GOA COLLEGE OF ENGINEERING
FARMAGUDI, GOA
DEPARTMENT OF ELECTRONICS & TELECOMMUNICATION ENGINEERING

2020 - 2021

DEVELOPMENT OF AI BASED VISUAL AID SYSTEM

by

Sahil Bhonsle (P.R.No.:201704606)
Crizel fernandes (P.R.No.:201705123)
Elaine Fernandes (P.R.No.:201704592)
Siddharth Gadkar (P.R.No.:201704608)

A project submitted
in partial fulfilment of the requirements
for the degree of
Bachelor of Engineering
in
Electronics & Telecommunication Engineering
GOA UNIVERSITY

under the guidance of

Dr. Shajahan M. Kutty
Assistant Professor,
Electronics & Telecommunication Department
Goa College of Engineering

CERTIFICATE

This is to certify that the project entitled

"DEVELOPMENT OF AI BASED VISUAL AID SYSTEM"

submitted by

Sahil Bhonsle	**P.R. No.:201704606**
Crizel Fernandes	**P.R. No.:201705123**
Elaine Fernandes	**P.R. No.:201704592**
Siddharth Gadkar	**P.R. No.:201704608**

has been successfully completed in the academic year 2020-2021 as a partial fulfilment of the requirement for the degree of BACHELOR OF ENGINEERING in Electronics & Telecommunication Department, at Goa College of Engineering, Farmagudi.

Internal Examiner
(Dr.Shajahan M.Kutty)

External Examiner

Head of Department,
Dr. H. G. Virani,
Professor, ETC Dept.

Place: Farmagudi, Ponda, Goa
Date:

PROJECT APPROVAL SHEET

The project entitled

"DEVELOPMENT OF AI BASED VISUAL AID SYSTEM"

by

Sahil Bhonsle	**P.R. No.:201704606**
Crizel Fernandes	**P.R. No.:201705123**
Elaine Fernandes	**P.R. No.:201704592**
Siddharth Gadkar.	**P.R. No.:201704608**

completed in the year 2020-2021 is approved as a partial fulfilment of the requirements for the degree of **BACHELOR OF ENGINEERING in Electronics & Telecommunication Engineering** and is a record of bonafide work carried out successfully under our guidance.

Project Guide,
Dr.Shajahan M.Kutty
Assistant Professor,
ETC Dept.

Head of Department
Dr. H. G. Virani
Professor, ETC Dept.

Principal
Dr. R B. Lohani
Goa College of Engineering

Place: Farmagudi, Ponda, Goa
Date:

Declaration

We declare that the project work entitled "DEVELOPMENT OF AI BASED VISUAL AID SYSTEM" submitted to Goa College of Engineering, in partial fulfilment of the requirement for the award of the degree of B.E. in Electronics and Telecommunication Engineering is a record of bonafide project work carried out by us under the guidance of Dr. Shajahan Kutty. We further declare that the work reported in the project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or other institute or university.

Sahil Bhonsle	Elaine Fernandes
Crizel Fernandes	Siddharth Gadkar

Place: Farmagudi, Ponda, Goa
Date:

Acknowledgement

This project would not have been possible without the guidance and support of our project guide Dr. Shajahan Kutty, whose deep knowledge and experience were essential for the development of this project.

We would also like to thank our principal, Dr. R B. Lohani and the Head of the Department, Dr H. G. Virani for providing us the infrastructure and facilities to carry out this project. We are also grateful to the faculty and staff for their help and support in this endeavour.

Contents

List of Figures

Abstract

Blind mobility is one of the major challenge encountered by visually impaired people in their daily life, greatly restricting a lot of activities. In our project we are developing an AI based smart system that focuses on face recognition of family members and object detection, integrated with an audio feedback.

We have designed a pair of smart glasses consisting of a camera module, a battery and a headset interfaced to the Raspberry pi. The camera continuously feeds images to Raspberry pi for object detection and face recognition followed by text to speech conversion. We have implemented the above capabilities that can be used as a visual aid for blind thus reducing the dependency and allowing them to do basic work by themselves.

Chapter 1

Introduction

1.1 Overview

According to the World Health Organization(WHO) fact sheet of 2018 [1] at least 2.2 billion people have a near or distance vision impairment. These people include those with moderate or severe distance vision impairment or blindness due to unaddressed refractive error (88.4 million), cataract (94 million), glaucoma (7.7 million), corneal opacities (4.2 million), diabetic retinopathy (3.9 million), and trachoma (2 million), as well as near vision impairment caused by unaddressed presbyopia (826 million).

The information acquired by a normal human being greatly depends on vision. Unfortunately, visually impaired people obtain this information through touch, listening and moving. Because of their reduced mobility and less knowledge about

the surroundings they cannot participate in various activities which creates social isolation for them. Blind people tend to depend on human assistance to do daily tasks. Wearable devices are advancing in terms of technology, functionality and size. They represent potential aid for people with physical and sensory disabilities that might lead to the improvements in their quality of life.

This project introduces a live object recognition system. The aim is to transform a visual world into an audio world with the potential to inform blind people about the objects in the surrounding. Our main focus is face recognition of the family member and object detection. We have tried to portray a cost effective and user-friendly device which would help a blind person navigate smoothly. Objects detected from the scene are represented by their names and converted into speech. Video is captured with a camera device and streamed for real time image recognition.

1.2 Machine Learning for Visual system

Machine Learning is the study of computer algorithms that improve automatically through experience and by the use of data. It is a subset of Artificial Intelligence. Machine Learning Algorithms build a model based on sample data, known as the "Training data", in order to make predictions or decisions without being explicitly programmed to do so. Computer Vision is an important branch of Machine Learning that deals with how computers can gain high-level understanding from

digital images or videos. It seeks to understand and automate the tasks the human visual system can do.

Computer Vision tasks include methods for acquiring, processing, analyzing and understanding digital images and extraction of high-dimensional data from real world and making decisions. In our project the images are acquired using a camera module and fed to the machine learning model for processing and decision making.

1.3 Single Shot Detector

The traditional target detection methods such as Local Binary Patterns (LBP) [], Scale Invariant Feature Transforms (SIFT) [], Histograms of Oriented Gradient (HOG) [], and Haar-like (Haar) [], are based on hand-crafted features. This feature extracted by the traditional target detection methods has obvious limitations. Firstly, the feature extraction is complex and the calculation speed is slow. Secondly, the artificial features largely limit the application scenarios of the algorithm. It is difficult to satisfy the needs of real-time detection on a complex and large dataset.

In recent years, a lot of target detection algorithms based on the convolutional neural network (CNN) have been proposed to solve the problem of poor accuracy and real-time performance of commonly used traditional target detection algorithms. Among various target detection methods, SSD is relatively fast and accurate because it uses multiple convolution layers of different scales for target

detection. SSD takes the Visual Geometry Group(VGG16) [] as the basic network, and adopts a pyramid structure feature layer group (multi-scale feature layer) for classification and positioning. It uses features extracted from shallow networks to detect smaller targets, and larger targets are detected by deeper networks features.

SSD algorithm is based on feed-forward convolutional neural network which produces collection of fixed-size bounding boxes of different aspect ratios and scores for the presence of objects in those boxes. First few layers are designed based on standard architecture which are used for high quality image classification — also called base layers – followed by auxiliary structure to produce detection with various features. Convolutional feature layers are added at the end of the base layer which decrease in size progressively and allow detections at multiple scales thus tackling the scale problem.

SSD algorithm does not resample pixels or features for bounding box hypotheses and gives accuracy as good as the algorithms which do, thus resulting in significant improvement in speed at high accuracy. Though SSD algorithm was not the first to do this, due to series of improvements – such as adding multiple layers of progressively decreasing size resulting in detection at multiple scales – SSD has achieved high-accuracy using relatively low-resolution input, further decreasing the detection speed.

1.4 Support Vector Machines

Face recognition is a major research area in computer vision. Algorithms must recognize faces by extrapolating from the training samples. Support vector machines are one of the supervised machine learning algorithms which can be used for both classification problems and regression problems. SVMs are one of the most robust prediction methods, being based on statistical learning framework. Since these are supervised learning algorithm, desired output is provided along with the training data.

