BLIND SUPPORT SYSTEM

A project report submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE & ENGINEERING

BY

Shubham Mathur BE/25005/15

Apoorv Gaurav Agrawal BE/25034/15

Prageet.N.Gupta BE/25068/15

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

BIRLA INSTITUTE OF TECHNOLOGY, MESRA
JAIPUR CAMPUS, JAIPUR

SP-2019

DECLARATION CERTIFICATE

This is to certify that the work presented in the project entitled "Blind Support System**"** in partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering in Computer Science and Engineering of Birla Institute of Technology, Mesra, Ranchi, Extension Center Jaipur is an authentic work carried out under my supervision and guidance.

To the best of my knowledge, the content of this project does not form a basis for the award of any previous degree to anyone else.

Date: 7th May 2019

Dr. Shripal Vijayvargiya
(Associate Professor)
Birla Institute of Technology, Mesra ,Ranchi
Extension Centre,Jaipur

ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the completion of our project. All that has been accomplished is only due to such supervision and assistance and we thank all of them with folded hands for their support.

First of all , We would like to thank Dr. Shripal Vijayvargiya , for providing us with an opportunity to do the project work on the topic “Blind Support System” and giving us all the required support and guidance which made it possible to complete the project duly. We are extremely thankful to him for providing the much needed support and guidance.

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of Computer Science and Engineering which helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in laboratory for their timely support.

Thanking you all.

Shubham Mathur BE/25005/15 :
Apoorv Gaurav Agrawal BE/25034/15 :
Prageet.N.Gupta BE/25068/15 :

TABLE OF CONTENTS

1. Introduction

With so many people around the world having some sort of visual impairment (The World Health Organization estimates that 285 million people are visually impaired worldwide: 39 million are blind and 246 million others have low vision), there arose a need to help them with understanding the world in a better way. Advancement in technology particularly in the field of Machine learning , Artificial intelligence and Computer vision has made it possible to help these people by helping them by creating a simulated proxy visual environment.This project has been developed with the aim to help assist the blind by creating a model that understands the objects in front of the camera and alerts the user of their presence based on its distance from the camera.It also has an integrated face recognition model that has been trained to recognise previously seen and saved faces and can be further used to detect the emotions on these faces.Basically , The project focuses on the development of a support system for the visually impaired to help them with their daily lives. People with vision disability have great difficulty in perceiving and understanding the physical reality in an unknown environment.The basic functions that the project would perform are Object detection, Object recognition and Target Distance Estimation. The project uses a camera to identify and recognize the object the person is looking at and provide relevant information as auditory output thus guiding the person of the obstacle in front. The system developed also uses distance estimation techniques to calculate the distance between the obstacle and person just like how eyes perceive the immediate surroundings so as to provide a more detailed description of the environment. This project is intended to make people with such disabilities more aware of their environment and hence improve their Quality of Life. The project requires an external camera source that takes input as a video stream and produces the final output as an auditory result alerting the user.The project draws its inspirations from a Microsoft product called Seeing AI which is designed for the blind and low vision community and harnesses the power of AI to describe people, text, currency, color, and objects.Seeing AI is a Microsoft research project that brings together the power of the cloud and AI to deliver an intelligent app designed to help you navigate your day. Point your phone's camera, select a channel, and hear a description of what the AI has recognized around you.With the intelligent camera app the user can just hold up your phone and hear information about the world around you. Seeing AI can speak short text as soon as it appears in front of the camera, provide audio guidance to capture a printed page, and recognizes and narrates the text along with its original formatting. The app can also scan barcodes with guided audio cues to identify products, recognize and describe people around you and their facial expressions, as well as describing scenes around you using the power of AI.

2. Objective

The objective of this project is to help the visually impaired person to better understand their surroundings.It acts as a audio-visual aid to the human senses and is basically a virtual eye reporting all objects(including faces) to the user for ease of understanding of their immediate environment.The inspiration of this project is a similar project been developed by Microsoft called Seeing AI. Several different modules can be added to the existing module to increase its functionality although speed has to be compromised for increased modularity.

3. Model Description

COCO was an initiative to collect natural images, the images that reflect everyday scene and provides contextual information. In everyday scene, multiple objects can be found in the same image and each should be labeled as a different object and segmented properly. COCO dataset provides the labeling and segmentation of the objects in the images.COCO annotations are inspired by the Common Objects in Context (COCO) dataset which we have used for object detection. COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features: Object segmentation, Recognition in context, Superpixel stuff segmentation, 330K images (>200K labeled), 1.5 million object instances, 80 object categories, 91 stuff categories, 5 captions per image, 250,000 people with keypoints.The "categories" object contains a list of categories (e.g. dog, boat) and each of those belongs to a supercategory (e.g. animal, vehicle). The original COCO dataset contains 90 categories which further is divided into subcategories.

```
"categories": [
    {"supercategory": "person","id": 1,"name": "person"},
    {"supercategory": "vehicle","id": 2,"name": "bicycle"},
    {"supercategory": "vehicle","id": 3,"name": "car"},
    {"supercategory": "vehicle","id": 4,"name": "motorcycle"},
    {"supercategory": "vehicle","id": 5,"name": "airplane"},
    ...
    {"supercategory": "indoor","id": 89,"name": "hair drier"},
    {"supercategory": "indoor","id": 90,"name": "toothbrush"}
]
```

