Each anchor box is given an aspect ratio and a zoom level in practice.

4.2.3 ZOOM LEVEL

The shapes of all items are not the same. Some are shorter, longer, or both, while others are narrower, wider, and shorter, longer, or both. This is made possible by the SSD architecture, which allows the anchor boxes to have pre-determined aspect ratios. At each zoom level, the ratios parameter of the anchor boxes coupled with each

grid can specify several aspect ratios. The anchor boxes may or may not have to be the same size as the grid cell. The user may be looking for both smaller and larger items within a grid cell. The zooms option sets the amount by which each pixel's anchor boxes should be changed up or down.

4.2.4 DEPTH ESTIMATION

The methods and procedures used to obtain a depiction of a scene's spatial structure are referred to as depth estimation, also known as extraction feature. To put it another way, it's utilized to figure out how far two objects are apart. In order to aid blind persons with obstacles on their path, it is important to determine the distance between the obstruction and the person in any given situation. When an object is detected, a rectangular box is created around it. If the object takes up the majority of the frame, the approximate distance between the object and the person is determined, taking into account some constraints. Following the detection of an object, the algorithm is utilized to calculate the object's relative distance from the human.

4.2.5 VOICE GENERATION MODULE

After detecting an object is done, it is necessary to notify the individual about the existence of an object. pyttsx3 is an essential element of the voice generation module as shown in figure 4.6. The Python conversion module pyttsx3 transforms text to speech. Python 2 and 3 are both supported by this library. pyttsx3 is a straightforward text-to-speech converter. To find hidden text in an image, we use Python-tesseract for character recognition. Optical Character Recognition (OCR) identifies textual data on images and converts it to a computer-readable format. The content is detected by scanning and analyzing the image. As a result, Python-tesseract recognizes and "reads" text encoded in images. These texts are referred to a pyttsx.

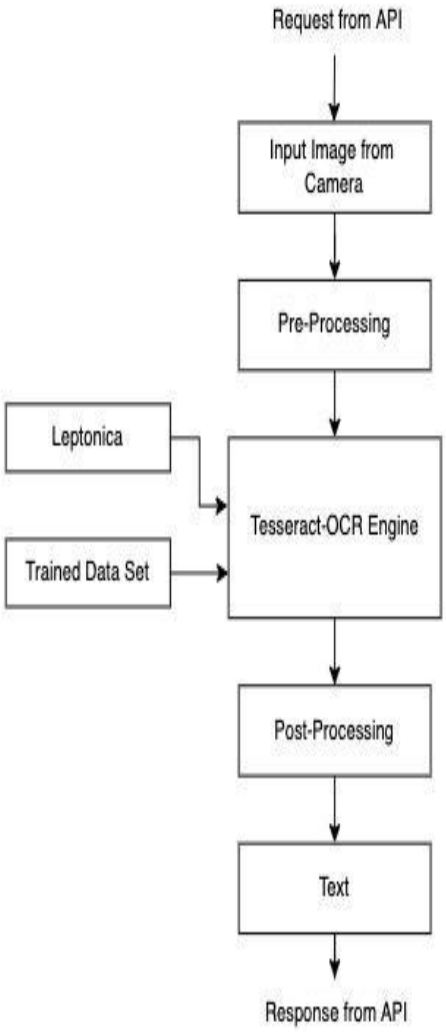

Figure 4.6 : Flowchart for OCR Process

CHAPTER 5

LOW LEVEL DESIGN

5.1 Object Detection

Input: input images through the network

Output: bounding box (x, y)-coordinates of each object in the image along with name of the detected object and their confident scores

```
for all vertical kernel positions
  for all horizontal kernel positions
    initialize the value at the output position to the
bias
    for all vertical positions in the kernel
      for all horizontal positions in the kernel
        add the product of the input value to that of
the output position
until number of boxes reaches max
  make first guess of two coordinates
  until number of guesses reaches max or matching
criteria is met
    evaluate guess
    remember guess and guess results
    improve on guess based on evaluation results and
          possibly injected randomness,
          excluding locations already covered
    if some intermediate criteria is met
      change the nature of the guessing, evaluation, and
improving
            as is appropriate for the criteria match
```

```
            (this covers approaches that have multiple
phases)
  if no guess matched criteria
    break
```

For a general understanding of how to create this single shot object detection technique, let's look at the pseudo-code implementation. The DNN (Deep Neural Network) module of OpenCV is used to load a pre-trained object identification network. As a result, we will be able to send input images over the network and generate the output bounding box (x,y) - coordinates for each object in the image. The code to print the name of the identified object and their confident scored is then written. Finally, we take a look at MobileNet Single Shot Detector's output for the images that served as our input.

5.2 Text-to-Speech Conversion

```
Input: Input text

Output: Smooth human speech output


repeat

{

   for each word

   {

       if word in base

           Generate speech

       else

           for each twoletters

           {
```

```
                    if twoletters in base
                        generate speech
                    else
                    {
                        for each letter
                            generate speech
                    }
                }
        }
}
```

Initially the whole text is converted to fit so that two-symbols are replaced with single characters. Abbreviations and numbers that are included in the database are converted with the texts and words as they are read/pronounced. The second step begins with the process of replacement of textual segments with corresponding sound files, which will be placed in a final sound file. Here, is applied algorithm, which takes the specific words of the text and examines them under the following procedure. If the word is found in the database, the appropriate sound file is taken and added to the final sound file. Otherwise, the word is fragmented into two-letter segments and for each segment is taken appropriate sound file and is placed in the final sound file. Similarly, is acted with numbers and abbreviations. If it happens that any digraph lacks in the database, sound files of separate characters/letters are used. We are analysing the texts in order to find the larger units of text segments composed of two or more words, so that the relevant sound files are used, what will certainly affect the growth of neutrality and clarity of the generated speech.

5.3 Depth Estimation Module

```
Input: RGB image

Output: Relative distance of an object or representation of
spatial structure of the scene or Distance Matrix


repeat

{

   for i,b in enumerate(boxes[0]):

      boxes[0][i][0] - y axis upper start coordinates

      boxes[0][i][1] - x axis left start coordinates

      boxes[0][i][2] - y axis down start coordinates

      boxes[0][i][3] - x axis right start coordinates

   end for

}
```

Further analysis iteration is done through the boxes. Boxes are an array inside of another array. The criteria for iteration is defined as shown above. The index of the boxes array is given the symbol i. By using an index, the box score is examined. Access to classes is another purpose for it. Now, the width of the identified object is determined. This is accomplished by requesting an object's width in pixels.

CHAPTER 6

IMPLEMENTATION

6.1 TENSORFLOW

The implementation is done by using TensorFlow APIs. TensorFlow object detection API is basically a structure built for creating a deep learning network that solves the problems for object detection.

import tensorflow as tf

There are trained models in their framework and they refer it as Model Zoo .This includes a collection of COCO dataset, the KITTI dataset, and the Open Images Dataset. Here, we use COCO DATASETS.There are a bunch of pre-trained models with Tensorflow . For a faster accuracy SSD DETECTION is being used.

MODEL_NAME = 'ssd_inception_v2_coco_2017_11_17'

SSD meaning, “Single Shot MultiBox Detector” divides the image using a grid and have each grid cell be responsible for detecting objects in that region of the image. Detecting objects basically means predicting the class and location of an object within that region. COCO dataset, meaning “Common Objects In Context”, is a set of challenging, high quality datasets for computer vision, mostly state-of-the-art neural networks.

6.2 DEPTH ESTIMATION

Depth estimation or extraction feature is nothing but the techniques and algorithms which aim to obtain a representation of the spatial structure of a scene. In simpler words, it is used to calculate the distance between two objects. The application is used to assist the blind people which aims to issue warnings to the blind people about the hurdles coming their way. In order to do this, we need to find how much distance the obstacle and person are located in any real time situation. After the object is detected, a rectangular box is generated around that object.

If that object occupies most of the frame then with respect to some constraints the approximate distance of the object from the particular person is calculated. Following code is used to recognize objects and to return the information of the distance and location.

```
(boxes, scores, classes, num_detections) = sess.run ([boxes, scores, classes, num_detections], feed_dict = {image_tensor : image_np_expanded})
```

Here, we have established a Tensorflow session comprising crucial features for Detection. So, for further analysis iteration is done through the boxes. Boxes are an array, inside of an array. So, for iteration we need to define the following conditions.

```
for i,b in enumerate(boxes[0]):

boxes[0][i][0] – y axis upper start coordinates

boxes[0][i][1] – x axis left start coordinates

boxes[0][i][2] – y axis down start coordinates

boxes[0][i][3] – x axis right start coordinates
```

Index of the box in boxes array is represented by i. Analysis of the score of the box is done by index. It is also used to access classes. Now the width of the detected object is measured. This is done by asking the width of an object in terms of pixels.

```
mid_x = (boxes [0][i][1] + boxes [0][i][3]) /2

mid_y = (boxes [0][i][0] + boxes [0][i][2]) /2

apx_distance = round (((1 - (boxes [0][i][3] - boxes [0][i][1]))**4) , 1)
```

In the above formula, mid_x is the center of the X axis and mid_y is the center of y axis. If the distance apx_distance < 0.5 and if mid_x > 0.3 and mid_x < 0.7 then it can be concluded that the object is too close to the particular person. With this code, the relative distance of the object from a particular person can be calculated. After the detection of an object the code is

used to determine the relative distance of the object from the person. If the object is too close then a signal or a warning is issued to the person through voice generation module.

6.3 VOICE GENERATION MODULE

After the detection of an object, it is utmost important to acknowledge the person about the presence of that object on the way. For the voice generation module PYTTSX3 plays an important role. Pyttsx3 is a conversion library in Python which converts text into speech.

Engine instance, a factory function called **pyttsx.init()** is invoked by an application. During construction, a **pyttsx.driver.DriverProxy** object is initialized by the engine which is responsible for loading a speech engine driver from the **pyttsx.drivers** module. After construction, an object created by an engine is used by the application to register and unregister event call-backs; produce and stop speech; get and set speech engine properties; and start and stop event loops.

Audio commands are generated as output. If the object is too close then it states "**Warning: The object (class of object) is very close to you. Stay alert!**". Else if the object is at a safer distance then a voice is generated which says that "**The object is at safer distance**". This is achieved with the help of certain libraries like **pytorch, pyttsx3, pytesseract and engine.io .**

CHAPTER 7

SYSTEM TESTING

In the area of app development, Third Party App offers simplicity and freedom. It increases effectiveness and facilitates quick output delivery. Third Party Software enables you to segment your work and aids in focusing on the essential elements of any system or app. This approach aids in the creation of high-quality software. We can give the system access to the Third Party App's features.

- The images are first delivered from the mobile phone to the laptop after building a connection between the two devices. We are initially taking real-time pictures from the back camera of the cellular phone of blind persons.
- A third-party app that the person has installed on their phone facilitates this connection. All of the real-time photographs that are taken with the mobile phone's rear camera are first transmitted to a third-party app on the device, and then those images are forwarded to a laptop where they are processed in order to draw additional conclusions.
- The laptop's system will test it using its APIs and SSD ALGORITHM, and it will determine the image's confidence accuracy. For other certain classes like scissors, we achieved a 100 percent accuracy as shown in Figure 6.5 .
- Following the testing of the images, we are producing an output on a laptop-based system, and our forecast is converted into voice using voice modules and communicated to the blind individual using wireless audio assistive devices.

7.1 Test Cases

Below are the screenshots of the test case taken during the system testing. Figure 7.1 shows a ValueError that was identified during the initial stages of testing. Figure 7.2 depicts the proper working of the model. Figure 7.3 explains distance calculated and shows its output on consoles and the blind person can hear these distance values along with object based warning. Figure 7.4 shows an umbrella is identified with an

accuracy of 64%. Figure 7.6 shows that the system is capable of detecting and recognizing multiple objects in a single frame. Figure 7.7 elucidates the alerts printed on the console along with the voice warnings. Figure 7.8 shows multiple objects are detected in the frame and the WARNING message is also displayed.

Figure 7.1 : An error (ValueError) occurred during the initial stages of testing

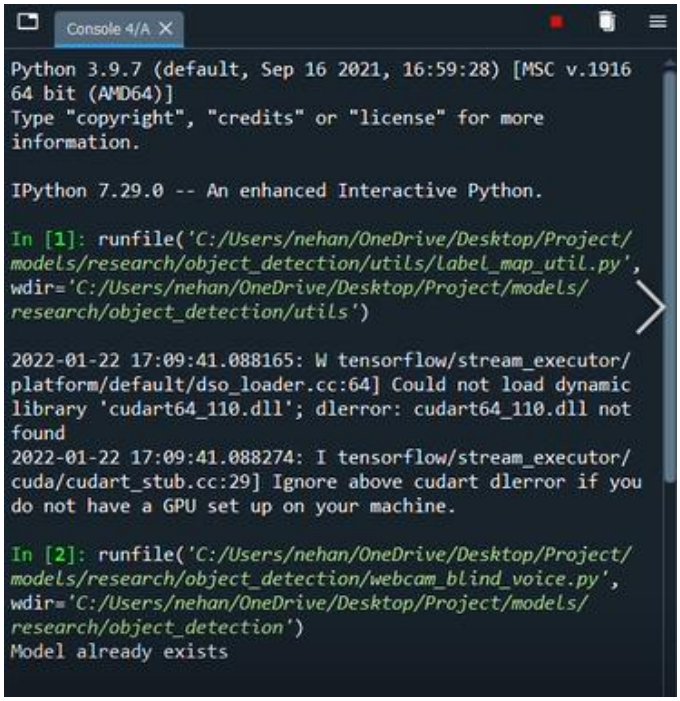

Figure 7.2 : The Model loaded correctly and showing the same on the console

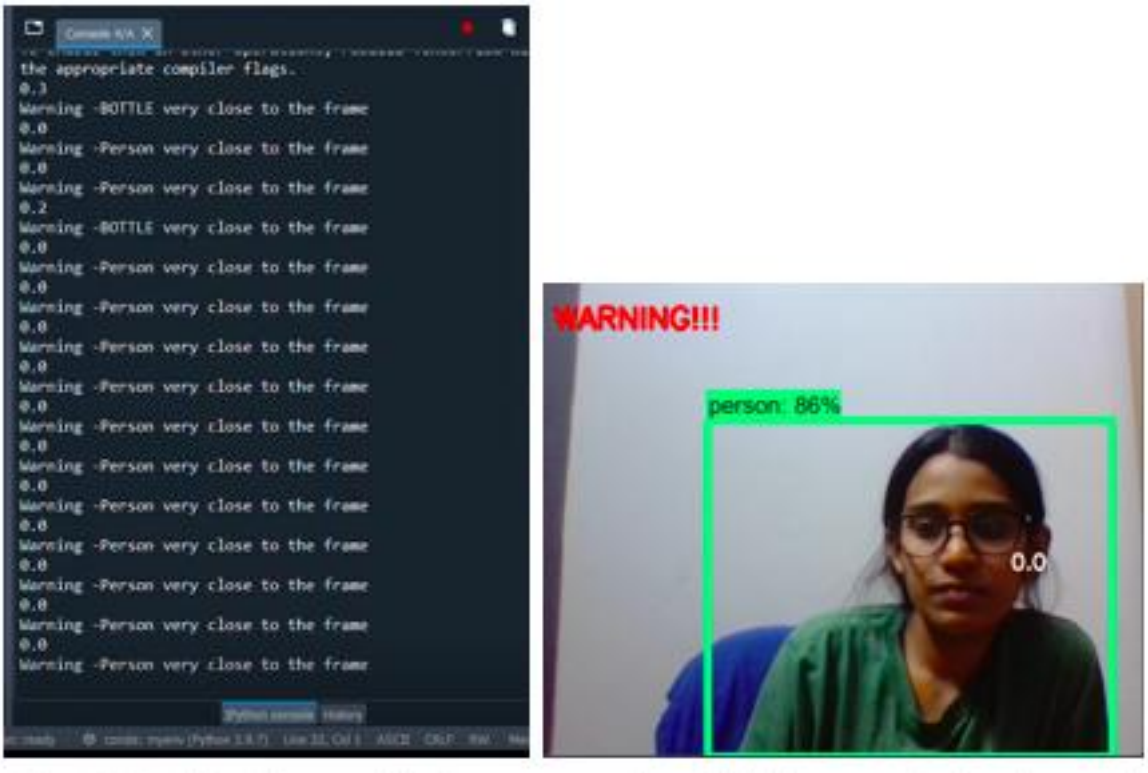

Figure 7.3 : It calculates distance and shows its output on consoles and the blind person can hear these distance values along with object based warning.

Figure 7.4 : The system identifies an umbrella with an accuracy of 64%

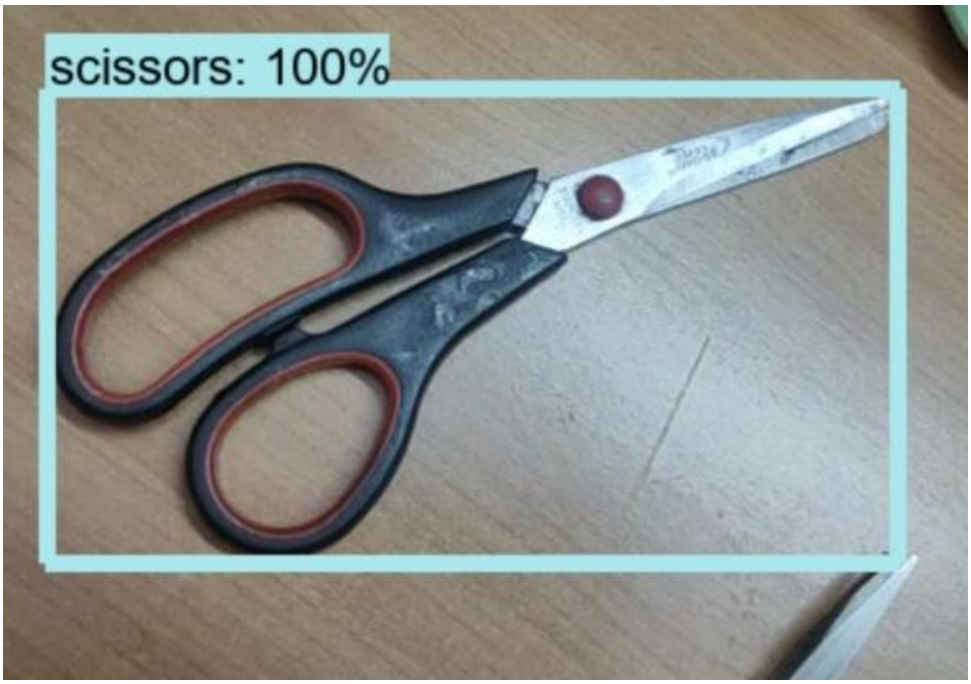

Figure 7.5 : Final accuracy of 100% was achieved for scissors

Figure 7.6 : The system can identify multiple objects in a single frame

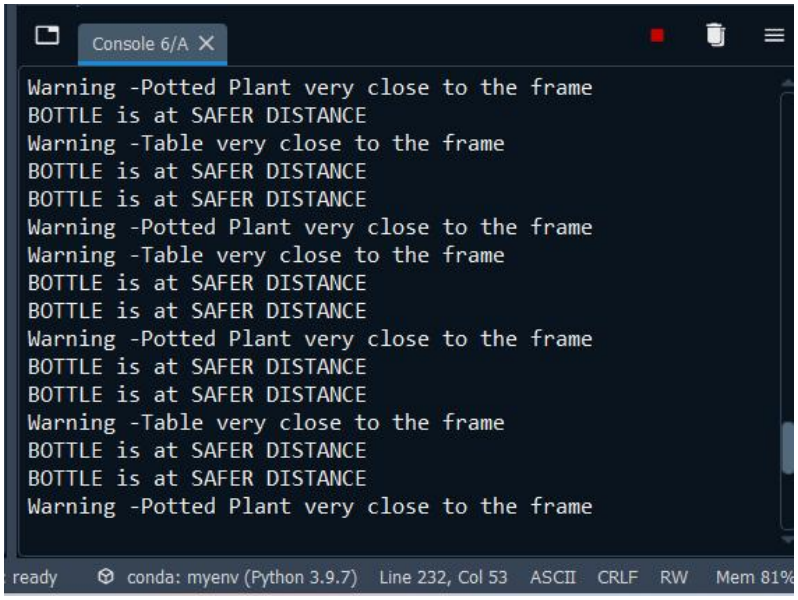

Figure 7.7 : Alerts generated on Console along with voice warnings

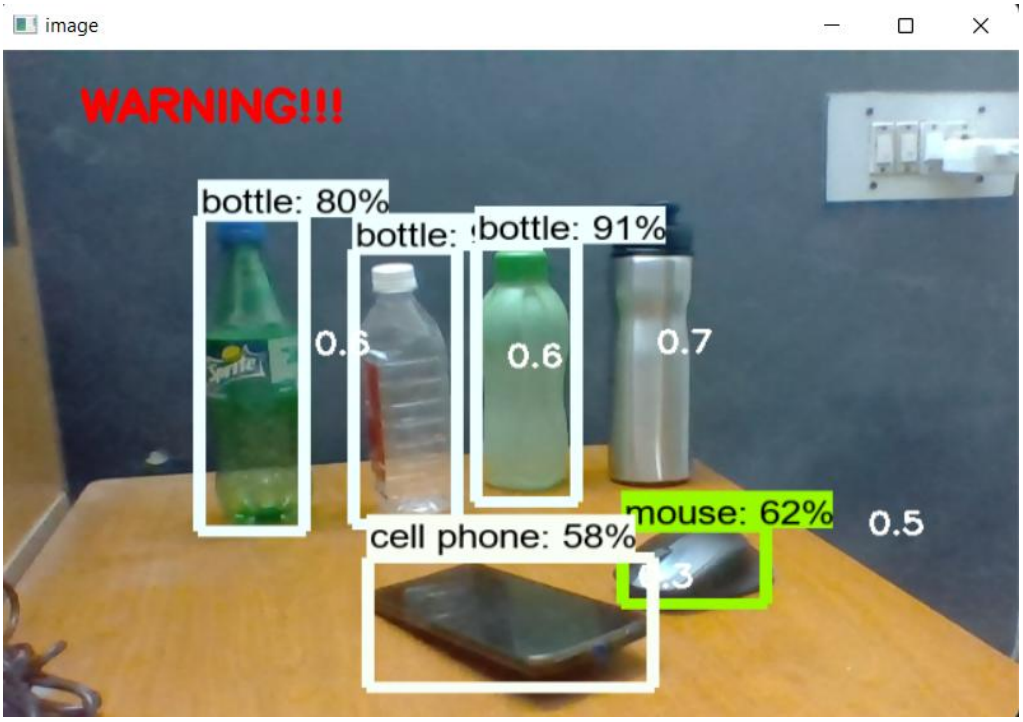

Figure 7.8 : Multiple objects detected along with the WARNING message

CHAPTER 8

CONCLUSION

The proposed system combines the capabilities of the various components to create a versatile smartphone for those who are blind or visually impaired. The application is simple to use because it makes use of cell phones, which are convenient and portable. Object identification and voice guiding are among the functions included in the programme, which is designed for visually impaired people. It contains a user-friendly interface developed exclusively for visually impaired people, as well as detection findings that are read out loud so the user can plainly hear them. As a result, their lives will be made easier. Our proposed solution might be implemented in a multilingual application, allowing users to use it in their local language with ease. Image recognition can be improved with more information about the image obtained. It also works with the Internet of Things.

CHAPTER 9

FUTURE SCOPE

Artificial Intelligence and machine learning is one of the most growing technology. These technologies are the playing vital role in the development of the IT sector. Here we have tried to use these technologies for the visually challenged people so that they can also live an independent and normal life. We can provide more accessibility to the user by further adding features such as text recognition and live translation. This will enable the user to interpret text detected in the image in front of them. The friendly chat with the bot Image recognition of the objects and surroundings. Currency recognition to help in the easy payment. The development of the proposed system if is completed, it can serve the visually challenged people with a better assistant. In coming days our proposed system can be applied in multilingual application so that a person can use the application in their own language without any trouble. In addition, our proposed system can be deployed with the IoT. In future our proposed system will be able interpret the textual description in a much better way. The Image recognition can be enhanced with much more details about the image captured through the camera. Enhancement to this system can be done by adding the features of currency recognition and route navigation. The existing methodology for image and currency recognition can be done with more accuracy.

CHAPTER 10

SCREENSHOTS

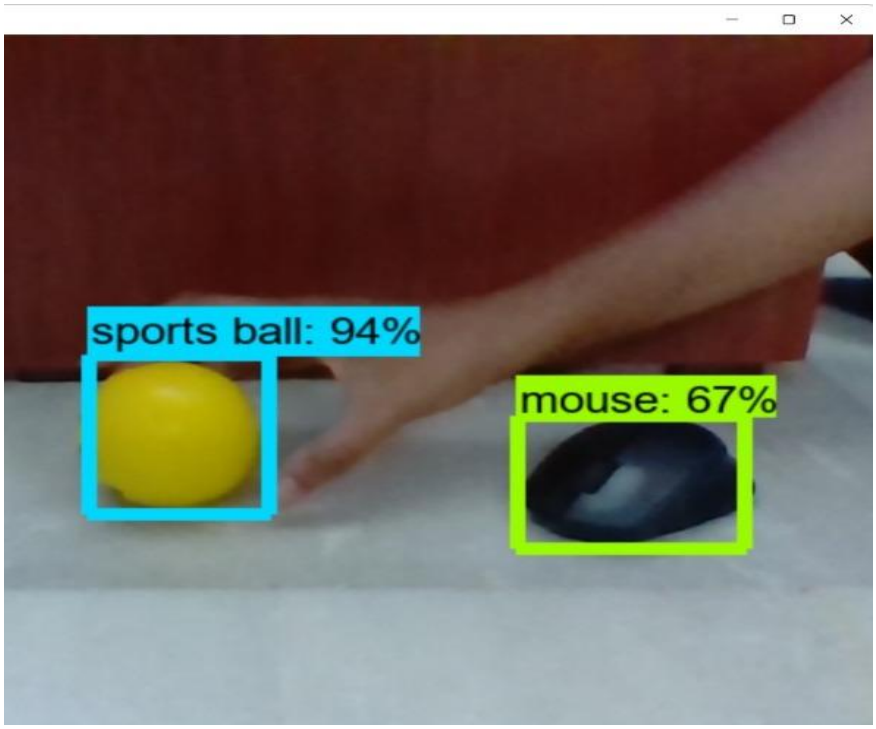

Figure 10.1 : System identifying sports ball and mouse with accuracy 94% and 67% respectively

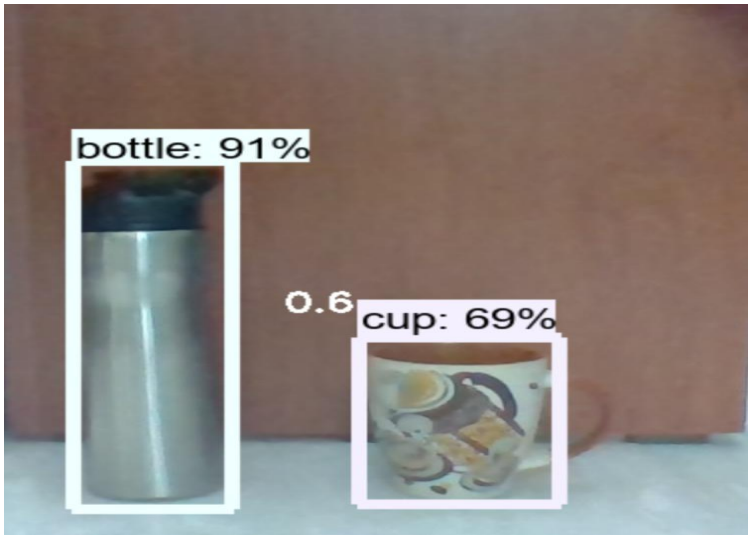

Figure 10.2 : System identifying bottle and cup with accuracy 91% and 69% respectively

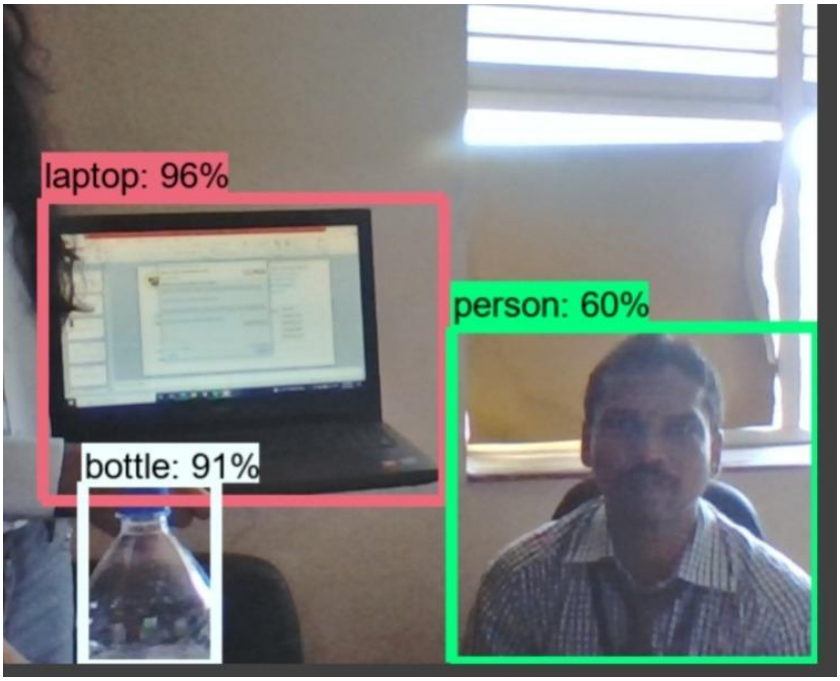

Figure 10.3 : System identifying laptop, bottle and person with accuracy 96%, 91% and 60% respectively

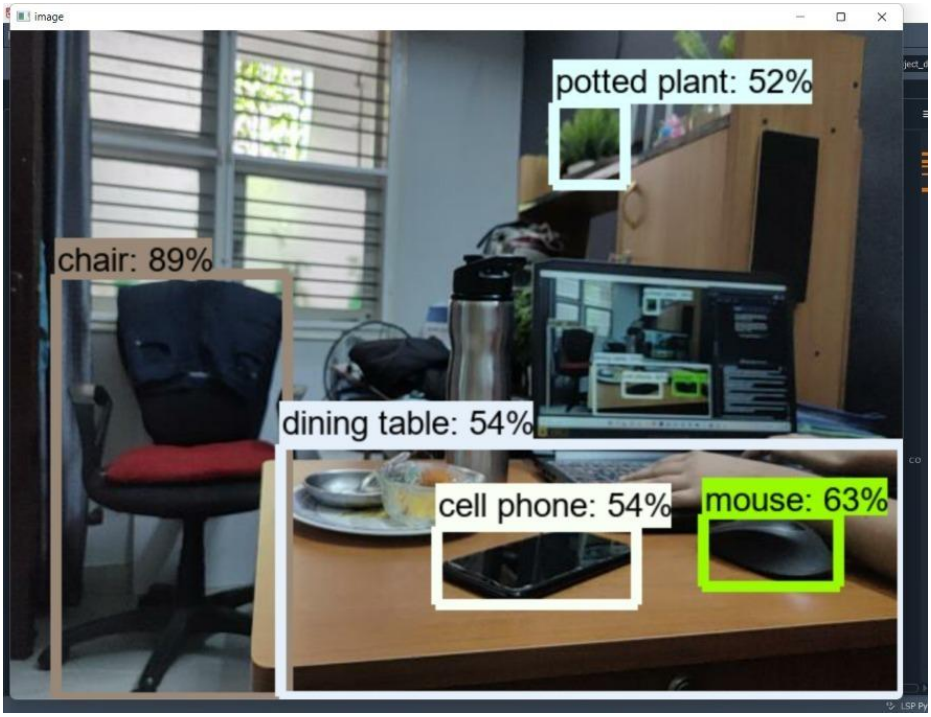

Figure 10.4 : System identifying chair, dinning table, cell phone, potted plant and mouse with accuracy

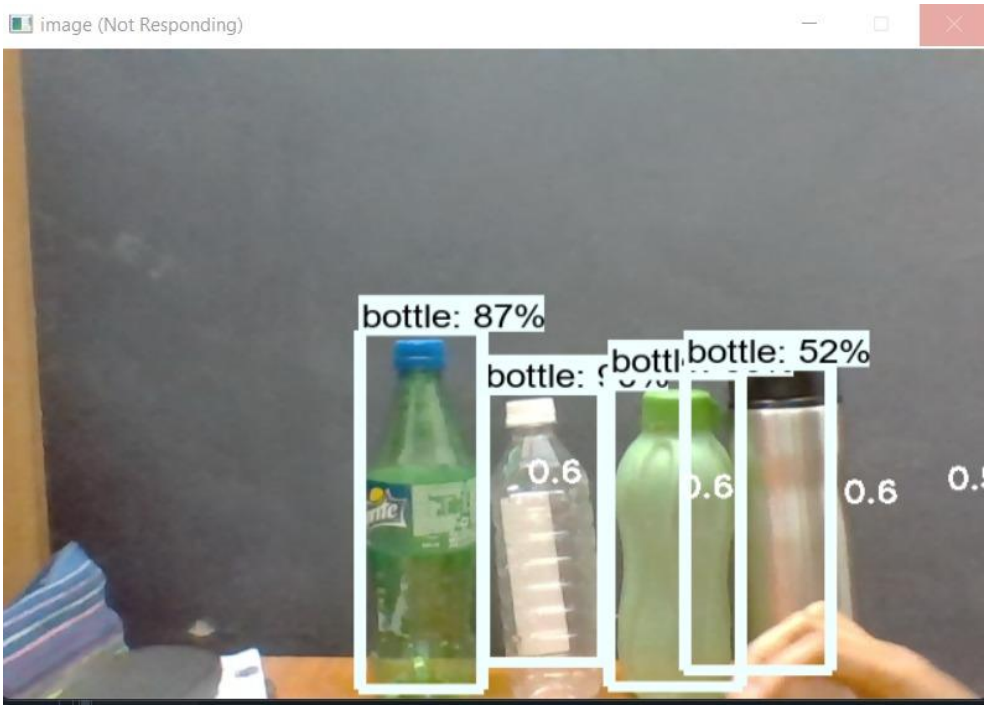

Figure 10.5 : System identifying different types of bottles with respective accuracy

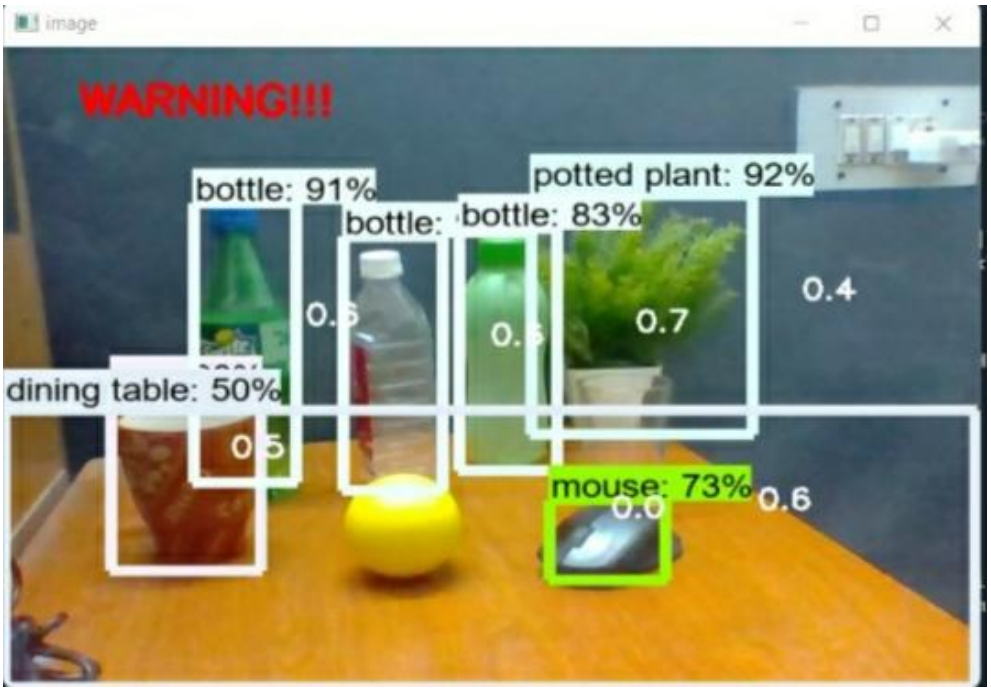

Figure 10.6 : System identifying multiple objects with respective accuracy

REFERENCES

[1] Joe Louis Paul I , Sasirekha S, Mohanavalli S, Jayashree C, Moohana Priya P, and Monika K, "Smart Eye for Visually Impaired-An aid to help the blind people". *Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019).*

[2] Divya .S, Shubham Raj, M. Praveen Shai, A. Jawahar Akash,Ms. V. Nisha **"**Smart Assistance Navigational System for Visually Impaired Individuals" *2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*.

[3] Kasthuri R,Nivetha B, Shabana S,Veluchamy M,Sivakumar S "Smart Device for Visually Impaired People". *2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)*,pp- 54-59.

[4] Arjun Pardasani, Prithviraj N Indi, Sashwata Banerjee, Aditya Kamal, Vaibhav "Smart Assistive Navigation Devices for Visually Impaired People ". 2*019 IEEE 4th International Conference on Computer and Communication Systems.*

[5] Shubham Melvin Felix, Sumer Kumar, and A. Veeramuthu, "A Smart Personal AI Assistant for Visually Impaired People". Proceedings of the 2*nd International Conference on Trends in Electronics and Informatics (ICOEI 2018).*

[6] Durgadevi, K.Thirupurasundari, C.Komathi and S.Mithun Balaji,"SMART MACHINE LEARNING SYSTEM FOR BLIND ASSISTANCE", *2020 2nd International Conference on Power, Energy, Control and Transmission Systems.*

[7] Mrs. Subhasini Shukla, Shreya Pimple, Shraddha Shetke, Namrata Gaikwad, Krutika Patil, "Smart Cap for Visually Impaired Person using Raspberry Pi", *International Research Journal of Engineering and Technology (ICSSIT 2020*), pp-2805-2809

[8] Agarwal A, D.Kumar, and A. Bhardwaj, "Ultrasonic Stick for Blind". *International Journal of Engineering and Computer Science, 2015*

[9] Andreas Wachaja, Pratik Agarwal Mathias Zink, Miguel Reyes Adame Knut Moller, Wolfram Burgard,"Navigating Blind People with a Smart Walker", *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Congress Center Hamburg, pp- 6014-6019