The main objective of SVMs is to divide the dataset into different classes by using a hyperplane/decision boundary. Hyperplane is a plane or a space which divides the dataset into different classes. This hyperplane is determined by support vectors. Support vectors are the datapoints of each class that are closest to the hyperplane. Support vectors are the supporting point of this algorithm. So, even if all the other datapoints are removed, the algorithm will not change. The hyperplane is equidistant from all the support vectors, that is, it has maximum margin. Sum of the distances between support vectors and the hyperplane is maximised. SVMs look at the extreme case which is very close to the boundary and it uses that to construct its analysis and this makes SVM algorithm very different to most of the other machine learning algorithm and most of the time SVMs perform better than other non-SVM algorithms.

Another method is Principal component analysis (PCA). PCA is the process of

computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest. PCA is used in exploratory data analysis and for making predictive models. According to the study [] which compares SVM's and PCA methods, it was found out that the identification performance for SVM is 77-78 per cent versus 54 per cent for PCA. For verification. The equal error rate is 7 per cent for SVM and 13 per cent for PCA. Thus our face recognition model will be based on SVM algorithm.

Chapter 2

Literature survey

2.1 CNN architecture

CNN [2]is the foundation block of object detection. CNN is of great importance. Let us learn about CNN Architecture. CNN model consists of many convolution and pooling layers. These layers help to extract features from the input image by convolving the input image with filters. Each convolution layer will have different filters(i.e predefined kernel) thus extracting different features. Initial layers will extract basic features and further layers will extract complex features. The pooling layer is down-sampling in order to reduce the complexity for further layers. Then the final layer is a fully- connected layer which connects each node in a fully connected layer to the previous and next layer.

Figure 2.1: CNN Architecture

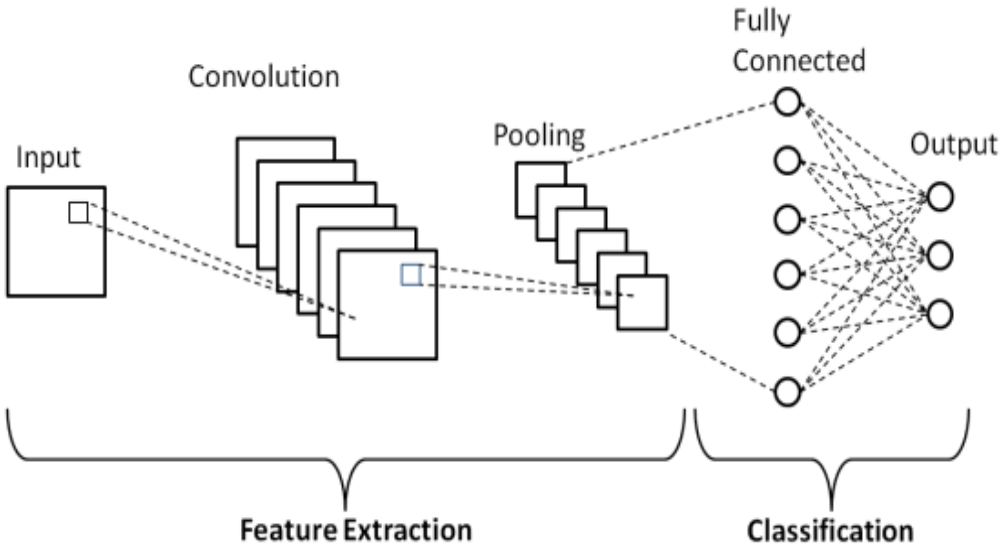


Figure 2.1: CNN Architecture

In the convolution layer the predefined kernel slides over the image thus producing the output layer. The kernel size is smaller compared to the size of the image.

2.2 Deep learning based object detection

There are many deep learning methods such as R-CNN , fast R-CNN , faster R-CNN, SSD and YOLO

- Region Proposal based Methods: R-CNN, fast R-CNN, faster R-CNN

- Regression based Methods: YOLO and SSD

2.2.1 Region Proposal based Method

Region Proposal based methods that first selects regions of interest by selecting many bounding boxes and then extracting features from each bounding box to find presence of object. Finally combining all overlapping boxes into one bounding box.

R-CNN means Region based CNN method. Girshick et al. [3] proposed this novel object detection method called R-CNN. R-CNN first generates region proposals using the Selection Search algorithm and then extracts features. Before feature extraction each proposal is wrapped to a fixed size bounding box and then given to the network for feature extraction. Feature extraction is done for each proposal making it computationally expensive.

Fast - RCNN [4] 9 times faster than R-CNN. It gives the image to CNN for feature extraction and output feature map is produced. Regions of interest are obtained on the feature map. The region is then given to the ROI pooling layer which is then given to the fully connected layer.

Faster-RCNN [5] generates a feature map through CNN. Faster-RCNN uses the Region proposal network to obtain the region of proposal. It reduces the computational cost of region proposals by using RPN. Thus making it suitable to function at real time processing.

2.2.2 Regression based Methods

YOLO [6] stands for You Only Looks Once is a single neural network predicts the class and confidence in one evaluation. YOLO consists of 24 convolutional layers and 2 fully connected layers. YOLO is fast and accurate thus making it suitable for real time processing. But it struggles with small objects that appear in group

SSD [7] short form of single shot detection runs the convolution layer once and gets the feature map and then several convolution layers for reducing sizes thus able detect objects of various scales. SSD architecture consists of VGG-16 for feature extraction and additional convolution layers are added which reduces size of feature map and reduces computation. SSD evaluates to multiple default boxes of different sizes. So find the default box that most overlaps with the ground truth box containing objects. The boxes containing objects are treated as positive and rest are negative. So SSD is faster than faster R-CNN and accuracy is similar to faster R-CNN thus we choose SSD.

2.3 Face recognition

Face recognition is a way of recognizing a human face through facial recognition system. The facial recognition system maps facial features from image or video. The algorithm will map the face, measuring the distance between eyes, nose, mouth and so on. Finally outputs a vector for each face which uniquely identifies

each face among all others in training set.

We use SVM [8] Support vector machines as it generates a hyperplane between points belonging to 2 classes thus avoiding the chance of miss classification. It finds the hyperplane which will have the maximum margin i.e. maximum distance between points of both classes. The points which are closer to the hyperplane are called support vectors. Also used for multi class recognition combining SVMs.

2.4 Related works

Many other assistive systems were developed to help visually impaired individuals such as Smart cane [9] which would recognize faces using Viola jones algorithm and give output in form of vibration pattern for each person. It was restricted to only few individual as it would become difficult for blind person to remember so many patterns and it does only face recognition not all time some individual will be there so when no individual is there it will do nothing.

Other project [10] for blind also involves development of glasses which employees ultrasonic sensors, depth sensors and cameras which detects obstacles and potholes thus helping the blind to navigate properly. It doesn't involve any Machine learning to detect any object in front or face. It can only be used to navigate blind. If any object or individual comes in front then it will not able to detect the object or may detect it as obstacles.

The stated project [11] uses EPSON BT-300 which has built in camera and voice output module. It uses YOLO v3 for object detection which runs on back-end server. The smart glasses captures image and uploads to back-end server and uses YOLO for object detection. The result are downloaded by the smart glasses and audio output is generated. Its produces good accuracy but the time required for uploading and downloading from back-end server is around 4 seconds. The blind individual have to wait for 4 seconds for output to produce. Which makes it unsuitable for real time application.

Chapter 3

Project Objective

3.1 Designing a pair of glasses

We aim at designing a pair of smart glasses incorporating a camera module, a battery and a headset interfaced to the Raspberry Pi. Since the raspberry pi is bulky, we will make use of a customized pi consisting of only those components which are of our concern. This will help us make the hardware portable as well as less expensive.

3.2 Machine learning based model for Face Recognition

The other objective of our model is to recognise the faces of the house members. In order to achieve this we will train a ML based face recognition model. The dataset used will have images of the group members as well as

some unknown faces for better results.

3.3 Achieving low latency Object Detection

Speed and latency are the major concerns of any object detection model. In our project we deal with real time object detection thus the model should have low latency and a high mean average precision. We intend to detect and classify maximum number of object captured by the frame.

Chapter 4

Design

In this chapter, we will discuss the design of the proposed model, various hardware and software requirements,along flow of entire algorithm.We use raspberry pi for our project mainly because it is easily available, low cost device and we can customize based on our requirements. Also, the libraries are open source which makes it cost effective. We will also describe the algorithm for the machine learning model used. Finally, we will discuss about the integration of all the subsystems to form the final prototype.