SYSTEM REQUIREMENT SPECIFICATION

3.1. Problem Statement

3.1.1. Purpose

The project focuses on the development of a support system for the visually impaired to help them with their daily lives. The World Health Organization estimates that 285 million people are visually impaired worldwide: 39 million are blind and over 246 million have low vision. People with vision disability have great difficulty in perceiving and understanding the physical reality in an unknown environment.The basic functions that the project needs to perform are Object detection, recognition and Distance Estimation. The project uses a camera to identify and recognize the object the person is looking at and provide relevant information as auditory output thus guiding the person of the obstacle in front. The system developed also uses distance estimation techniques to calculate the distance between the obstacle and person just like how eyes perceive the immediate surroundings so as to provide a more detailed description of the environment. This project is intended to make people with such disabilities more aware of their environment and thus improve their Quality of Life.

3.1.2. Document Conventions

The document uses the standard system of writing Software Requirements specification. The font used is Times New Roman. All headings have a standard size of 14 pts and have a bold typeface whereas all other information in paragraphs have a size of 11 pts. Heading points are separated by 2 newline spaces.All headings are present in order as they appear in the Table of Contents Page. References are attached as links to the Documents referred on the Internet. Other than that a list of some books referred for making the project are also mentioned.

3.1.3. Intended Audience and Reading Suggestions

The intended audience for this Document is all the developers who need information as to what the system does and how it does it so that they can understand the dynamics of this system and further suggest changes thereby helping improve the quality of the System. The system has been developed under the mentorship of our college professor as a part of our final year project and can be used as a starting point by developers interested in developing similar Systems.

3.1.4. Product Scope

This project focuses on development of a support system for visually impaired people which will perceive the world and supply the gathered information through auditory means so as to aid the blind person in carrying about daily chores and improve the quality of life of such a person. An ideal form of the project will include all the functions performed by the eyes namely object detection, object recognition, distance estimation, face recognition, character recognition, text to speech processing etc. Currently some of the modules namely object detection and Recognition, distance estimation and face recognition are being developed. Various other modules can also be implemented in the support of the project improving it further.

3.2. Overall Description

3.2.1. Product Perspective

The Product basically assists the blind person in taking day to day decisions by making them more aware of their surroundings. The input received from camera will help in identifying and recognizing the object, analyzing the distance of the object in front of the user and inform the user of this analyzed information through an in ear audio device as an auditory output. The system also contains a facial recognition module thereby helping the user identify any acquaintance present in front of the camera.

3.2.2. Product Functions

- ❏ After installing the android app, it takes in the input through the camera and processes the required input thereby providing information to the user about the object present in front of the camera.
- ❏ Alert the user about an incoming object using distance estimation module and help user take better decisions based on the voice output received.
- ❏ Recognizing faces and informing the user of the any acquaintance present in front of the camera.

3.2.3. User Classes and Characteristics

- ❏ Blind people can use this product to enhance their quality of life , for easing some day to day activities such as identifying obstacles and other objects. The generally use their other senses such as that of touch intuition and smell to recognize the objects. Using this product they can effectively save a lot of time since recognizing common objects would become easier.

3.2.4. Operating Environment

- ❏ Operating System : Android
- ❏ Language used : Java, Kotlin.
- ❏ Libraries used : JavaCV , OpenCV.
- ❏ Others : Tensorflow.

❑ Addition Libraries :-

1. Android.graphics :- It is used to draw graphics in android. It provides methods to draw oval, rectangle, picture, text, line etc.
2. Utils :- This library is basically used for pre defined data structures. The frames and recognition parameters are stored in Priority queue so as to provide a good flow of data. It also Contains the collections framework, legacy collection classes, event model, date and time facilities, internationalization, and miscellaneous utility classes.
3. Android.widget :- A widget is a small gadget or control of your android application placed on the home screen. Widgets can be very handy as they allow you to put your favourite applications on your home screen in order to quickly access them. You have probably seen some common widgets, such as music widget, weather widget, clock widget e.t.c.
4. Android.os :- This Library is used on for providing the system clock in the application so as to track the time for all the events and to plan the timeout for those events accordingly.
5. Android.Media :- Provides classes that manage various media interfaces in audio and video. The voice input and output modules uses this library and manages the operation using the function from this library. The media interface may include the recording of the voice and voice output.

3.2.5. Design and Implementation Constraints

❑ The camera should be of high resolution so that images can be better perceived and recognized.
❑ The phone should be kept at front pocket properly with camera facing out.
❑ It is viable to use the system only in safe conditions where incoming object speeds are slow unlike on Busy Roads and crowded public places. The system analyses the current information and informs the user on any incoming object but considering the obvious machine lag the system is better suited safe places such as inside the house where incoming speed of the objects is slow.

3.2.6. Assumptions and Dependencies

❑ A high resolution camera is present in the android phone and is placed properly.
❑ The incoming speed of objects is slow, preferably objects are stationary, so that spaced alerts can be generated.
❑ For recognizing the face, required information must be present in the database.
❑ Other senses such as auditory and olfactory work fine since alerts generated are audio based.

3.3. External Interface Requirements

3.3.1. User Interfaces

The product is designed for visually impaired and thus does not contain any Graphical Interface for its users. The users will interact with the app with the help of mic installed in the spectacle frame and a ear piece for the output of the processed information. The android app will be actively listening to the voice input from the user.

3.3.2. Hardware Interfaces

The Project is to be installed on an Android phones having Android version greater than 4.2 with 2GB RAM. The phone must have camera of at least 5MP. An active internet connection is required for the face recognition. The video captured by the camera will by processed in a series of images from which different objects will be detected.

3.3.3. Software Interfaces

High-level Overview includes the functionalities and the flow of data. The live feed recorded through the camera is continuously passed through the object detection and recognition module which recognizes the objects present in the scene and storing them in a list. As a new object appears, it is added to the list and the object no longer in the scene is removed from the scene. The object detection module has been trained using MS COCO Dataset which contains approximately 3 lakhs images of 90 most-commonly found objects. Trainer modules are tested and best module is selected after prioritizing the speed and accuracy. The identified image part of the object is continuously passed to the distance estimation module, which reports the estimated distance in a fixed time interval. The output from this module is converted to voice output using the auditory module.

3.3.4. Communications Interfaces

The modules associated with the project are combined in form of an integrated software which will be available as Android APK which will be ready to be installed. All the communication will take place with a help of a mic and a ear piece only.

3.4. System Features

The basic functions that the project needs to perform are Object detection, recognition and Distance Estimation. Various modules can also be implemented to further develop the project, increasing the standard of the project.

3.4.1. Object Detection and Recognition

Objects in a series of images will be located and identified i.e. for an image containing a table and chair, this feature will help in identifying both of them.
Creating accurate machine learning models capable of localizing and identifying multiple objects in a single image remains a core challenge in computer vision. For this purpose, we are incorporating TensorFlow Object Detection API, which is an open-source framework built on top of TensorFlow that makes it easy to construct, train and deploy object detection models. Various leading organizations like Google, Microsoft, etc find this codebase to be useful for computer vision needs. The API processes the image data recorded through the camera and process it, recognizing the objects with an accuracy parameter. This accuracy parameter is then compared with an optimal value, which will be set through trial run, for eliminating any false positives.

3.4.2. Distance Estimation

The identified objects are then passed through a module which is responsible for estimating the approximate distance from the source, the person. The basic idea behind the distance estimation is implementation of a module often used in automated car technologies which helps estimating the distance between the cars.The module is implemented through the image processing, roughly estimating the dimensions of the object with the help of camera properties.

3.4.3. Auditory Output

Output has to be transmitted through sound for the visually impaired people. All the processed output is passes to a module as a list which converts the textual information to auditory information which can be easily interpreted by the target audience.

3.5. Other Nonfunctional Requirements

3.5.1. Product Requirements

The product can be used in a very simple manner by just installing the android application and placing the phone properly while starting the application. Once the camera starts capturing the surrounding environment the user would get continuous information as output through the in-ear device. The user is advised to place the phone properly and also movements should be relatively spaced so that the input from the camera can be monitored properly for further analysis.

3.5.2. Performance Requirements

The response time of the product is limited by the quality of the camera, orientation of the camera with respect to the object and database stored for recognition purposes. Since, The product functions using a camera the input affects the output directly and for best performance it is necessary for the user to

orient the camera properly with respect to the object that has to identified or the person whose face has to identified. Alerts would be generated for identified objects that are too close to the person to prevent toppling and provide faster response time.

3.5.3. Safety Requirements

While the software processes or executes any particular task the user must wait for the existing task to complete by facing the camera in the required direction and waiting for any output. The user is advised to use this system in relatively safer places which are less crowded. Since the output perception is based directly on the input received from the camera the quality of camera affects the results and can give certain false positives. The user must not at any point of time press the power button toggle because it would result in shutting down the system and no inferences can be made for that particular duration until pressing the button again so as to start the system again for usage.

3.5.4. Software Quality Attributes

- ❏ The software cannot afford any unexpected failures. In order to avoid any failures occurrence the specifications have been respected and followed diligently. The following attributes have been taken into considerations:
- ❏ Maintainability : The code has been written following the standard coding practices so as to ensure that any changes and/or improvements can be made easily as and when required.
- ❏ Reliability: The system has been designed to provide maximum reliability.
- ❏ Flexibility: The Layout/architecture of the system will be flexible enough for some later requirements change or application enhancement.
- ❏ Usability: The usage for the device is really simple. Android application has to be installed and can be directly started. This has been included so as to ensure ease of use.

SOFTWARE DESIGN SPECIFICATION

4.1. Description

4.1.1. Description

There has been a significant advancement in the field of technology and science over past few decades. Various Smart device were developed which made it possible for human being to live their lives with more luxury. Now, efforts are being made to replicate the working of the brain. With the help of machine learning, various machines are being trained to adapt and perform like humans in the changing environments.

Humans can expand their knowledge to adapt the changing environment. To do that they must “learn”. Learning can be simply defined as the acquisition of knowledge or skills through study, experience, or being taught. Although learning is an easy task for most of the people, to acquire new knowledge or skills from data is too hard and complicated for machines. Moreover, the intelligence level of a machine is directly relevant to its learning capability. The study of machine learning tries to deal with this complicated task. In other words, machine learning is the branch of artificial intelligence that tries to find an answer to this question: how to make computer learn?

When we say that the machine learns, we mean that the machine is able to make predictions from examples of desired behavior or past observations and information. More formal definition of machine learning by Tom Mitchell is A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. The definition also indicates the main goal of machine learning: the design of such programs.

The ultimate goal of Machine Learning is for the program to adapt and perform like humans in changing environment. This leads to an interesting idea, Can the machines be used to replace the sensual impairments present in in humans? The answer is definitely positive since we can design a system as such. This project also focuses on development of a support system for visually impaired people which will perceive the world and supply the gathered information through auditory means. An ideal form of the project will include all the functions performed by the eyes namely object recognition, distance estimation, face recognition, character recognition, etc. Various modules can be implemented in the support of the project developing it further.

4.1.2. System Overview

The project focuses on the development of a support system for the visually impaired to help them with their daily lives. The basic functions that the project needs to perform are Object detection, recognition

and Distance Estimation. Various modules can also be implemented to further develop the project, increasing the standard of the project.

4.1.2.1. Object Detection and Recognition

Creating accurate machine learning models capable of localizing and identifying multiple objects in a single image remains a core challenge in computer vision. For this purpose, we are incorporating TensorFlow Object Detection API, which is an open-source framework built on top of TensorFlow that makes it easy to construct, train and deploy object detection models. Various leading organizations like Google, Microsoft, etc find this codebase to be useful for computer vision needs. The API processes the image data recorded through the camera and process it, recognizing the objects with an accuracy parameter. This accuracy parameter is then compared with an optimal value, which will be set through trial run, for eliminating any false positives.

4.1.2.2. Distance Estimation

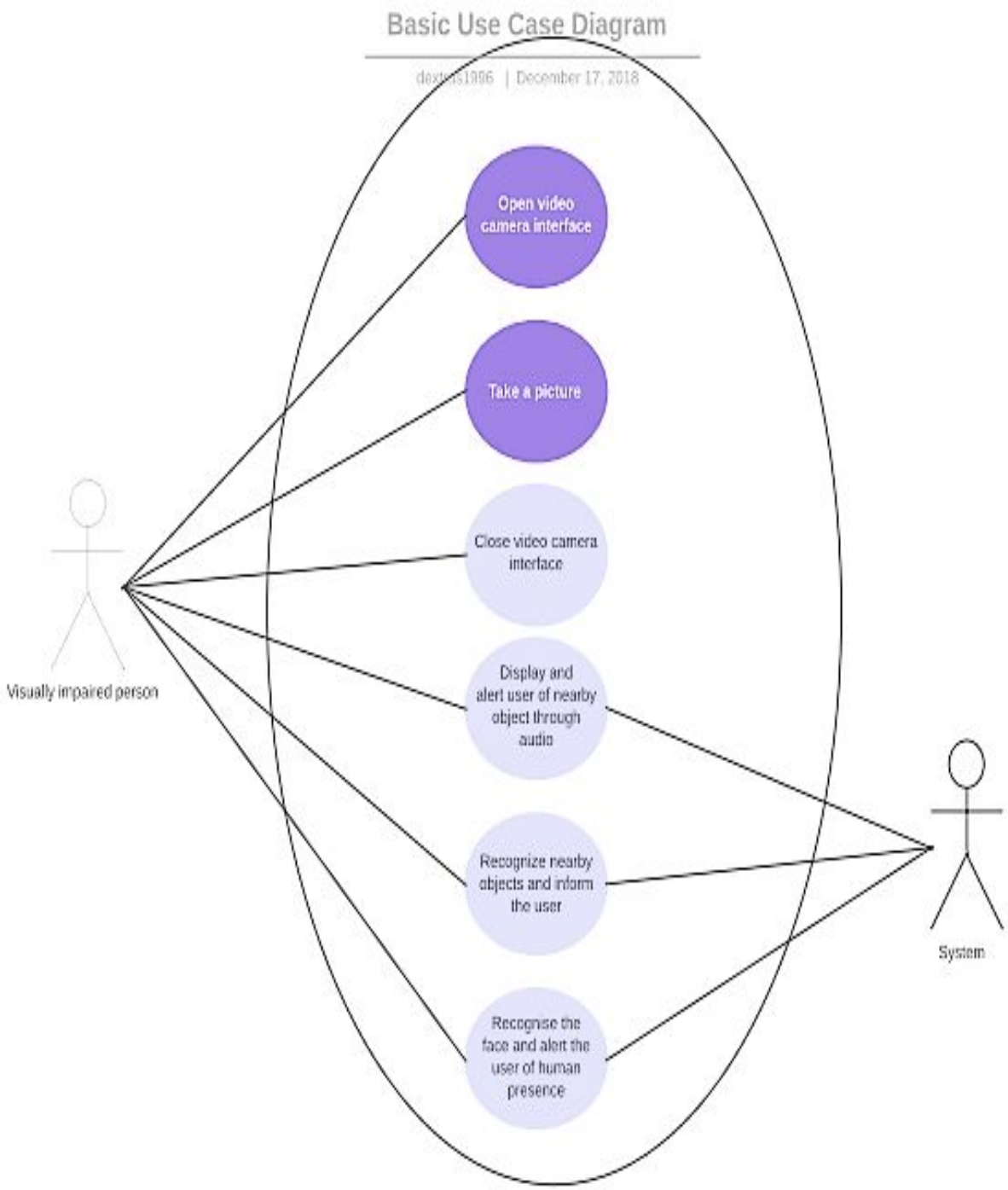
The identified objects are then passed through a module which is responsible for estimating the approximate distance from the source, the person. The basic idea behind the distance estimation is implementation of a module often used in automated car technologies which helps estimating the distance between the cars. The module is implemented through the image processing, roughly estimating the dimensions of the object with the help of camera properties.

4.1.2.3. Auditory Output

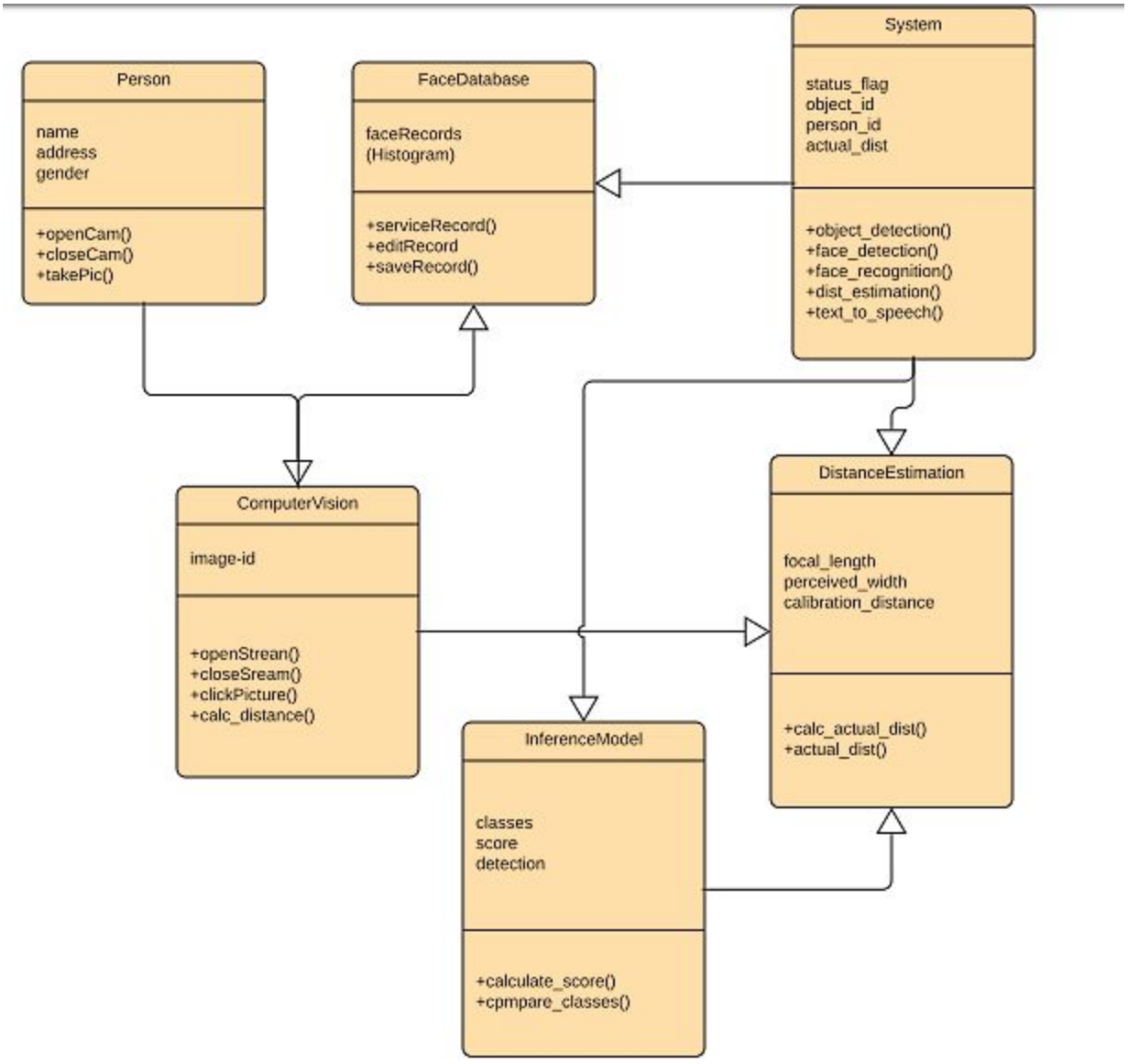
Output has to be transmitted through sound for the visually impaired people. All the processed output is passes to a module as a list which converts the textual information to auditory information which can be easily interpreted by the target audience.

4.2 Design Consideration

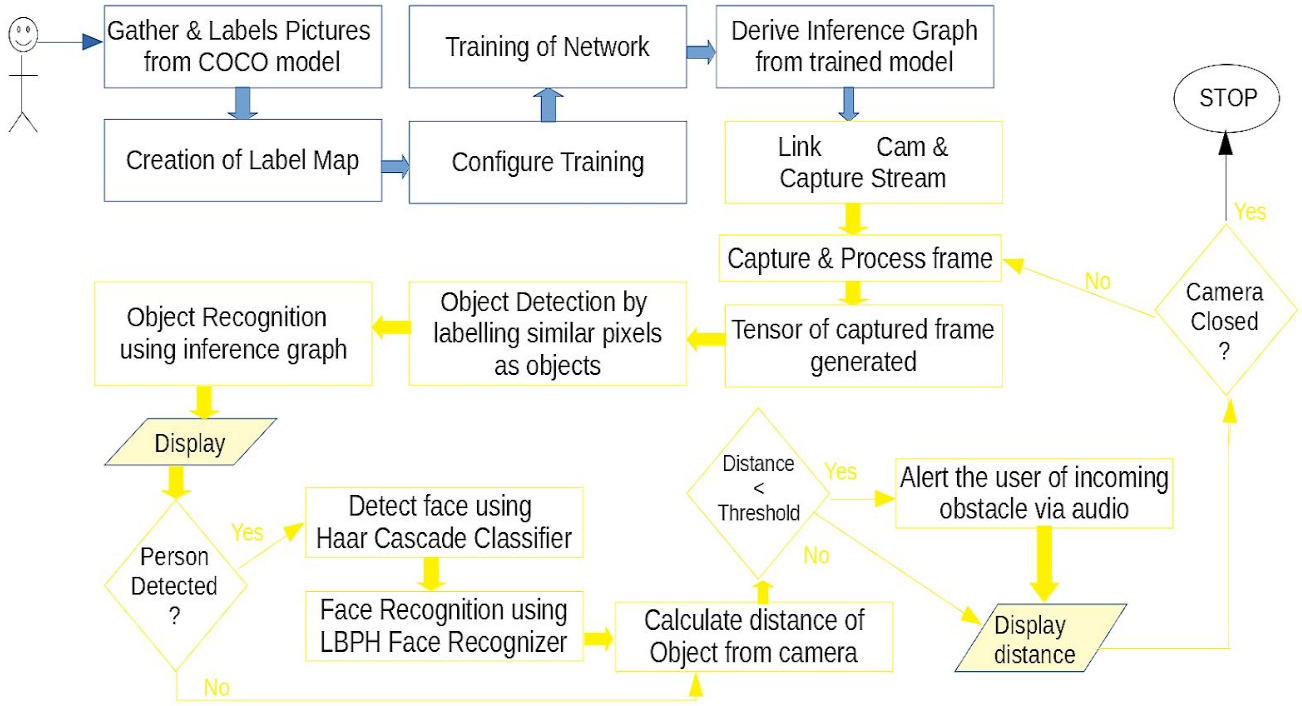
4.2.1 Use Case Diagram



4.2.2 Class Diagram



4.2.3 Flowchart



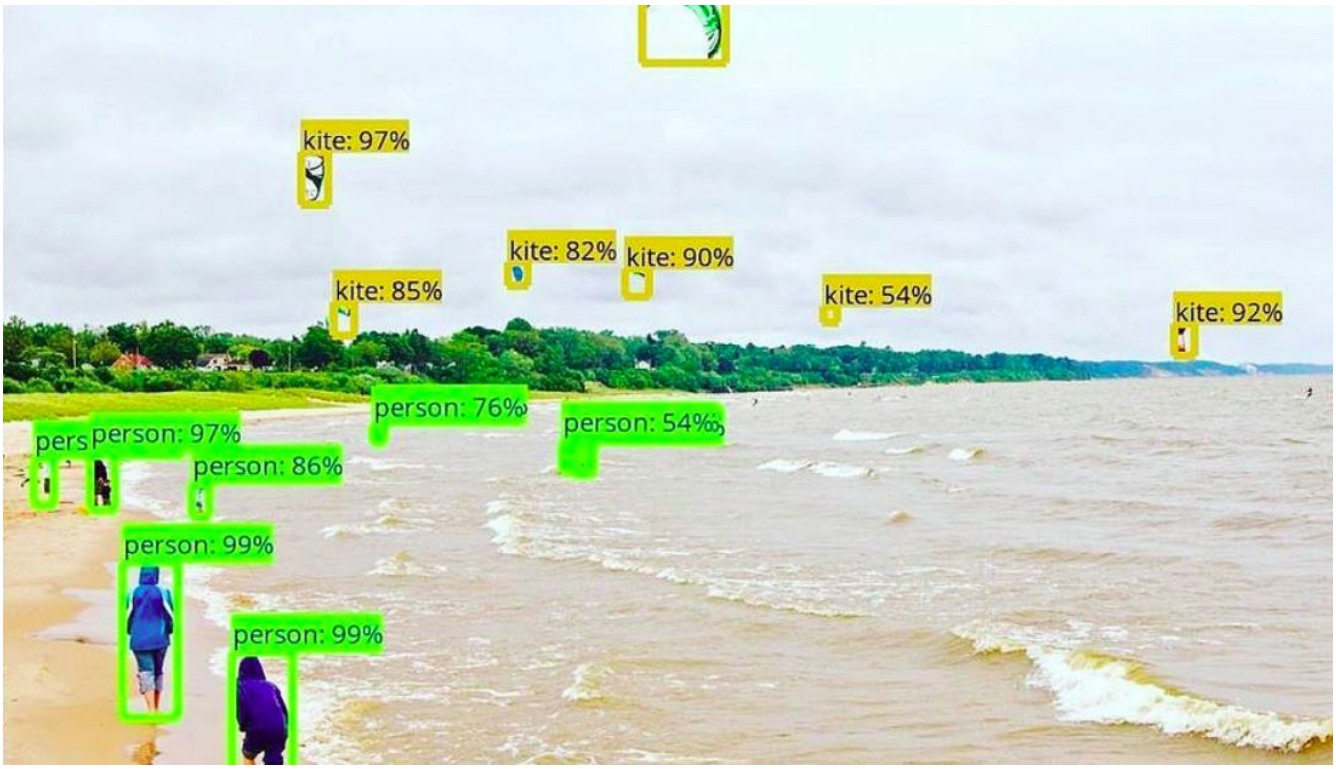
4.3. System Architecture and Strategies

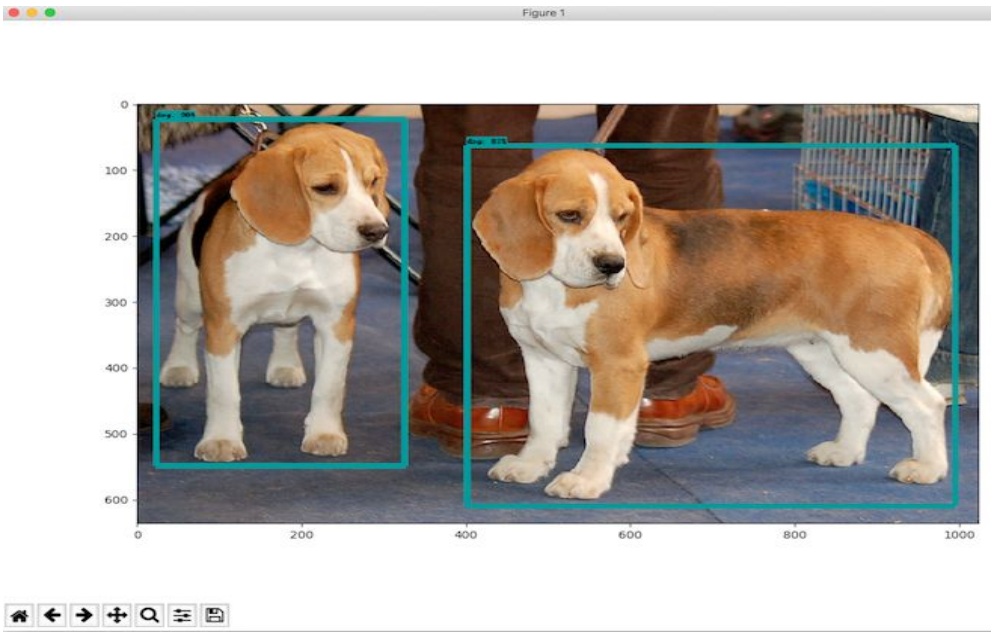
Since the project isn't sponsored, the project will be using a free datasets available on the internet. Hence, many functions of the android app will be limited. The project will be made available as an Android APK which will only need an installation and application can be directly run afterwards. Earplugs will be provided for providing the output to the target audience. The live feed will be continuously processed upon and continuous output will be generated.

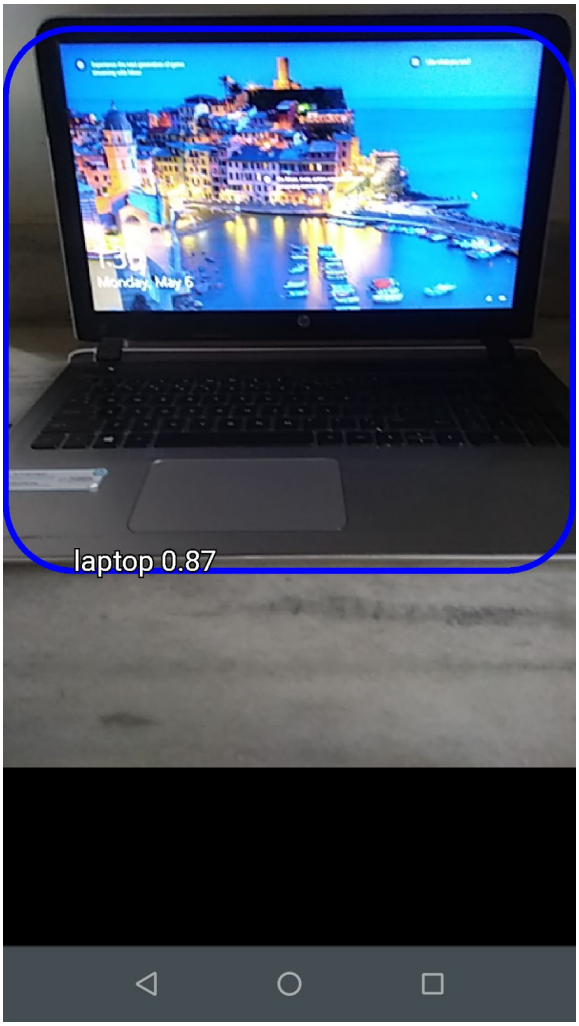
High-level Overview includes the functionalities and the flow of data. The live feed recorded through the camera is continuously passed through the object detection and recognition module which recognizes the objects present in the scene and storing them in a list. As a new object appears, it is added to the list and the object no longer in the scene is removed from the scene. The object detection module has been trained using MS COCO Dataset which contains approximately 3 lakhs images of 90 most-commonly found objects. Trainer modules are tested and best module is selected after prioritizing the speed and accuracy. The identified image part of the object is continuously passed to the distance estimation module, which reports the estimated distance in a fixed time interval. The output from this module is converted to voice output using the auditory module**.**

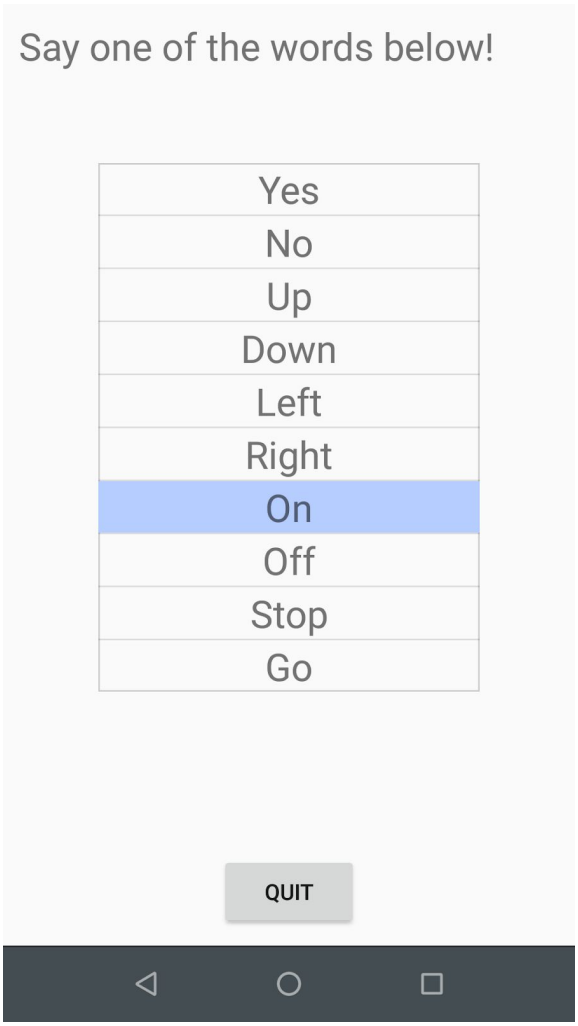
5. Output Screen & Implementation

OBJECT DETECTION AND RECOGNITION









6. