**OBJECT DETECTION AND IDENTIFICATION FOR BLIND AND DEAF PEOPLE IN A VIDEO SCENE**

#### A PROJECT REPORT

***Submitted by***

#### DEVARAKONDA RUPASREE [REGISTER NO:211418104047] LEBASINI RR [REGISTER NO:211418104139] DHARMARAJU HARIKA RAMAKRISHNA

**[REGISTER NO:211418104049]**

***in partial fulfillment for the award of the degree of***

### BACHELOR OF ENGINEERING

**IN**

#### COMPUTER SCIENCE AND ENGINEERING



**PANIMALAR ENGINEERING COLLEGE**

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

**MAY 2022**

### BONAFIDE CERTIFICATE

Certified that this project report **“OBJECT DETECTION AND IDENTIFICATION FOR BLIND AND DEAF PEOPLE IN A VIDEO SCENE”** is the bonafide work of “**DEVARAKONDA RUPASREE (211418104047), LEBASINI RR(211418104139), DHARMARAJU HARIKA RAMAKRISHNA(211418104049)”** who carried out the project work under my supervision.

#### SIGNATURE SIGNATURE

**Dr.S.MURUGAVALLI,M.E.,Ph.D Dr. JOSEPHINE LEELA, M.E,Ph.D., HEAD OF THE DEPARTMENT SUPERVISOR**

**PROFESSOR**

DEPARTMENT OF CSE, DEPARTMENT OF CSE,

PANIMALAR ENGINEERING COLLEGE, PANIMALAR ENGINEERING COLLEGE, NASARATHPETTAI, NASARATHPETTAI,

POONAMALLEE, POONAMALLEE,

CHENNAI-600 123. CHENNAI-600 123.

Certified that the above candidate(s) was/ were examined in the Anna University Project Viva-Voce Examination held on...........................

**INTERNAL EXAMINER EXTERNAL EXAMINER**

### DECLARATION BY THE STUDENT

We **DEVARAKONDA RUPASREE** (211418104047), **LEBASINI RR**

(211418104139),**DHARMARAJU HARIKA RAMAKRISHNA**(211418104049)

hereby declare that this project report titled “OBJECT DETECTION AND IDENTIFICATION FOR BLIND AND DEAF PEOPLE IN A VIDEO SCENE”,

under the guidance of **Dr. JOSEPHINE LEELA, M.E, Ph.D.,** is the original work done by us and we have not plagiarized or submitted to any other degree in any university by us.

#### DEVARAKONDA RUPASREE

**LEBASINI RR**

### DHARMARAJU HARIKA RAMAKRISHNA

**ACKNOWLEDGEMENT**

We would like to express our deep gratitude to our respected Secretary and Correspondent **Dr.P.CHINNADURAI, M.A., Ph.D.** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We express our sincere thanks to our Directors **Tmt.C.VIJAYARAJESWARI**, **Dr.C.SAKTHI KUMAR,M.E.,Ph.D** and **Dr.SARANYASREE SAKTHI KUMAR**

**B.E.,M.B.A.,Ph.D.,** for providing us with the necessary facilities to undertake this project.

We also express our gratitude to our Principal **Dr.K.Mani, M.E., Ph.D.** who facilitated us in completing the project.

We thank the Head of the CSE Department, **Dr. S.MURUGAVALLI , M.E.,Ph.D.,**

for the support extended throughout the project.

We would like to thank my **Project Guide Dr. JOSEPHINE LEELA,M.E,Ph.D.,** and all the faculty members of the Department of CSE for their advice and encouragement for thesuccessful completion of the project.

### ABSTRACT

Obstacle detection and warning can improve the camera as well as the safety of visually impaired and deaf people especially in unfamiliar environments. For this, firstly, obstacles are detected and localized and then the information of the obstacles will be sent to the visually impaired and deaf people by using different modalities such as voice, vibration. We present an assistive system for visually impaired people based camera. This system consists of two main components: environment information acquisition and analysis and information representation. The first component aims at capturing the environment by using a camera and analyzing it in order to detect the predefined obstacles for visually impaired and deaf people, while the second component when there is an obstacle to the right of the user; he or she feels strong frequent pulses on the right side through the haptic strap. For the user, this leads to the perception of an obstacle to the right. The user can then avoid the obstacle by stepping left. After stepping left, the pulses on the right side stop and the user perceives that there is no longer an obstacle on his or her rightside.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** | VI |
|  | **LIST OF TABLES** | IX |
|  | **LIST OF FIGURES** | X |
|  |  |  |
| **1.** | **INTRODUCTION** |  |
|  | 1.1 Overview | 2 |
|  | 1.2 Problem Definition | 3 |
| **2.** | **LITERATURE SURVEY** | 5 |
| **3.** | **SYSTEM ANALYSIS** |  |
|  | 3.1 Existing System | 11 |
|  | 3.2 Proposed system | 12 |
|  | 3.3 Feasibility Study | 14 |
|  | 3.4 Hardware Environment | 14 |
|  | 3.5 Software Environment |  |
| **4.** | **SYSTEM DESIGN** |  |
|  | 4.1. ER diagram | 16 |
|  | 4.2 Data Flow Diagram | 17 |
|  | 4.3 UML Diagrams | 18 |

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
| **5.** | **SYSTEM ARCHITECTURE** |  |
|  | 5.1 Module Design Specification | 26 |
|  | 5.2 Algorithms | 30 |
| **6.** | **SYSTEM IMPLEMENTATION** | 33 |
| **7.** | **PERFORMANCE ANALYSIS** |  |
|  | 7.1 Results and discussions | 39 |
|  | 7.2 Performance Analysis | 43 |
| **8.** | **CONCLUSION** |  |
|  | 8.1 Conclusion | 45 |
|  | 8.2 Future Enhancements | 46 |
|  | **APPENDICES** |  |
|  | A.1 Sample Screens | 47 |
|  | **REFERENCES** | 52 |

### LIST OF TABLES

|  |  |  |
| --- | --- | --- |
| **TABLE.NO** | **TITLE** | **PAGE.NO** |
| 7.3 | Training parameters and results of the three detection networks. SGD = stochastic gradient descent. | 39 |
| 7.4 | Detection results of the trained MobileNet- SSD algorithm. | 39 |
| 7.5 | Correct detection rate, training time and the detection time per image of each network. | 40 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TITLE** | **PAGE NO** |
| 4.4 | ER Diagram | 16 |
| 4.5 | Data Flow Diagram | 17 |
| 4.6 | Use Case Diagram | 18 |
| 4.7 | Class Diagram | 19 |
| 4.8 | Sequence Diagram | 20 |
| 4.9 | Activity Diagram | 21 |
| 4.10 | State Chart Diagram | 22 |
| 4.11 | Component Diagram | 23 |
| 4.12 | Deployment Diagram | 23 |
| 5.3 | System Architecture Diagram | 25 |
| 7.6 | Chair | 41 |
| 7.7 | Tv | 41 |
| A.2 | A person is detected with an accuracy of 99.51% | 47 |
| A.3 | TV is detected in frame with accuracy of 92.61% | 48 |
| A.4 | Chair is detected in frame with accuracy of 98.31% | 49 |
| A.5 | Audio Output | 50 |
| A.6 | Vibration Output | 51 |

# CHAPTER 1

## INTRODUCTION

### 1.INTRODUCTION

#### OVERVIEW

This project investigates methods and procedures to construct an efficient system to assist blinds in their everyday life. In particular, various technologies that can be utilized to build a wearable system are examined. The machine vision and the communication component of the blind navigation and guidance is designed not only to search the surroundings environment but also to determine a safe. This work highlights the importance and also provides the instructions to blinds for efficient navigation and safe guidance by incorporating object/pedestrian detection in real- time. Real time object detection is widely used Open CV concept that mainly concentrates on identifying, detecting objects belonging to different classes in the input image. Detecting the objects along with its unique features and background details is done using different techniques and methods. One among them is creating bounding boxes so that pixels of all the aspects of the objects are found and detection is done accordingly with certain computations. The proposed system’s goal is to detect objects that are shown in front of it and convert the detected object to speech.Object detectors can be made as a portable device that can be fitted or carried anywhere to detect and recognize objects. The system requires a portable device which integrates the code, camera and sensors required for object detection. The main working model runs in python code in the laptop and the output is obtained through the web camera on the laptop or any personal device. Before running the code, the necessary packages must be installed and the pre-built model should be properly imported. The datasets are present in the model and its completely trained.

#### PROBLEM DEFINITION

Blindness is a state of lacking the visual perception due to neurological or physiological factors. The partial blindness represents the lack of integration in the growth of the optic visual or nerve center of the eye, and total blindness is the full absence of the visual light perception. In this work, cheap, a simple friendly user, smart blind guidance system is designed and implemented to improve the mobility of both blind and visuallyimpaired people in a specific area. The proposed work includes a wearable equipment consists of light weight blind stick and sensor based obstacle detection circuit is developed to help the blind person to navigate alone safely and to avoid any obstacles that may be encountered, whether fixed or mobile, to prevent any possible accident. The main objective of this project is to develop an application for blind people to detect the objects in various directions, detecting pits and manholes on the ground to make free to walk Detecting objects using image processing can be used in multiple industrial as well as social applications. This project is proposing to use object detection for blind people and give them audio/ vocal information about it. We are detecting an object using the mobile camera and giving voice instructions about the direction of an object.

# CHAPTER 2

## LITERATURE SURVEY

### 2.LITERATURE SURVEY

**[1]Title** -A Smart Personal AI Assistant for Visually Impaired People

**Year** – 2018

**Author** -Shubham Melvin Felix, Sumer Kumar, and A. Veeramuthu

**Concept-**In today’s advanced hi-tech world, the need of independent living is recognized in case of visually impaired people who are facing main problem of social restrictiveness. They suffer in strange surroundings without any manual aid. Visual information is the basis for most tasks, so visually impaired people are at 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 using Artificial Intelligence, Machine Learning, Image and Text Recognition.

**Technique** -Impaired, AI, Machine Learning, Voice Assistant, Chat bot, Image recognition.

**Future scope**- Smart portable assistant system for blind

**Drawback** -Involves high level data processing need to use costly hardware

**[2]Title** - A vision and speech enabled, customizable, virtual assistant for smart

environments.

**Year** – 2018

**Author** - Giancarlo Iannizzotto,Lucia Lo BelloAndrea,Nucita,Giorgio MarioGrasso

**Concept**- Recent developments in smart assistants and smart home automation are lately attracting the interest and curiosity of consumers and researchers.

Speech enabled virtual assistants (often named smart speakers) offer a wide variety of network oriented services and, in some cases, can connect to smart environments, thus enhancing them with new and effective user interfaces.

However, such devices also reveal new needs and some weaknesses. In particular, they represent faceless and blind assistants, unable to show a face, andtherefore an emotion, and unable to ‘see’ the user.

**Technique** - Smart home, virtual assistant, computer vision, deep learning.

**Future scope**- Interactive smart home.

**Drawback** - Request response based system.

**[3]Title** - Digital Assistant for The Blind

**Year** – 2017

**Author** -Prince Bose, Apurva Malpthak, Utkarsh Bansal, Ashish Harsola

**Concept** -The blind and the visually impaired have little to no internet presence because of the absence of cheap solutions to get them online which can be both, hardware and software. Existing technology used for enabling the blind or visually impaired to use the internet or any digital form of information is dependent on Braille displays and keyboards which are expensive and scarce.

Another shortcoming of existing technology is that out of all the visually impaired population, less than 2% know how to interpret Braille. Hence a voice controlled system for the blind and the visually impaired was designed, which transceivers information in the form of audio.

**Technique** - Voice assistant, Speech Recognition, Low cost, Internet, Speech Visually Challenged, Raspberry Pi

**Future scope**- Navigation for the visually challenged using voice instructionsand haptic feedbacks can be added to improve the existing system.

**Drawback** - Uses audio feedback

**[4]Title**- Android Assistant EyeMate for Blind and Blind Tracker

**Year** – 2015

**Author** - Md. Siddiqur Rahman Tanveer, M.M.A. Hashem and Md. Kowsar Hossain Concept - At present many blind assistive systems have been implemented but there is no such kind of good system to navigate a blind person and also to track the movement of a blind person and rescue him/her if he/she is lost. In this paper, we have presented a blind assistive and tracking embedded system. In this system the blind person is navigated through a spectacle interfaced with an android application. The blind person is guided through Bengali/English voice commands generated by the application according to the obstacle position. Using voice command a blind person can establish voice call to a predefined number without touching the phone just by pressing the headset button.

**Technique** – Arduino, Android, RFID, Ultrasonic, GSM, GPS, Server,

JSON, Microcontroller, Pulse, Analog, Modulation.

**Future scope**- Interactive smart home.

**Drawback**- Request response based system.

**[5]Title-** Navigation Assistive System for the Blind using a Portable Depth Sensor

**Year** – 2015

**Author** - Kumar Yelamarthi

**Concept** - The lightweight and low-cost 3-dimensional depth sensors have gained much attention in the computer vision and robotics community. While its performance has been proven successful in the robotics community, these sensors have not been utilized successfully for many assistive devices. Leveraging on this gap, this paper presents the design, implementation, and preliminary evaluation of a haptic feedback system for the blind using 3-D depth sensors. The proposed portable system interprets the visual scene using the depth sensor, converts it into distance map, processes, and evaluates this information using a tablet computer.

**Technique** - depth sensor, haptic feedback, blind, navigation assistance.

**Future scope**- Assistant system for blind people in rough terrain.

**Drawback** - Sensors are not compatible for a mobile system.

# CHAPTER 3

## SYSTEM ANALYSIS

### SYSTEM ANALYSIS

#### EXISTING SYSTEM

The common way for navigating of visionless person is using a walking stick cane or walking cane. The walking cane is a simple and mechanical device dedicated to detect static obstacles on the ground, uneven surfaces, and holes via simple tactile-force feedback. This device is light, portable but range limited and it is not usable for the protection from obstacles near to head area. Another option that provides the best travel aid for the blind is the guide dogs. The disabled owner and his dog, the training and therelationship to the animal are the keys to success for this method. The dog is able to detect and analyse complex situations: cross walks, stairs, potential danger, know pathsand more.

**DISADVANTAGES**

* + - Walking sticks can only help to a level
      * Previous existing systems uses range finder sensor that can only detect obstacle in general not the type of obstacle
      * No audio feedback available

#### PROPOSED SYSTEM

Our proposed visual substitution system is based on the identification of objects around the blind and deaf person. We propose a system that recognize and locate 2D in the video. This system should find the invariant characteristic of objects, provide the recognition and reduce the complexity of detection. We propose a method based on object extraction and matching in video. A comparison between query frame and database objects is made to detect object in each frame. For each object detected an audio file containing the information about it is activate.

#### ADVANTAGES

* + - Convolutional Neural Network is very advanced deep learning algorithm
    - Number of data set is high hence accuracy is also high
    - Uses simple USB camera for video capture
    - No 3D camera is used
    - cost effective

#### FEASIBILITY STUDY

* + - ECONOMICAL
    - TECHNICAL
    - SOCIAL

#### ECONOMICAL FEASIBILITY:

The Raspberry Pi is perfect for adaptive technology and it is able to display images or play videos. Object detection can be made more faster by using a high end Single board computer with GPU. This microcomputer is useful for small business that run on a lower budget to use their product or to invent new technology that embeds the product.

#### TECHNICAL FEASIBILITY:

**PYTHON:**

Python is used here for predictive modeling because Python-based frameworks give us results faster and also help in the planning of the next steps based on the results. The product does not require user to have extensive programming experience since it is aimed for the younger generation to learn about programming.

#### SOCIAL FEASIBILITY:

This website will be helpful for the whole society. The Raspberry Pi is perfect for adaptive technology and it is able to display images or play videos. Object detection can be made more faster by using a high end Single board computer with GPU.

#### HARDWARE ENVIRONMENT

* Raspberry pi
* Camera
* Haptic strip
* RAM - 8 GB (min)
* Hard Disk - 500 GB
* Key Board - Standard Windows Keyboard
* Mouse - Two or Three Button Mouse
* Monitor - LCD

#### SOFTWARE ENVIRONMENT

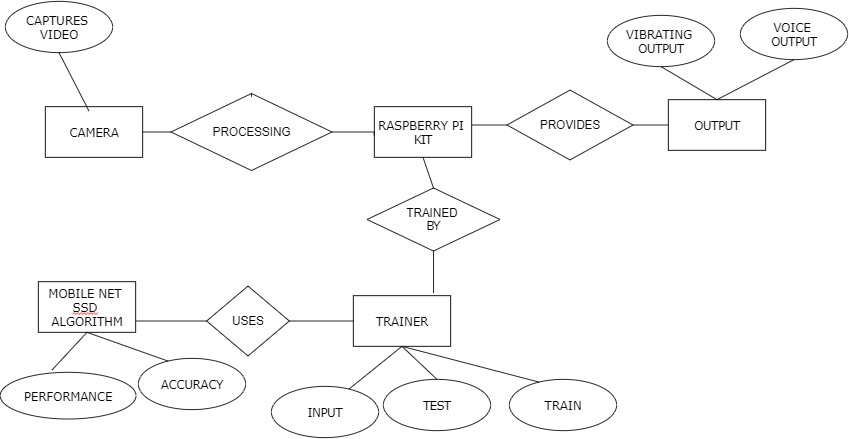
* Operating System: Windows/7/10
* Raspberry pi OS: Raspbian stretch
* Programming platform: python 3 IDLE
* Programing language: python 3
* Library: OpenCV

# CHAPTER 4

## SYSTEM DESIGN

### SYSTEM DESIGN

#### 4.4 ER DIAGRAM



**FIG NO.4.4 ER Diagram**

An entity relationship diagram (ERD), also known as an entity relationship model, is a graphical representation of an information system that depicts the relationships among people, objects, places, concepts or events within that system. An ERD is a modeling technique that can help define business processes and be used as the foundation for relational database. Entity relationship diagrams provide a visual starting point for database design that can also be used to help determine information system requirements throughout an organization.

#### 4.5 DATA FLOW DIAGRAM



Capture Video

Feature Extraction

Choose model

Train model



Test model



Matching Features

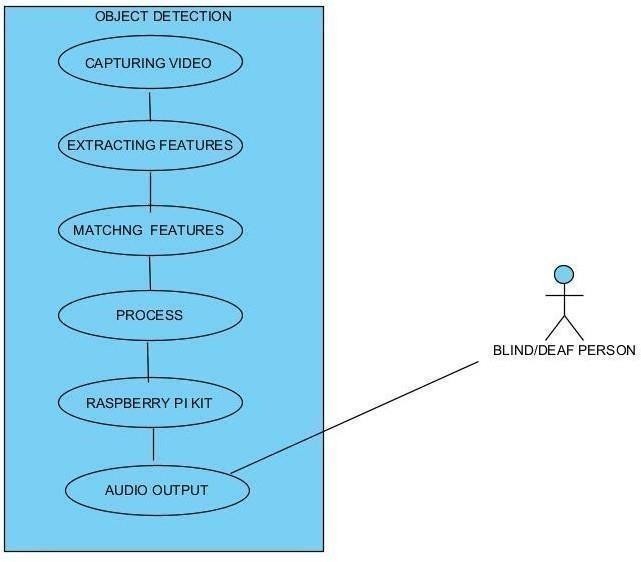
Vibration and Audio output

**FIG NO.4.5 DATA FLOW DIAGRAM**

It needs data gathering have lot of past data’s. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data can’t be used directly. It’s used to preprocess then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

#### UML DIAGRAMS

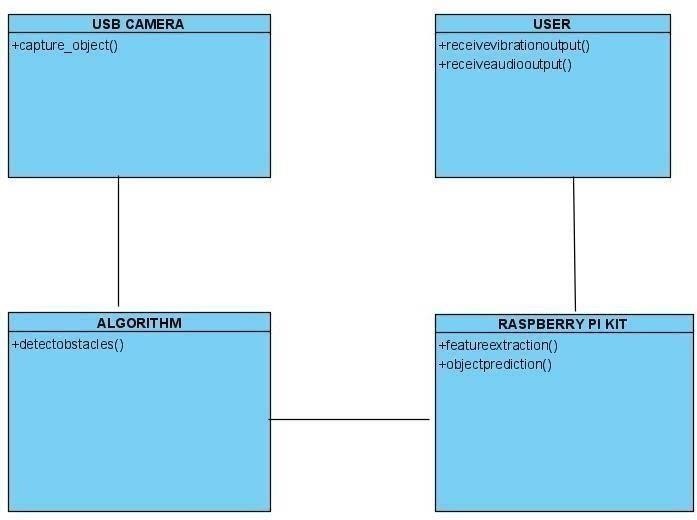
**4.6USE CASE DIAGRAM:**



**FIG NO.4.6 USE CASE DIAGRAM**

Use case diagrams are considered for high level requirement analysis of a system. So when the requirements of a system are analyzed the functionalities are captured in use cases. So, it can say that uses cases are nothing but the system functionalities written in an organized manner.

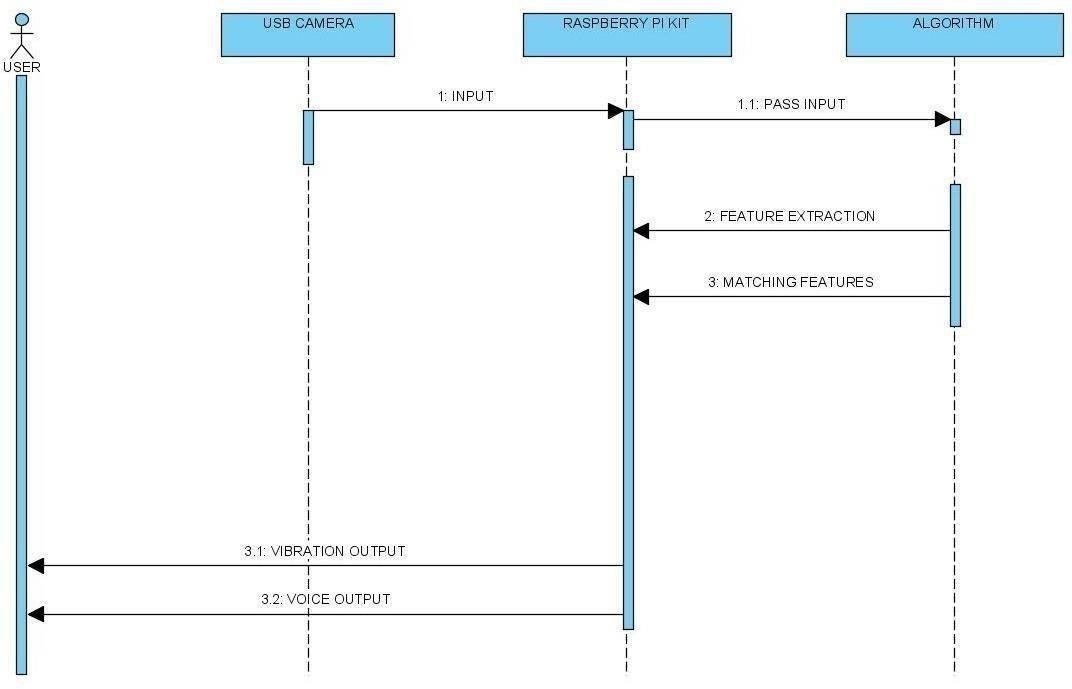
#### 4.7CLASS DIAGRAM



**FIG NO.4.7 CLASS DIAGRAM**

Class diagram is basically a graphical representation of the static view of the system and represents different aspects of the application. So a collection of class diagrams represent the whole system. The name of the class diagram should be meaningful to describe the aspect of the system. Each element and their relationships should be identified in advance Responsibility (attributes and methods) of each class should be clearly identified for each class minimum number of properties should be specified and because, unnecessary properties will make the diagram complicated.

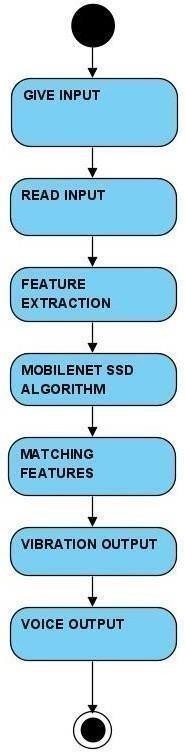
#### SEQUENCE DIAGRAM



**FIG NO.4.8 SEQUENCE DIAGRAM**

Sequence diagrams model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and are commonly used for both analysis and design purposes. Sequence diagrams are the most popular UML artifact for dynamic modeling, which focuses on identifying the behavior within your system. Other dynamic modeling techniques include activity diagramming, communication diagramming, timing diagramming and interaction overview diagramming.

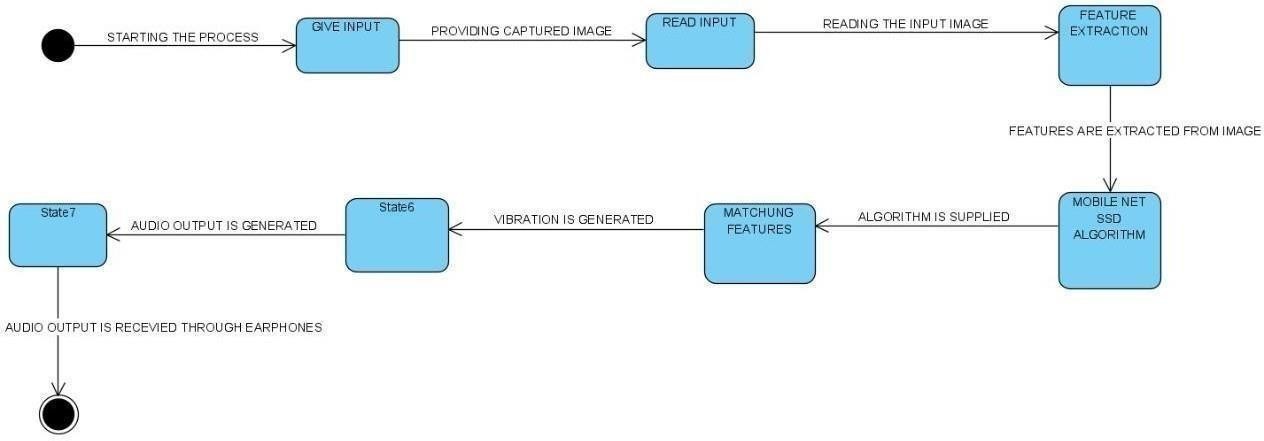
#### ACTIVITY DIAGRAM



**FIG NO.4.9 ACTIVITY DIAGRAM**

Activity is a particular operation of the system. Activity diagrams are not only used for visualizing dynamic nature of a system but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in activity diagram is the message part. It does not show any message flow from one activity to another.

#### STATE CHART DIAGRAM:

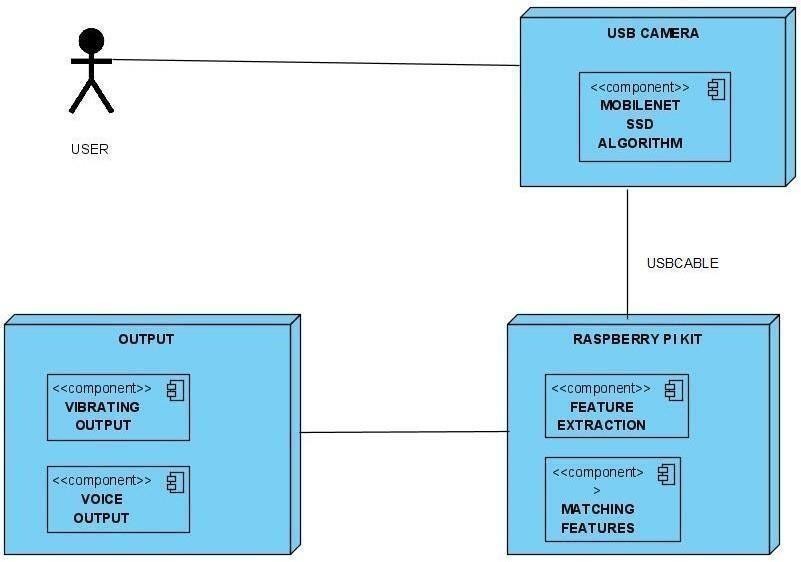


**FIG NO.4.10 STATE CHART DIAGRAM**

State chart diagram describes the flow of control from one state to another state. States are defined as a condition in which an object exists and it changes when some event is triggered. The most important purpose of State chart diagram is to model lifetime of an object from creation to termination.

State chart diagrams are also used for forward and reverse engineering of a system. However, the main purpose is to model the reactive system.

### DEPLOYMENT DIAGRAM:



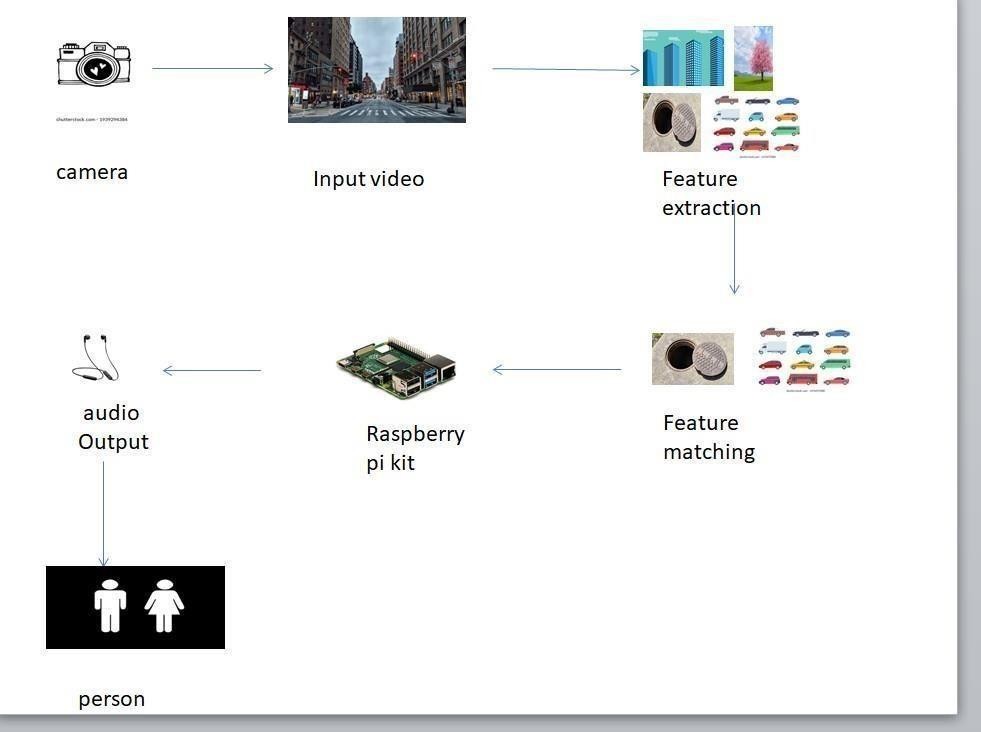
**FIG NO.4.11 DEPLOYMENT DIAGRAM**

Deployment diagrams are used to visualize the topology of the physical components of a system, where the software components are deployed. Deployment diagrams are used to describe the static deployment view of a system. Deployment diagrams consist of nodes and their relationships. The term Deployment itself describes the purpose of the diagram. Deployment diagrams are used for describing the hardware components, where software components are deployed. Component diagrams and deployment diagrams are closely related. Component diagrams are used to describe the components and deployment diagrams shows how they are deployed in hardware.

# CHAPTER 5

## SYSTEM ARCHITECTURE

### 5.SYSTEM ARCHITECTURE



**FIG NO.5.3 SYSTEM ARCHITECTURE DIAGRAM**

A system architecture or systems architecture is the conceptual model that defines the structure, behaviour, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviours of the system. This system provides majorly four modules, which are mentioned below. The system provides user friendly GUI and data flows to all modules in dynamic .

#### MODULE DESIGN SPECIFICATIONS

**LIST OF MODULES**

* + - Feature extraction
    - Object detection
    - Training
    - Testing

#### MODULE-01

#### FEATURE EXTRACTION

The main purpose of using features instead of raw pixel values as the input to a learning algorithm is to reduce/increase the in-class/out of class variability compared to the raw input data, and thus making classification easier. Open CV systems generally exploit a single camera to capture image data. Recognition is then performed based on various features extracted from that data. In addition, features extraction is the process by which certain features of interest within an image are detected and represented for further processing. It is a critical step in most computer vision and image processing solutions because it marks the transition from pictorial to non-pictorial (alphanumerical, usually quantitative) data representation . Types of features that can be extracted from image depend on the type of image, the level of granularity desired, and the context of the application. Once the features have been extracted, theirs representation depends on the technique used. The features extraction process should be precise, so that the same features are extracted on two images showing the same object .

#### MODULE-02

**OBECT DETECTION**

Object detection is a computer technology related to computer vision and image

processing that deals with detecting instances of semantic objects of a certain class

(such as humans, buildings, or cars) in digital images and videos.Well-researched

domains of object detection include face detection and pedestrian detection. Object

detection has applications in many areas of computer vision, including image

retrieval and video surveillance

#### MODULE-04

**TRAINING:**

The objects of video captured is collected. The collected image is preprocessed and features like distance, object type are extracted. Based on the features extracted the model is trained.

#### MODULE-05

**TESTING:**

The object from the camera is collected. The collected image is preprocessed and required features like distance, object type is extracted from the image. Based on the trained data, the system tests the feature extracted and predicts the object distance**.**

#### SPECIFICATIONS:

**PROCESSOR**

Broadcom BCM2387 chipset. 1.2GHz Quad-Core ARM Cortex-A53

802.11 b/g/n Wireless LAN and Bluetooth 4.1 (Bluetooth Classic and LE)

#### GPU

Dual Core VideoCore IV Multimedia Co-Processor. Provides Open GL ES 2.0, hardware-accelerated OpenVG, and 1080p30 H.264 high-profile decode. Capable of 1Gpixel/s, 1.5Gtexel/s or 24GFLOPs with texture filtering

and DMA infrastructure

#### MEMORY

1GB LPDDR2

#### OPERATING SYSTEM

Boots from Micro SD card, running a version of the Linux operating system or Windows 10 IoT

#### DIMENSIONS

85 x 56 x 17mm

#### POWER MICRO

USB socket 5V1, 2.5A

#### ETHERNET

10/100 Base T Ethernet socket

#### VIDEO OUTPUT

HDMI (rev 1.3 & 1.4

Composite RCA (PAL and NTSC)

#### AUDIO OUTPUT

Audio Output 3.5mm jack, HDMI USB 4 x USB 2.0 Connector

#### ALGORITHM

**MOBILE NET SSD ALGORITHM:**

Object detection is one of the most important fields of exploration in computer vision today. It is an augmentation of image classification the objective is to identify one or more classes of objects in a picture and with the help of bounding boxes locate their presence. Consequently, object detection carries an important role in many real-world applications like image recovery and video surveillance.

The main purpose of our analysis is to elaborate the accuracy of an object detection technique SSD and the pre- trained deep learning model MobileNet and additionally feature a portion of the notable elements that make this method stand out.

The trial results show that the Average Precision (AP) of the algorithm to recognize various classes as vehicle, person and chair is 99.76%, 97.76% and 71.07%, separately. This improves the accuracy of behavior detection at a handling speed which is needed for the real-time location and the necessities of day by day observing indoor and outside. The mix of MobileNet into the SSD framework forms one of the center parts of our work.

Our proposed model depends on the MobileNet- SSD architecture. One reason why we chose this architecture is on it gives good object detection accuracy while being quicker than different architectures. Especially, this is valid when attempting to detect object in real time in low computing devices as in our system. MobileNet-SSD permits to lessen the detection time by addressing the model utilizing 8-bit integers rather than 32-bit floats.

OpenCV is a library of programming functions basically focused on real time computer vision. OpenCV is an open-source library which is useful for computer vision applications like CCTV film analysis, video analysis and image analysis. It is an incredible tool for image processing and performing computer vision tasks.

Input will be given through Realtime video by camera or webcam, based on streamlined MobileNet Architecture which uses depth-wise separable convolutions to build light weight deep neural Networks. The input video divided into frames and pass it to MobileNet layers. The job of the MobileNet layers is to change over the pixels from the input image into highlights that describe the contents of the image. Then it passes to MobileNet-SSD model to determine the bounding boxes and corresponding class (label) of objects. After that the only last step is to show or display the Output.

The SSD architecture is a single convolution network that learns to predict bounding box locations and classify these locations in one pass. Hence, SSD can be trained end- to-end. The SSD network consists of base architecture (MobileNet in this case) followed by several convolution layers

By using SSD, we only need to take one single shot to detect multiple objects within the image, while regional proposal network (RPN) based approaches such as R-CNN series that need two shots, one for generating region proposals, one for detecting the object of each proposal. Thus, SSD is much faster compared with two-shot RPN-based approaches.

# CHAPTER 6

## SYSTEM IMPLEMENTATION

### 6.SYSTEM IMPLEMENTATION

# USAGE

# python real\_time\_object\_detection.py --prototxt MobileNetSSD\_deploy.prototxt.txt

--model MobileNetSSD\_deploy.caffemodel

# import the necessary packages

from imutils.video import VideoStream from imutils.video import FPS import numpy as np

import argparse import imutils import time import cv2

import subprocess ##import lcddriver import urllib.request import RPi.GPIO as GPIO

GPIO.setmode(GPIO.BCM) GPIO.setwarnings(False) GPIO.setup(21,GPIO.OUT) GPIO.setup(20,GPIO.OUT)

GPIO.output(21,False) GPIO.output(20,False)

# construct the argumen parse and parse the arguments ap = argparse.ArgumentParser()

ap.add\_argument("-p", "--prototxt", required=True, help="path to Caffe 'deploy' prototxt file") ap.add\_argument("-m", "--model", required=True,

help="path to Caffe pre-trained model") ap.add\_argument("-c", "--confidence", type=float, default=0.7,

help="minimum probability to filter weak detections") args = vars(ap.parse\_args())

##display = lcddriver.lcd()

# initialize the list of class labels MobileNet SSD was trained to

# detect, then generate a set of bounding box colors for each class

CLASSES = ["background", "aeroplane", "bicycle", "bird",

"bus", "car", "cat", "chair", "cow", "diningtable",

"dog", "horse", "motorbike", "person", "pottedplant", "sheep", "sofa", "train", "tvmonitor"]

COLORS = np.random.uniform(0, 255,size=(len(CLASSES),3))

# load our serialized model from disk print("[INFO] loading model...")

net = cv2.dnn.readNetFromCaffe(args["prototxt"], args["model"])

# initialize the video stream, allow the cammera sensor to warmup, # and initialize the FPS counter

print("[INFO] starting video stream...") ##vs = VideoStream(src=0).start()

# vs = VideoStream(usePiCamera=True).start() ##time.sleep(2.0)

fps = FPS().start() url="http://172.20.10.2:8080/shot.jpg" # loop over the frames from the video stream

while True:

# grab the frame from the threaded video stream and resize it # to have a maximum width of 400 pixels

## frame = vs.read() imgPath=urllib.request.urlopen(url) imgNp=np.array(bytearray(imgPath.read()),dtype=np.uint8) img=cv2.imdecode(imgNp,-1)

frame = imutils.resize(img, width=400) ## display.lcd\_clear()

## display.lcd\_display\_string("Waiting...", 1)

# grab the frame dimensions and convert it to a blob (h, w) = frame.shape[:2]

blob = cv2.dnn.blobFromImage(cv2.resize(frame, (300, 300)),

0.007843, (300, 300), 127.5)

# pass the blob through the network and obtain the detections and # predictions

net.setInput(blob) detections = net.forward()

# loop over the detections

for i in np.arange(0, detections.shape[2]):

# extract the confidence (i.e., probability) associated with # the prediction

confidence = detections[0, 0, i, 2]

# filter out weak detections by ensuring the `confidence` is # greater than the minimum confidence

if confidence > args["confidence"]:

# extract the index of the class label from the

# `detections`, then compute the (x, y)-coordinates of # the bounding box for the object GPIO.output(21,True) GPIO.output(20,True)

idx = int(detections[0, 0, i, 1])

box = detections[0, 0, i, 3:7] \* np.array([w, h, w, h]) (startX, startY, endX, endY) = box.astype("int")

# draw the prediction on the frame

label = "{}: {:.2f}%".format(CLASSES[idx], confidence \* 100) print(label)

data=str(CLASSES[idx])+" in fron of you" subprocess.call('echo '+data+'|festival --tts', shell=True) print(startX,endX) cv2.rectangle(frame, (startX, startY), (endX, endY), COLORS[idx], 2)

y = startY - 15 if startY - 15 > 15 else startY + 15 cv2.putText(frame, label, (startX, y), cv2.FONT\_HERSHEY\_SIMPLEX, 0.5, COLORS[idx], 2)

time.sleep(1) GPIO.output(21,False) GPIO.output(20,False)

# show the output frame cv2.imshow("Frame", frame) key = cv2.waitKey(1) & 0xFF

# if the `q` key was pressed, break from the loop if key == ord("q"):

break

# CHAPTER 7

## PERFORMANCE ANALYSIS

### 7.PERFORMANCE ANALYSIS

#### 7.1RESULTS AND DISCUSSIONS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DEFECT TYPE** | **SAMPLE NUMBER** | **SUCCESSFUL DETECTION NUMBER** | **LEAKAGE NUMBER** | **ERROR DETECTION NUMBER** | **POSITIVE RATE(%)** |
| BREACH | 30 | 30 | 0 | 0 | 100.00 |
| DENT | 30 | 27 | 2 | 1 | 90.00 |
| BURR | 30 | 28 | 1 | 1 | 93.33 |
| ABRASION | 30 | 29 | 1 | 0 | 96.67 |
| TOTAL | 120 | 114 | 4 | 2 | 95.00 |

**TABLE NO:7.3 Detection results of the trained MobileNet-SSD algorithm.**

|  |  |  |  |
| --- | --- | --- | --- |
| **PARAMETER** | **VGG-16** | **MobileNet** | **MobileNet-SSD** |
| BASIC LEARNING RATE | 0.003 | 0.003 | 0.003 |
| NUMBER OF ITERATIONS OF TRAINING | 10,000 | 10,000 | 10,000 |
| VERYFYING THE NUMBER OF ITERATIONS | 50 | 50 | 50 |
| BATCH QUANTITY | 128 | 128 | 128 |

|  |  |  |  |
| --- | --- | --- | --- |
| OPTIMIZATION ALGORITHM | SGD | SGD | SGD |
| NETWORK PARAMETER(MILLION) | ~12.30 | ~1.76 | ~1.92 |
| AMOUNT OF CALCULATION(MILLION) | ~1325.00 | ~127.54 | ~157.32 |
| MEAN OF LOSS FUNCTION | 0.14 | 0.19 | 0.09 |
| MEAN VALUE | 93.91 | 92.33 | 96.73 |

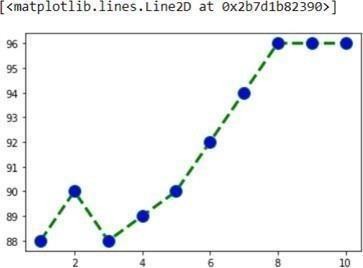
**TABLE NO:7.4 Training parameters and results of the three detection networks.**

**SGD = stochastic gradient descent.**

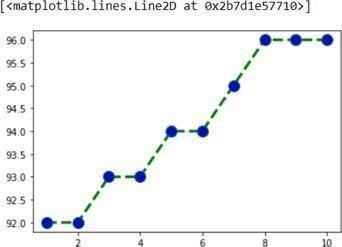
We have plot a graph of Performance Metrics w.r.t Accuracy . It can be seen as follows:

X-Axis: Time Frame(in Sec.)

Y-AXIS:Accuracy Metrics(in percentage)



#### FIG NO:7.5 CHAIR



**FIG NO:7.6 TV**

The proposed system successfully detects 90 objects, labels them and also shows its accuracy. The model also calculates the distance from the object to the camera and gives a voice feedback as when the person with the camera is approaching the object.

The dataset was tested on two different models, SSD Mobilenet V1 and SSD Inception V2. However the SSD Mobilenet V1 model showed less latency and was faster in detecting objects.

Adjusting Minimum Probability: By default, objects detected with a probability percentage of less than 50 will not be shown or reported. You can increase this value for high certainty cases or reduce the value for cases where all possible objects are needed to be detected.

Custom Objects Detection: Using a provided Custom Object class, you can tell the detection class to report detections on one or a few number of unique objects.

Detection Speeds: You can reduce the time it takes to detect an image by setting the speed of detection speed to “fast”, “faster” and “fastest”.

Input Types: You can specify and parse in file path to an image, Numpy array or file streamof an image as the input image.

Output Types: You can specify that the detect objects from image function should return the image in the form of a file or Numpy array.

#### PERFORMANCE ANALYSIS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **NUMBER** | **MODEL** | **POSITIVE RATE** | **TRAINING TIME(DAY)** | **DETECTION TIME(S)** |
| 1 | SqueezeNet | 85.83% | 2 | 0.31 |
| 2 | Faceness-Net | 90.83% | 3 | 0.59 |
| 3 | MobileNet | 90.83% | <1 | 0.54 |
| 4 | MTCNN | 91.67% | 3 | 0.64 |
| 5 | PVANet | 94.17% | 2 | 0.25 |
| 6 | MobileNet-  SSD | 95.00% | <1 | 0.12 |

**TABLE NO:7.5 Correct detection rate, training time and the detection time per image of each network.**

# CHAPTER 8

## CONCLUSION

### 8.CONCLUSION

#### 8.1CONCLUSION

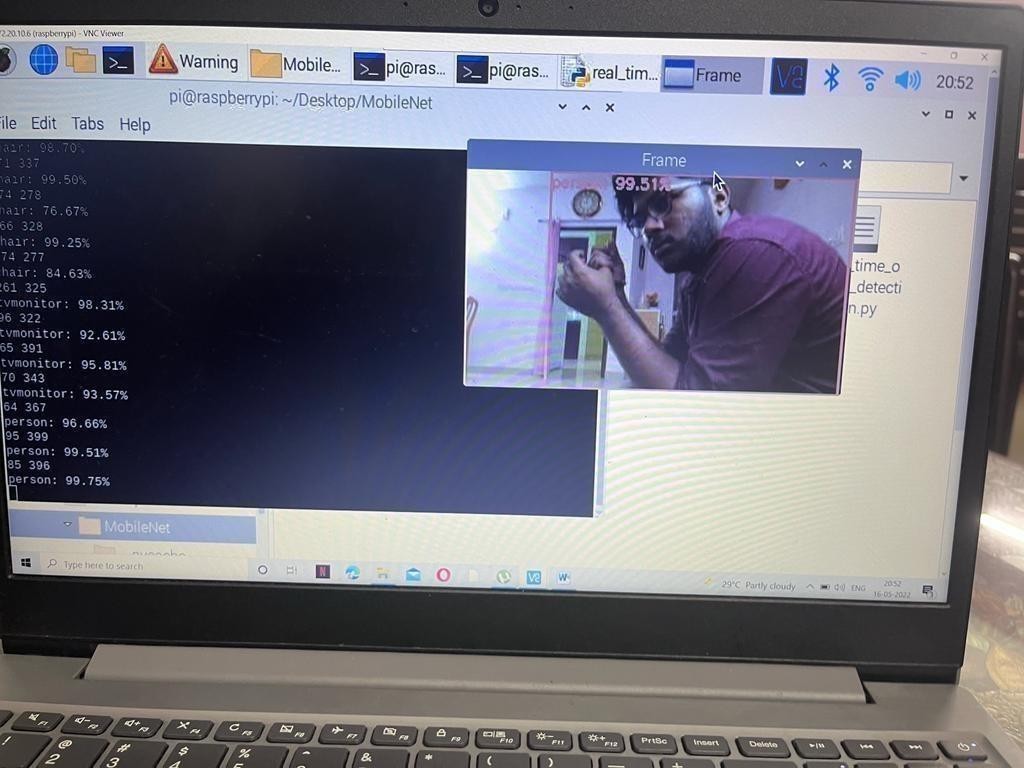
This project provides a means of active interaction and navigation between the system and the blind person. Its features like interaction or commands over voice andreliable obstacle detections make its future scope huge. With the advancement in technology and our dependency on it, this project would be proved to be high in demand in the recent future. The test results prove that this system is more reliable than any other existing system, however, some more improvements can make it even more faithful and accurate. Also, the cost of this project is optimum and proves to be the best aid for blind people, and it will help them to be independent of their instincts and other people.

#### 8.2FUTURE ENHANCEMENT

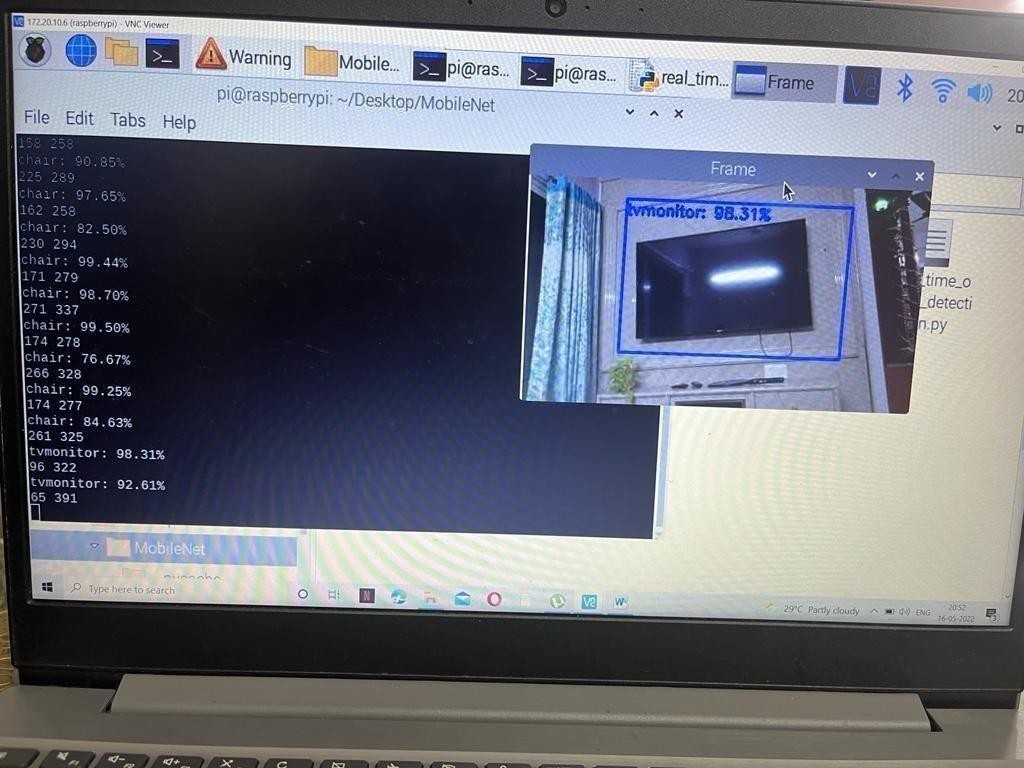
We present a visual substitution system for blind people based on object recognition in video scene. This system uses SIFTS key points extraction and features matching for object identification. We devote the experimental part to test the application in order to detect some objects in some video scene with different conditions. In this stage of works, we address the recognition of each object in the scene as an individual task, we do not consider the relationships between many objects. Thus, in future works, we will consider this relationship for scene understanding or detecting everything that belongs to a given place or location. Finally, in order to help bind people and to provide from the new technologies, a mobile application can be the best solution.

### APPENDICES

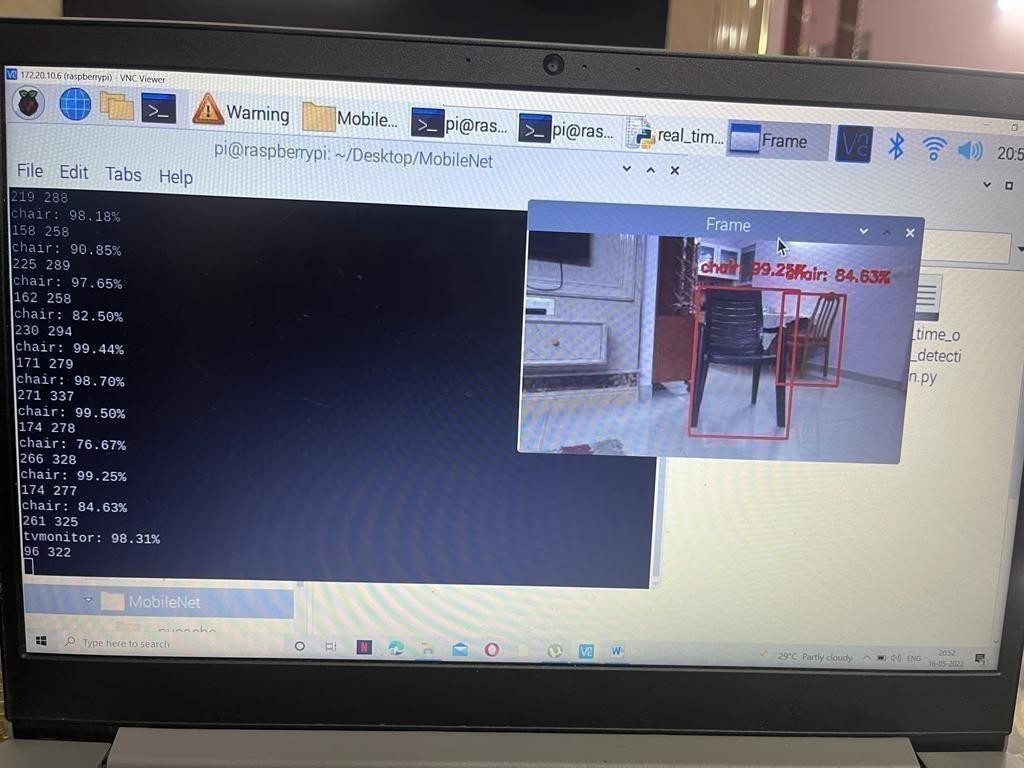
#### A.1 SAMPLE SCREENS



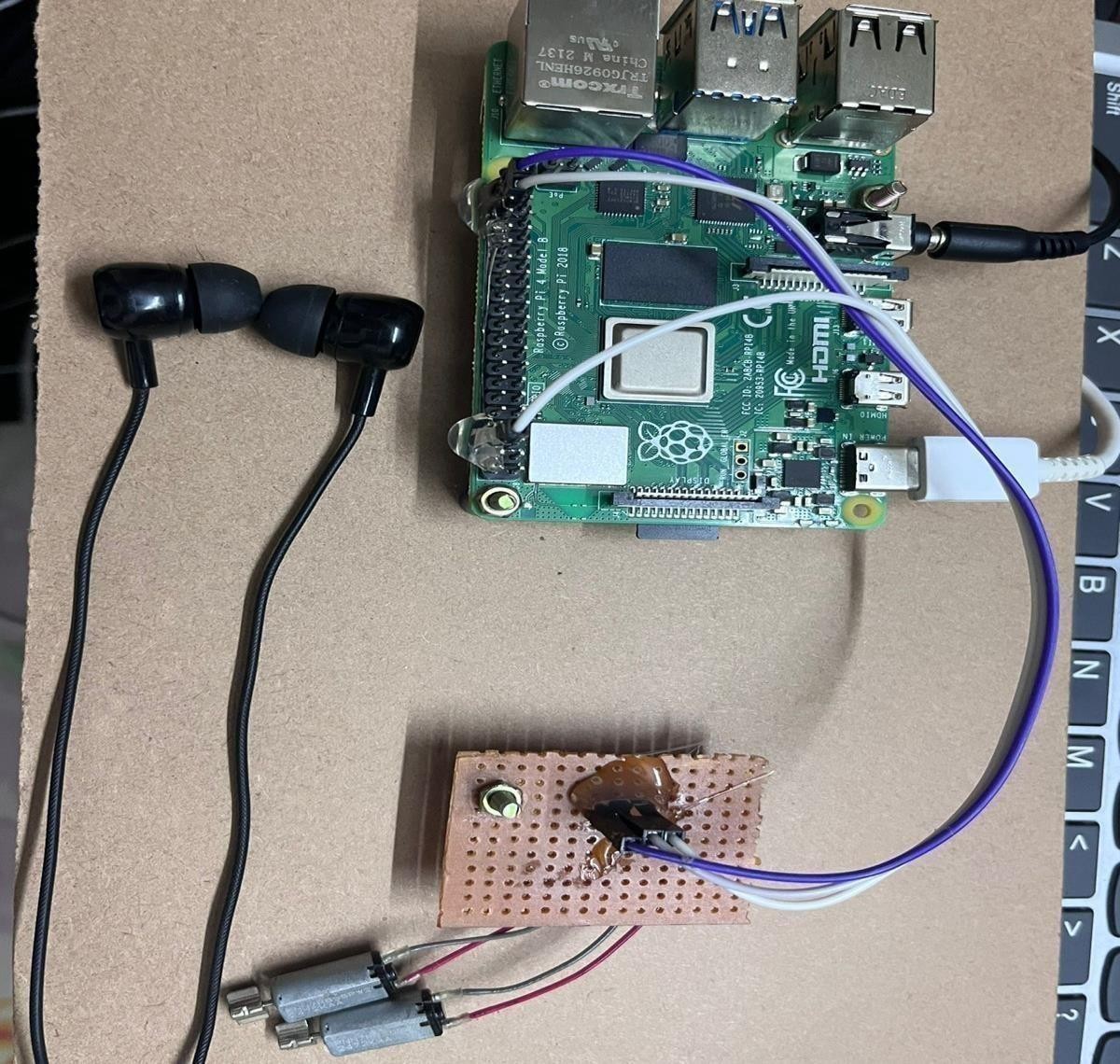
**FIG NO.A.2 A person is detected with an accuracy of 99.51%**



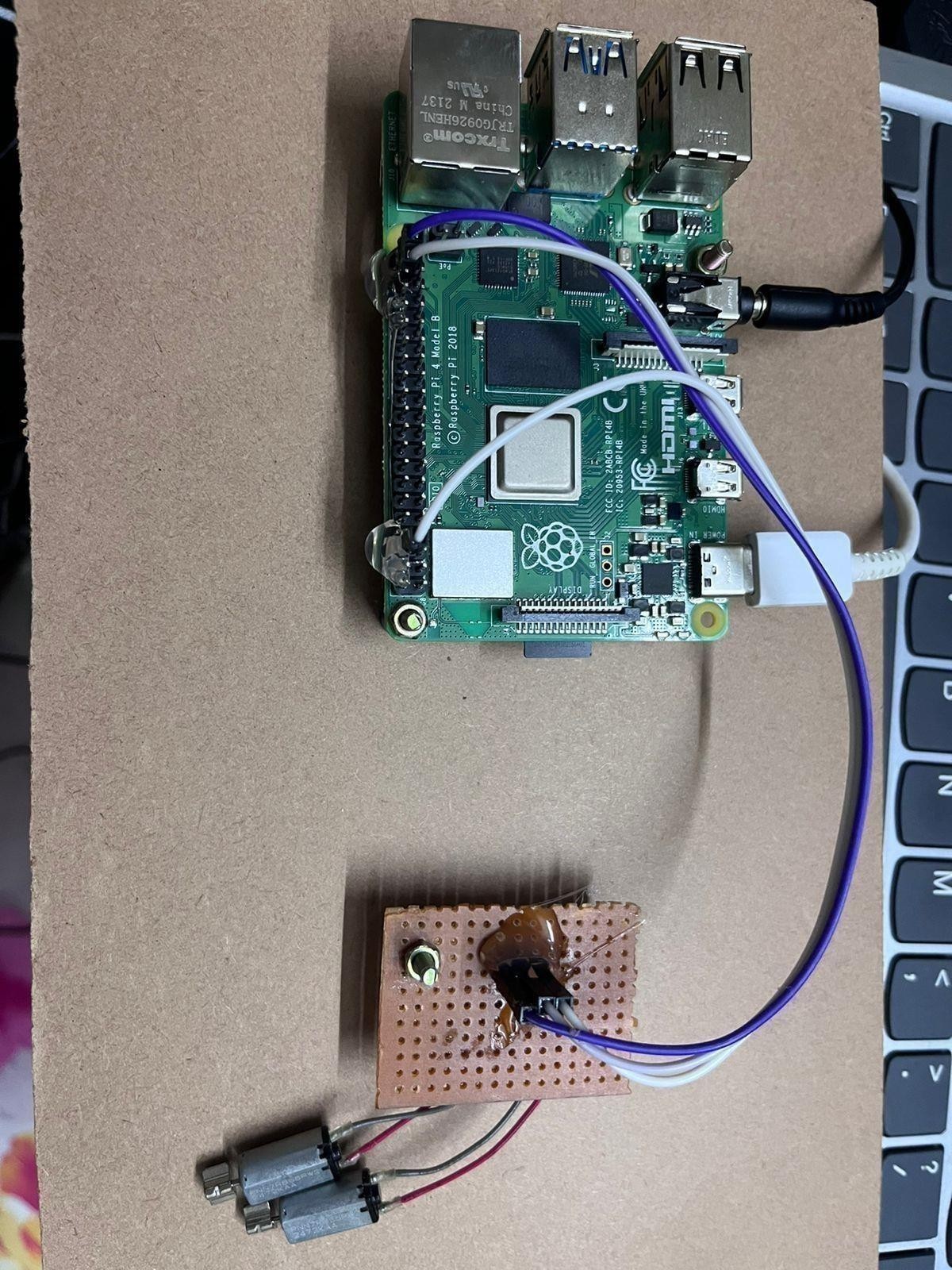
#### FIG NO.A.3 TV is detected with frame with accuracy of 92.61%



**FIG NO.A.4 Chair is detected with frame with accuracy of 98.31%**



#### FIG NO.A.5 AUDIO OUTPUT



**FIG NO.A.6 VIBRATION OUTPUT**

### REFERENCES

1. Shubham Melvin Felix, Sumer Kumar, and A. Veeramuthu, A Smart Personal AI Assistant for Visually Impaired People,2018.
2. Giancarlo Iannizzotto,Lucia Lo BelloAndrea,Nucita,Giorgio Mario Grasso, A vision and speech enabled, customizable, virtual assistant for smart environments,2018.
3. Prince Bose, Apurva Malpthak, Utkarsh Bansal, Ashish Harsola, Digital Assistant For The Blind,2017.
4. Md. Siddiqur Rahman Tanveer, M.M.A. Hashem and Md. Kowsar Hossain, Android Assistant EyeMate for Blind and Blind Tracker ,2015.
5. Kumar Yelamarthi, Navigation Assistive System for the Blind using a Portable Depth Sensor,2015.

[6] D. and J. H. Martin, “Logistic Regression,” in Speech and Language Processing, 2015

[7]J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, “Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[8]P.Y.Kumbhar,M.Attaullah,S.Dhere,andS.Hipparagi,“RealTime Face Detection and Tracking Using OpenCV,” International Journal for Research in Emerging Science And Technology, vol. 4, no. 4, pp. 39–43, Apr. 2014.

[9]B.-S. Lin, C.-C. Lee, and P.-Y. Chiang, “Simple Smartphone-Based Guiding System for Visually Impaired People,” Sensors, vol. 17, no. 6, p. 1371, 2013