4.1 Model Architecture

Our project will designing a pair smart glasses consisting of a camera module, a battery and a headset interfaced to the Raspberry pi.Since the final model is

a wearable, it should be light and comfortable. The underlining hardware used a Raspberry Pi microcontroller. The Raspberry Pi is bulky with respect to the frame, so to account for this we will be customizing the raspberry pi containing only those components which are of use to us. This will make the hardware lighter as well as cost effective. The camera module is mounted on the center of the frame. The captured images will be continuously fed to Raspberry pi. The trained model will analyse the captured frame, followed by text to speech conversion. The battery is chosen such that it is light and has a good power profile. We will be using a Li-po rechargeable battery.

4.1.1 Block Diagram

The block diagram of the proposed system is shown below.

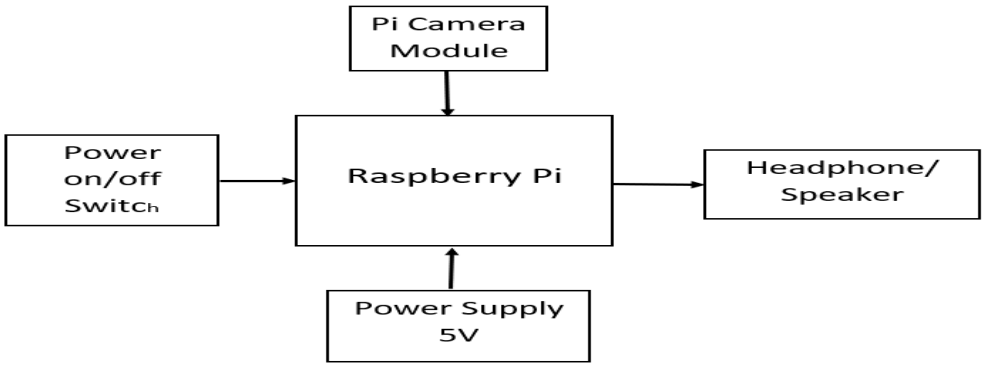


Figure 4.1: Block Diagram of the Proposed System

In this project we are using Raspberry Pi 4 Model B. The Raspberry Pi is powered using 5V DC power supply. A camera module is connected to Raspberry Pi via CSI

port. There is a push button which will turn on/off the Raspberry Pi.When the push button is turned on, the Raspberry Pi will power up. Continuous frames will be captured by the interfaced camera module and will be fed to Pi. The captured frames will be then processed for Object Detection and Face Recognition. This is followed by the text to speech conversion.

4.1.2 3D Model

The figure below shows the 3D view of the Model. This was done using Fusion 360 software. The customized raspberry Pi is mounted on the left arm of the glasses, the camera module is mounted in the centre of the frame and the battery on the right arm.

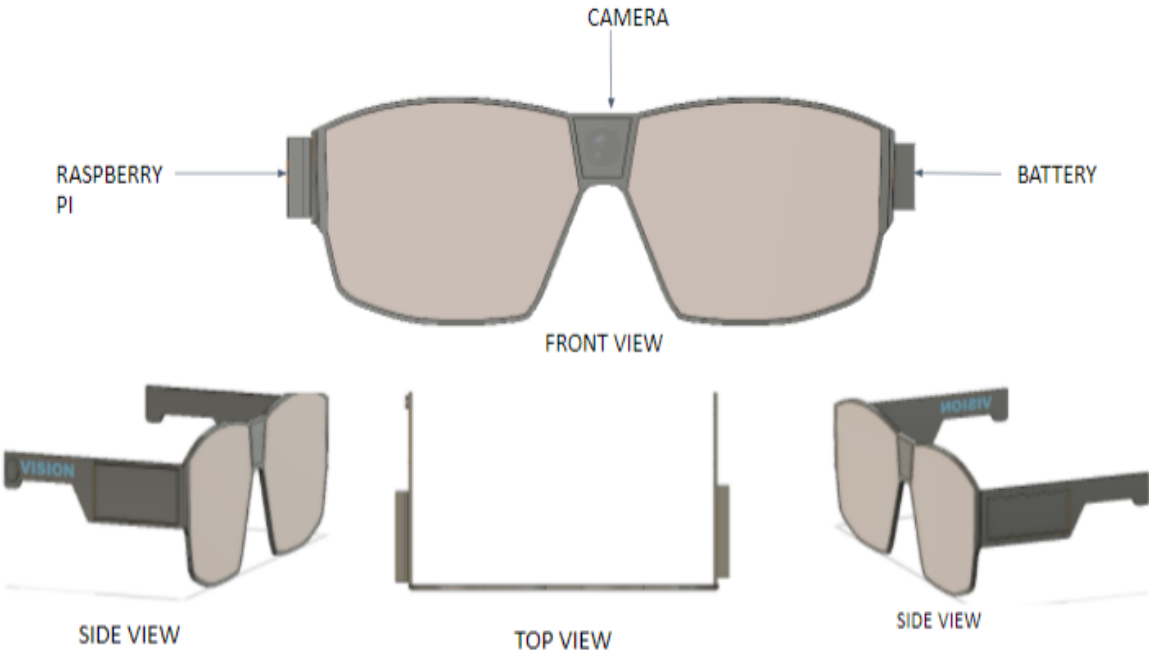


Figure 4.2: 3D Model

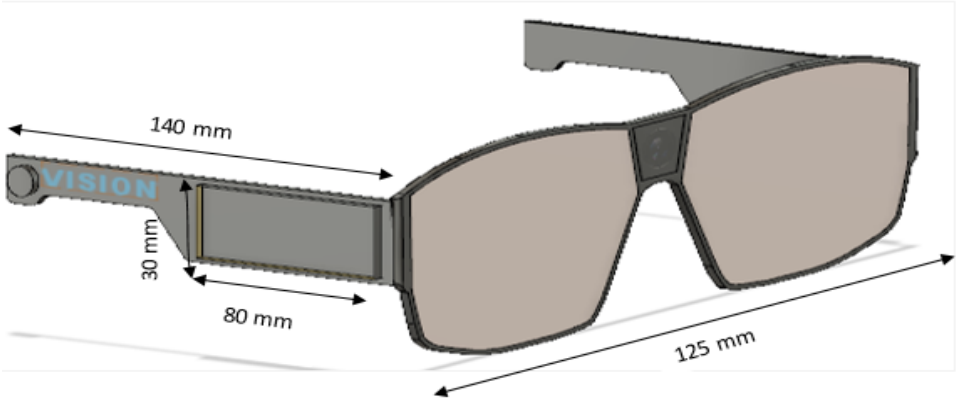


Figure 4.3: 3D Model Dimensions

4.1.3 Modified Raspberry Pi Schematic

The Raspberry Pi is bulky with respect to the frame, so to account for this we will be customizing the raspberry pi containing only those components which are of use to us. The various components included in the modified Raspberry Pi are as follows:

1. USB- C connector- To connect a 5V DC power supply.

2. 40 Pin GPIO Header-To connect physical input/output devices like Light Emitting Diode(LED) and switches to the Pi board to control the programms.The pins are split into 4 main types they are 3.3V, 5V , GPIO 12, Ground.

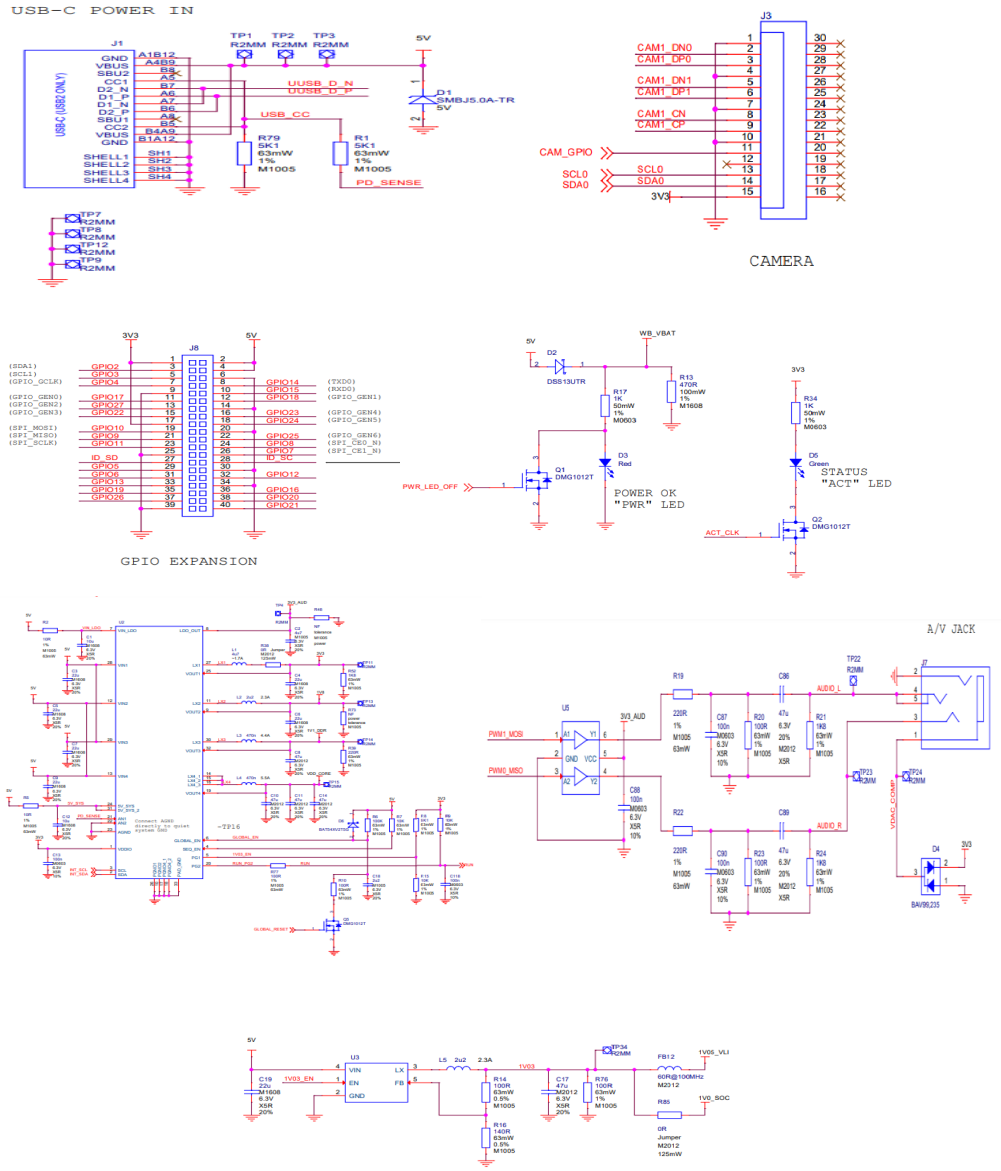


Figure 4.4: Modified Schematic of Raspberry Pi 4 Model B

3. Camera Port- 2 Lane MIPI CSI camera port is used to connect the Raspberry Pi Camera Module v2 that captures images continuously which are then fed to Raspberry pi.

4. Audio/Video Jack- The analog audio output can drive 32 Ohm headphone directly The text to speech conversi on output is obtained via the audio/video Jack.

5. Power LED(Red)- Indicates that power has been provided to the board.

6. Status LED(Green)-Indicates SD card activity.Flashes when read or write is in progress.

7. Run, Global EN Header

4.1.4 Power Supply

Voltage provided by batteries are typically 1.2V, 3.7V, 9V and 12V. However, the voltage required by the raspberry pi is 5V and hence we need a mechanism to provide a consistent 5V supply. There are two ways to achieve this: a) Voltage Divider b) Voltage Regulator

a) Voltage Divider

The Voltage divider circuit is given in figure 4.5. The voltage source is 11.1V.

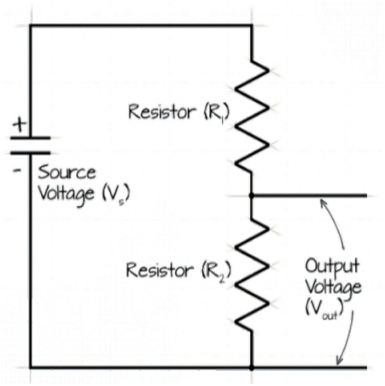


Figure 4.5: Voltage Divider Circuit

We fix the resistor R1 to 1000 ohm and the output voltage to 5V using the voltage divider equation given below we get the value of R2 equal to 820 Ohm:

$$V_{out} = \frac{V_s * R_2}{R_1 + R_2}$$

However this circuit has some disadvantages. There is always current flowing through the resistors, thus as the load current varies the output voltage will also vary and not be fixed to 5V.

b) Voltage Regulator

LM7805 is a linear voltage regulator that produces a regulated 5V output. It can deliver up to 1.5A of current (with heat sink). A voltage regulator, maintains a fixed output voltage irrespective of the change in input voltage or load current. The voltage regulator circuit is as shown below.

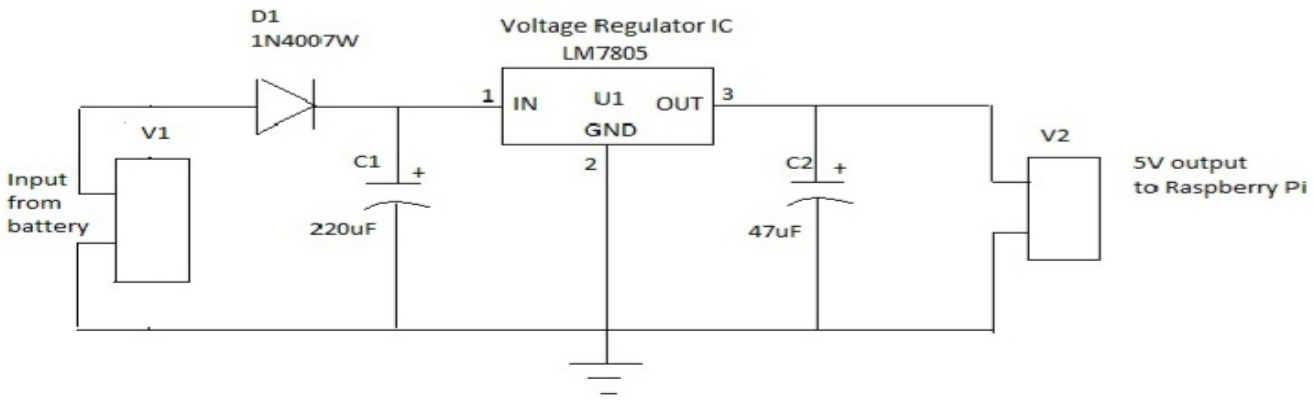


Figure 4.6: Voltage Regulator Circuit

The advantage of using a voltage regulator over a voltage divider is that as time progresses, the battery drains which will charge the input voltage, but this will have minimum effect on the regulated voltage and we get a constant output voltage of 5V. LM7805 also adds a provision for heat sink.

4.2 Hardware Components

The different hardware components required to design the smart glasses are as follows:

4.2.1 Raspberry Pi

We need cheap, easy to use hardware to run the object detection and face recognition program and Raspberry Pi turns out to be most suitable among other single board computers. It captures frames from the camera and process the frames

for the necessary output. For our project, we have used Raspberry Pi 4 Model B(Figure) which comes with Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz with 8GB LPDDR4-3200 SDRAM, standard 40 pin GPIO header and requires 5V/2.5A DC power supply via USB-C connector.

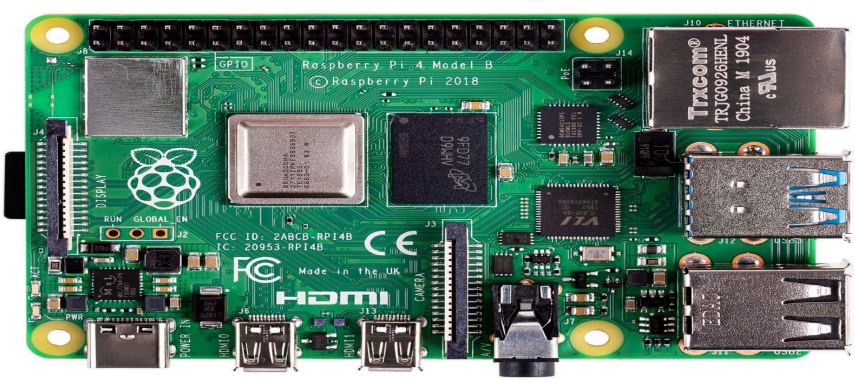


Figure 4.7: Raspberry Pi 4 model B

4.2.2 Raspberry Pi Camera Module v2

The camera that we have used for this project is Raspberry Pi Camera Module v2 which comes with a Sony IMX219 8-megapixel sensor. This sensor is capable of capturing videos at 1080p30, 720p60 and VGA90 modes, as well as still capture. It can be connected to Raspberry Pi via CSI port.

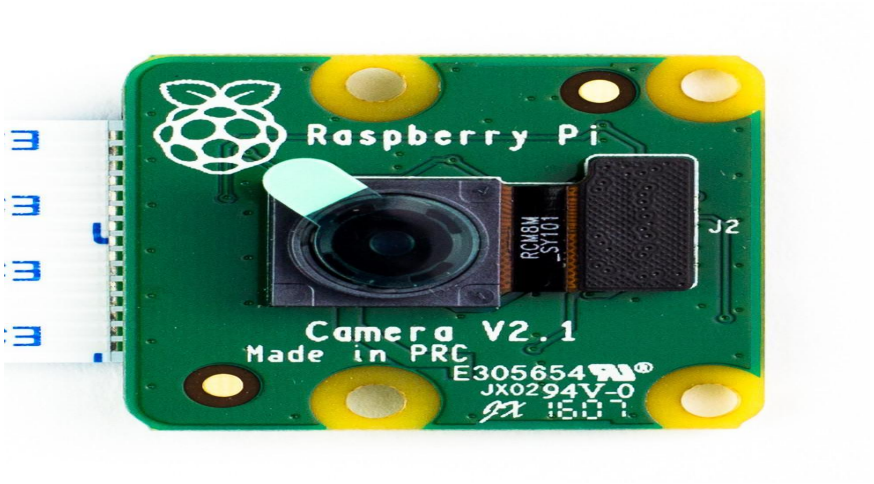


Figure 4.8: Raspberry Pi Camera Module V2

4.2.3 Battery

We are using a 12V Lithium ion (Li-ion) 1860 Rechargeable Battery pack 3S1P. It is compatible for Cameras, CCTV, Disks. The battery weighs 45gm and has dimensions of 68 x 55 x 20 (mm).It has a battery capacity of 2000 Milli-Amp Hours.

Figure 4.9: Li-ion Battery

It delivers a normal voltage of around 11.1V. It has a discharge cutoff of 9V and a full charge voltage of 12.6V. The DC output current is about 1-3A. the continuous discharge current is 5A and instantaneous discharge current is 7A.This is a 3S1P battery, 3 sets in series and 1 set in parallel. At a temperature of 25 degree Celcius, the deviation between the pure resistance discharge test result and the nominal data is controlled within 10 per cent. The battery is charged using a CC CV, S Lithium -ion battery charger.

4.2.4 LM7805

LM7805 is a three terminal voltage regulator with input pin for accepting incoming DC voltage, ground pin for establishing ground for the regulator and output pin that supplies the positive 5volts with output current of up to 1.5A.This regulator provides internal thermal-overload protection, high power dissipation and internal short circuit current limiting capability. LM7805 Ic is shown below.

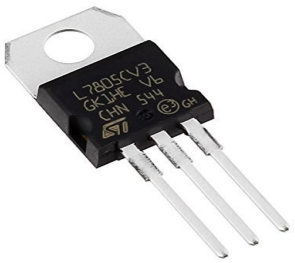


Figure 4.10: LM7805

4.2.5 Push Button

Push Button is a basic input device that can be connected to any controller or processor like Arduino or Raspberry Pi.A Push Button in its simplest form consists of four terminals.For our project we connect the Raspberry Pi's general purpose input output ports (GPIO) to a momentary tactile push button switch.When the button is pushed the raspberry camera module will capture the images followed by further processing.

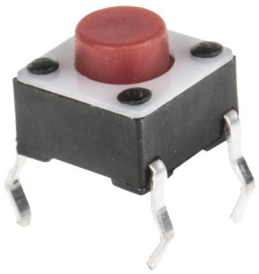


Figure 4.11: Push Button

We connect one side of the switch to an input pin on the Raspberry Pi, in this case we use pin 10. The other side of the switch we connect to 3.3V on pin 1 using a resistor.

4.3 Software Components

The following are the software components that we used for the project:

4.3.1 Raspberry Pi OS

Raspberry Pi OS: Raspberry Pi OS (Raspbian) is a 32-bit debian based Linux distribution operating system designed for Raspberry Pi. It is open-source distribution. After the initial build was completed in June 2012, which was started as an independent project; it has been officially provided by Raspberry Pi foundation as their primary OS for the Raspberry Pi SBCs since 2015. As of 2021, the 64-bit version of this OS is in beta and is not suitable for general use. The default UI is LXDE and has apt package manager with Monolithic 5.4.83 kernel.

4.3.2 OpenCV, NumPy and Imutils

OpenCV, NumPy and Imutils: Open Source Computer Vision Library was library mainly aimed at real-time computer vision applications. It is a cross-platform(Linux, Windows, MacOS, Android, iOS, etc) library originally developed by Intel. Primarily written in C++, now has bindings in Python, Java, MATLAB, etc.

NumPy is a open-source python library designed for supporting multi-dimensional arrays with a large amount of high level mathematical functions to operate on these arrays.

Imutils is a python library which has a various functions for basic image processing such as resizing, rotation, etc. along with OpenCV.

4.3.3 PuTTY

PuTTY is a free and open-source terminal emulator, serial console and network file transfer application. It is an easy-to-use telnet and SSH program for Windows.The IP of the RapsberryPi required for the PuTTY setup can be found obtained using an IP scanner. It can also connect to a serial port.

4.3.4 VNC

VNC is a graphical desktop sharing system that allows you to remotely control the desktop interface of one computer (running VNC Server) from another computer or mobile device (running VNC Viewer). VNC connect is included with Raspbian for RaspberryPi. VNC has to be enabled in the RaspberryPi configuration. VNC Viewer transmits the keyboard and either mouse or touch events to VNC Server, and receives updates to the screen in return.

4.3.5 Fusion 360

Fusion 360 Autodesk Fusion 360 is a cloud-enabled design platform which has all the tools needed to go from design to fabrication. It is free for hobbyist, students, educator, etc. Fusion 360 offers CAD, CAM and CAE features. In our project, we used this software to make a 3D model of the custom glasses which can hold our Raspberry Pi, its power source and the Pi Camera.

4.4 Machine Learning for Detection

In this section, we will discuss about the machine learning model, flowchart and working.

4.4.1 Algorithm design for face recognition using SVM

Algorithm used to build face recognition model is as follows:

Input: Image Dataset

1. begin model
2. define training dataset
3. define test dataset
4. preprocess the training dataset
5. extract feature embeddings and labels from training dataset
6. define classifier
7. train model(define epochs, batch size)
8. generate confusion matrix using test dataset.

In line 2, location where the training dataset is located is defined followed by line 3, where the location of test dataset is defined. In line 4, image dimensions like width, height, type of image to be given to the feature extractor are defined. In line 5, all the feature embeddings are extracted along with its corresponding labels. After this, in line 6, classifier is defined(in our case its Support Vector

Classifier) followed by training the model(in line 7) with the feature embeddings and its corresponding labels extracted from step 5. Finally, confusion matrix is generated using the test dataset to calculate the accuracy.

4.4.2 Flowchart for SSD model

Flowchart for SSD model Single shot multibox detector is based on feedforward convolutional network which produces bounding boxes and scores for the presence of objects. The initial network layers are called base network(based on a standard standard architecture used for high quality image classification). After this, convolutional feature layers are added to decrease the size progressively and predict the detection at multiple scales. Each layer added produces a fixed set of detection predictions using set of convolutional layers.

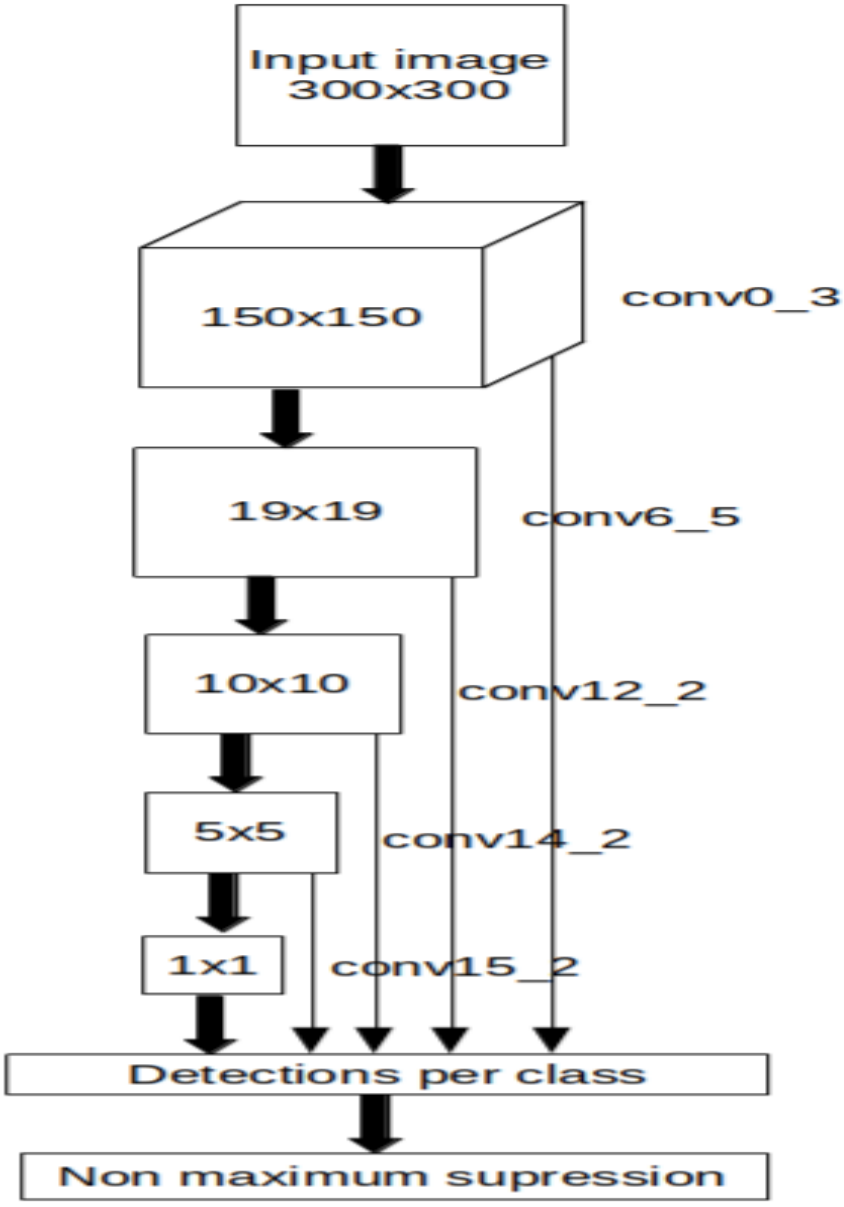


Figure 4.12: Flowchart for SSD model

MobileNet SSD is based on depthwise separable convolutions which are a form of factorized convolutions. These factorize a standard convolution into a depthwise convolution and a 1x1 convolution called a pointwise convolution.

1. Depthwise convolutional filters: applies single filter to each input chan-

nel(usually 3x3).

2. Pointwise convolutional filters: applies 1x1 convolution to combine the outputs of the depthwise convolution.

Main difference between traditional object detection methods and SSD is that those methods use standard convolution both filters and combines inputs into a new set of outputs in one step. While, on the other hand, depthwise separable convolution splits this into two layers - a separate layer for filtering and a separate layer for combining. This factorization has the effect of drastically reducing computation and model size.Standard input image size is 300x300 and the hidden layers have ReLU(Rectified Linear Unit) activation function and the output layer has softmax activation function.

4.4.3 Flow of entire algorithm

When the program starts running, it will capture images through the camera. The algorithm works by processing each image separately. So, when the first frame is captured, it will go through the pre-processing phase i.e. resizing the image to 300x300, converting it into blob, etc. After this phase, the processed image is given to the object detection algorithm.

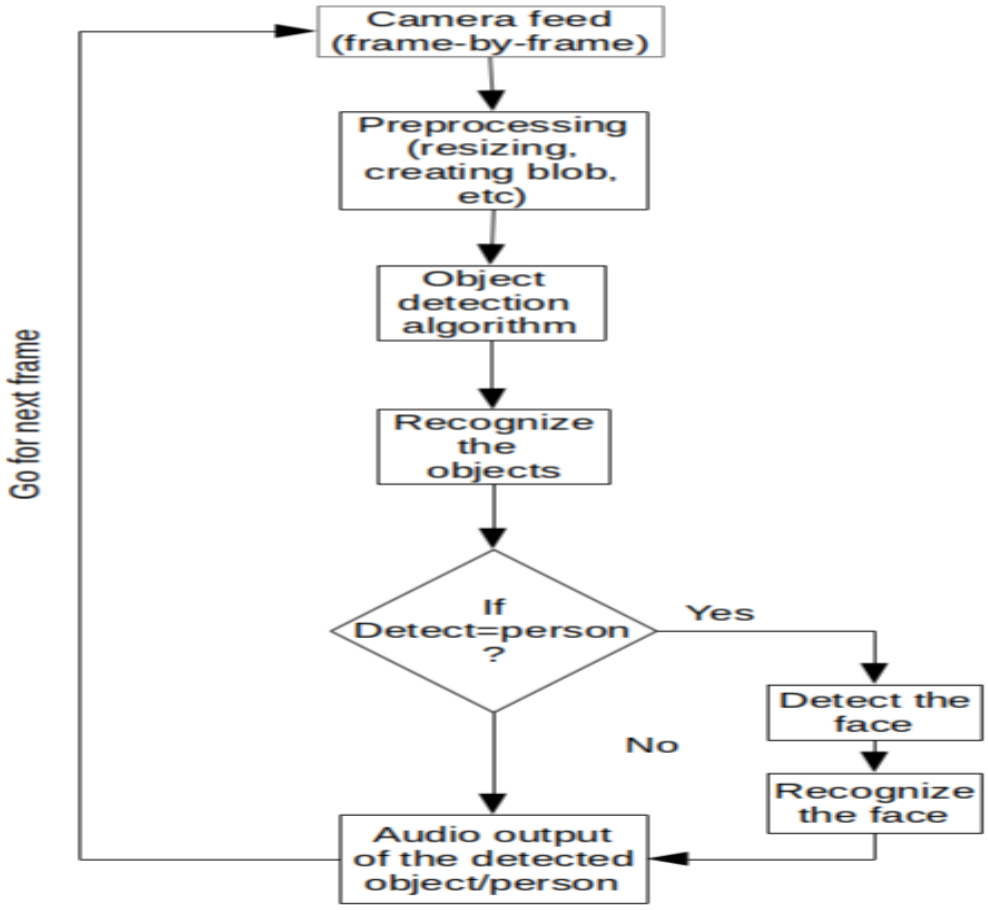


Figure 4.13: Flow of the entire Algorithm

The object detection algorithm will detect the objects from the image. Object detection is done by the the MobileNet SSD model. It takes input as 300x300 blob image, processes it, generates the output, then resizes the image and again generates the output, and this done till the image size is 1x1.Resizing is done in such a way that all the features are preserved and the reason why resizing is done is in order to tackle scale problem. After the objects are detected, it goes to check if any of the detected object is a person. If there is a person in the frame, then it branch from the main flow and go to detect the face, and then recognize the

person in that image. After that it will go to the audio output phase where audio output of the detected person/object will be generated.

If the detected object is not a person, then it will stay on main flow and the audio output will be generated. The algorithm will then go for the next frame.

4.5 System Integration

Once the smart glasses are assembled, the machine learning model is trained and is uploaded to the Raspberry Pi, we can put all of the components together as shown in fig below.

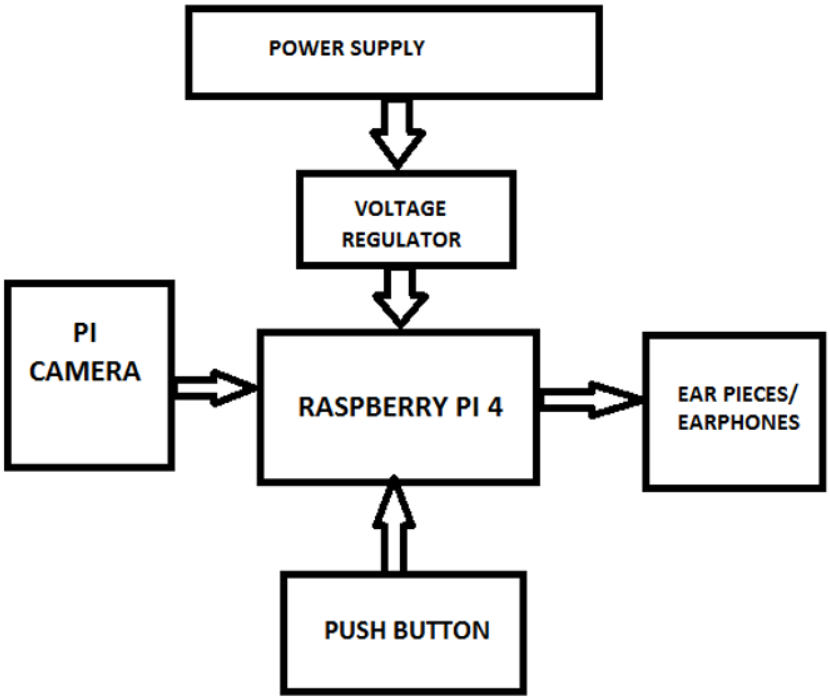


Figure 4.14: Block Diagram of Integrated System

We are using a 12 V power supply to power the Pi. Since the Raspberry Pi requires 5 V, we use a voltage regulator circuit that keeps the voltage level at 5 V. As the Pi does not have any on/off button we have configured a push button to turn the Pi on/off and run the detection model. The Pi camera module will continuously capture images and send them to the Raspberry Pi which will feed it to the object detection and face recognition model. Based on the output of model it will generate the text for the detected object or face, the Pi will convert this text to speech and will play the audio through the earphones connected to the pi.The push button is connected to 3 V (1) and to GPIO 15 (10) as shown in fig below.

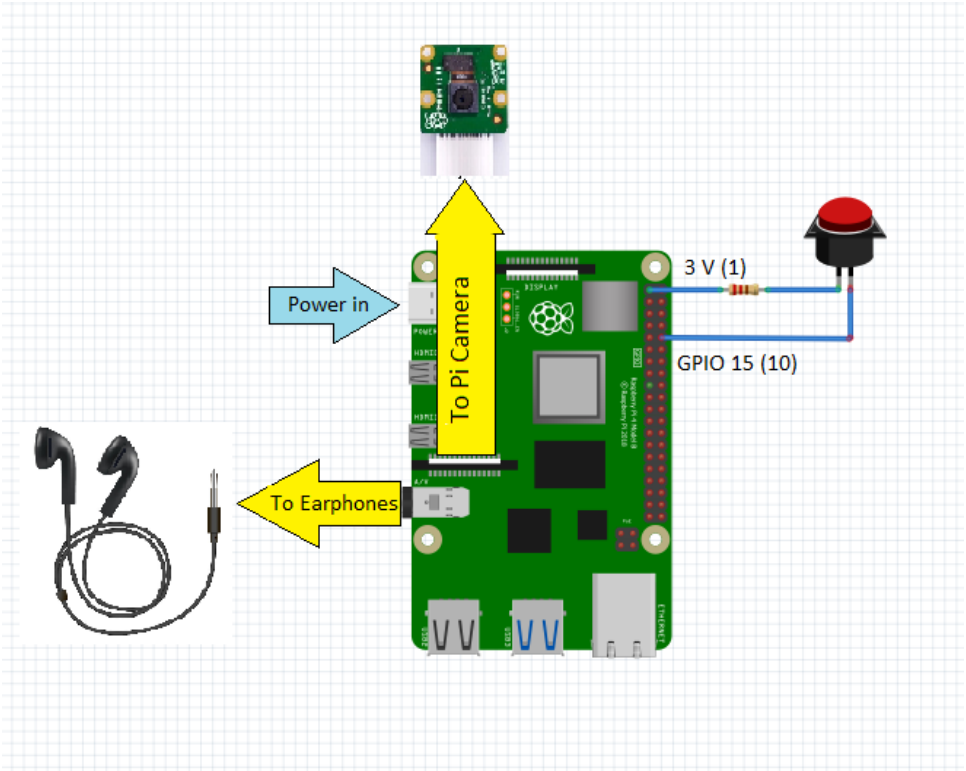


Figure 4.15: Internal Wiring of Integrated System

GPIO 15 pin is set as input pin. When the button is pushed the Pi will turn on and

will run the model loaded on the Pi and start detecting. The Pi Camera module is connected to the Pi via the CSI (Camera Serial Interface).The earphones are connected to the Pi via the 3.5 mm headphone jack as shown.

Chapter 5

Results and Discussions

In this chapter we will evaluate the machine learning models based on various parameters. The outputs produced by the Raspberry Pi have also been included.

5.1 The Face Recognition Model

5.1.1 Training the Face Recognition Classifier

The SVM based face recognition model was trained using a custom dataset consisting of images of group members as well as some unknown images. The custom dataset has 500 images of each member of the project group plus 500 images of unknown people. Therefore, the total images were 2,500 images.

80 per cent of the total images were used to train the SVM classifier and remaining 20 per cent were used for testing.

5.1.2 Evaluation Metrics of The Face Detection Classifier

Intersection of Union (IoU) is the ratio of the area of overlap of the predicted and ground truth boxes, to the area of their union. The IoU threshold was kept 0.5. An image with an IoU value greater that 0.5 was classified as True Positive (TP), else it was classified as False Positive (FP). No detection resulted in classifying the image as False Negative (FN).

The table below shows the confusion matrix:

313 (TP)	153 (FN)
153 (FP)	1777 (TN)

Table 5.1: Confusion Matrix

These classifications were used to calculate the Precision, Recall, Specificity and Accuracy values. Precision refers to the number of correct positive results divided by the number of positive results predicted by the classifier

$$Precision = \frac{TP}{TP + FP}$$

Recall refers to the number of correct positives results divided by the number of all relevant sample.

$$Recall = \frac{TP}{TP + FN}$$

Specificity or the true negative rate refers to the fraction of negative samples correctly predicted as negative by the classifier.

$$Specificity = \frac{TN}{TN + FP}$$

Accuracy gives us a measure of the total fraction of samples that were correctly classified by the classifier.

$$Accuracy = \frac{TP + FN}{TP + FP + TN + FN}$$

Using the above equations, the results were as follows:

Parameter	Value
Precision	67.2 per cent
Recall	67.2 per cent
Specificity	92.1 per cent
Accuracy	87.2 per cent

Table 5.2: Evaluation Matrices

5.2 The Object Detection Model

Latency in machine learning refers to the time taken to process one unit of data at a time. The unit of latency in seconds (time unit). Latency is an important measure as it is directly tied with the real time performance of the system. Lesser the latency better is the performance.

A Frame is a single image that captures a single static instance of a naturally occurring event. The machine learning model processes one frame at a time. The total time taken to capture a frame, process it and produce the output will be the response time and hence the latency of that particular frame.

The Python Time Module was imported to measure the performance of the model in terms of latency. The Time module measures the elapsed time for each frame. The response time of 200 consecutive frames was calculated and is plotted.

The graph below shows the Latency obtained for 200 consecutive Frames.

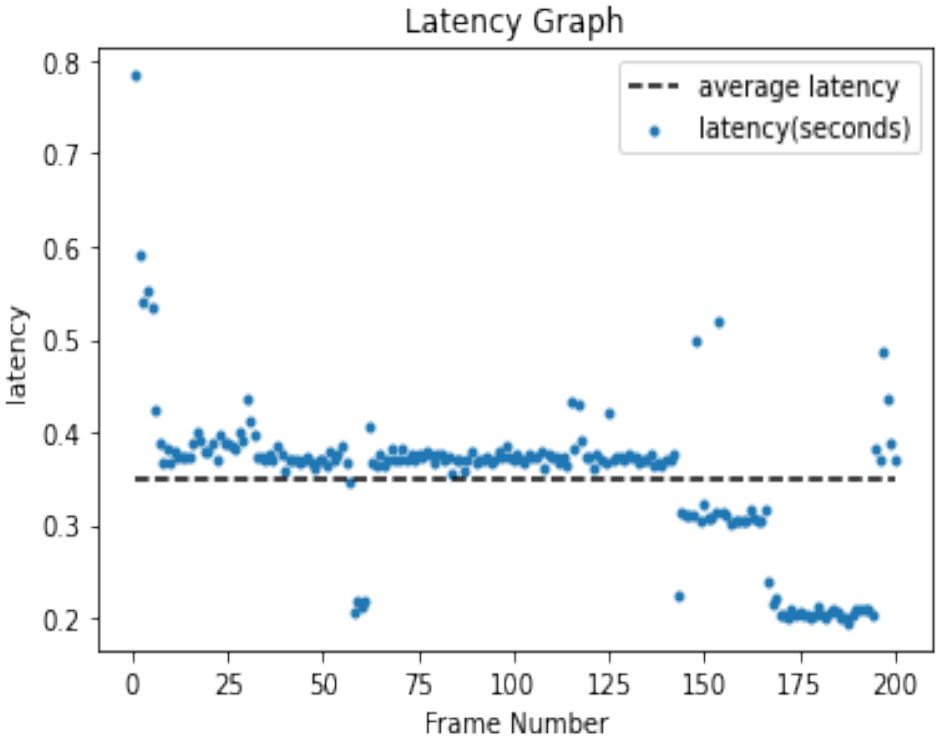


Figure 5.1: Graph of Latency v/s Frame Number

From this statistics the average latency was found out to be 0.34954248992499976 seconds.

5.3 Real Time Analysis

The output of the object detection model trained using Single Shot Detector algorithm predicted most of the objects in the captured frame. The model also displays the confidence score. Figures 5. Shows some of the detected frames.

Figure 5.2: Raspberry Pi output using real time camera input

The model was tested using real time input. It was successfully implemented on the Raspberry Pi and tested using real time camera feed.

Chapter 6

Conclusion

6.1 Challenges

The ultimate purpose of object detection is to locate important items, draw rectangular bounding boxes and determine the class of each item discovered. One of the challenges with object detection is viewpoint variation. An object viewed from different angles may look completely different. The classifier needs to be trained appropriately for maximum viewpoints. The object of interest could be occluded. Only a small portion of the object, just a few pixels can be visible.

Effects of illumination are drastic on the pixel level. This affects the capability of the detector to detect objects robustly. objects of interest may blend into the background, making them hard to identify. For creating a robust detector, one must ensure a good variation on training data, for different viewpoints, illumina-

tion conditions, and objects in different backgrounds.

Balancing the battery capacity with power requirement is a major challenge in term of hardware as high-power batteries are bulky. This leads to improper load balancing on the wearable.

6.2 Future Work

There are a lot of advancements that could be done to this project. Voice assistant integration could be done to automate most the things. This could also be used for providing directional and warning messages. Integration of google Maps could also be a great advancement. This could serve as a guide to unknown locations. A GPS tracker could also be incorporated such that the family member could monitor the patient.

Nowadays Many wearables like watches have an activity tracker such as heart rate, pulse rate, calories burnt etc. Incorporating this into our project could be plus point.

In terms of hardware advancements, we could go for sophisticated cameras having a wider-angle view.

Bibliography

[1] https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017

[3] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142-158, 1 Jan. 2016

[4] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015

[5] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017,

[6] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016

[7] Liu et al "SSD: Single Shot MultiBox Detector", 2016.

[8] Guodong Guo, S. Z. Li and Kapluk Chan, "Face recognition by support vector machines," Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, 2000

[9] Yongsik Jin, Jonghong Kim, Bumhwi Kim, Rammohan Mallipeddi, Minho Lee"Smart Cane: Face Recognition System for Blind ",Kyungpook National University Daegu, Korea, 2015,

[10] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," in IEEE Transactions on Consumer Electronics, vol. 63, no. 3, pp. 258-266, August 2017.

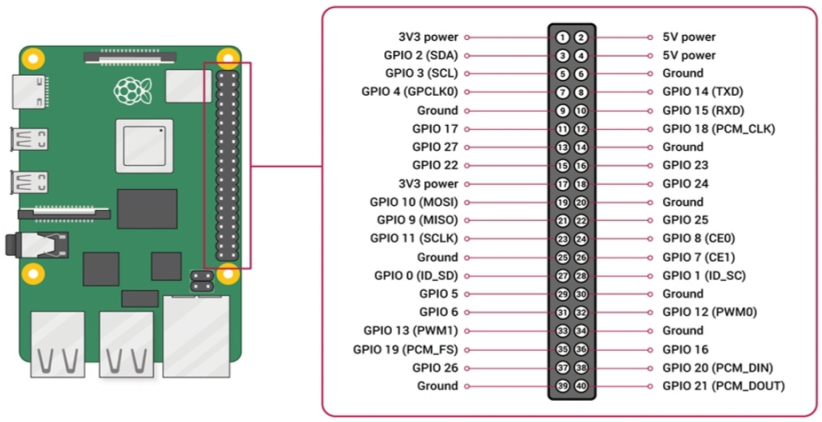
[11] J. -Y. Lin, C. -L. Chiang, M. -J. Wu, C. -C. Yao and M. -C. Chen, "Smart Glasses Application System for Visually Impaired People Based on Deep Learning," 2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), Rajpura, India, 2020

Appendix A

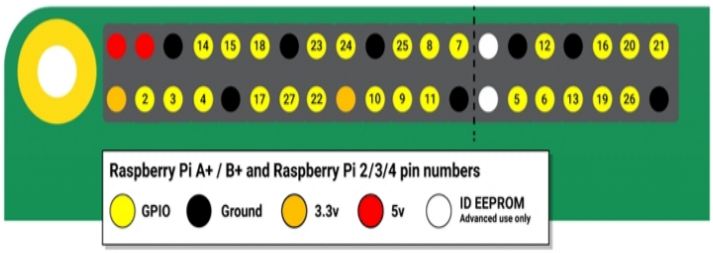
Appendix

GPIO Pins of Raspberry Pi 4 B

A powerful feature of the Raspberry Pi is the row of GPIO (general-purpose input/output) pins along the top edge of the board. A 40-pin GPIO header is found on all current Raspberry Pi boards (unpopulated on Pi Zero and Pi Zero W). Prior to the Pi 1 Model B+ (2014), boards comprised a shorter 26-pin header.



Any of the GPIO pins can be designated (in software) as an input or output pin and used for a wide range of purposes.



Note: the numbering of the GPIO pins is not in numerical order; GPIO pins 0 and 1 are present on the board (physical pins 27 and 28) but are reserved for advanced use (see below).

1. Voltages

Two 5V pins and two 3V3 pins are present on the board, as well as a number of ground pins (0V), which are unconfigurable. The remaining pins are all general purpose 3V3 pins, meaning outputs are set to 3V3 and inputs are 3V3-tolerant.

2. Outputs

A GPIO pin designated as an output pin can be set to high (3V3) or low (0V).

3. Inputs

A GPIO pin designated as an input pin can be read as high (3V3) or low (0V). This is made easier with the use of internal pull-up or pull-down resistors. Pins GPIO2 and GPIO3 have fixed pull-up resistors, but for other pins this can be configured in software.

4. More

As well as simple input and output devices, the GPIO pins can be used with a variety of alternative functions, some are available on all pins, others on specific pins.

- PWM (pulse-width modulation)

  – Software PWM available on all pins – Hardware PWM available on GPIO12, GPIO13, GPIO18, GPIO19

- SPI

  – SPI0: MOSI (GPIO10); MISO (GPIO9); SCLK (GPIO11); CE0 (GPIO8), CE1 (GPIO7) – SPI1: MOSI (GPIO20); MISO (GPIO19); SCLK (GPIO21); CE0 (GPIO18); CE1 (GPIO17); CE2 (GPIO16)

- I2C

  – Data: (GPIO2); Clock (GPIO3) – EEPROM Data: (GPIO0); EEPROM Clock (GPIO1)

- Serial

  – TX (GPIO14); RX (GPIO15)

Appendix

Appendix B

Data Sheets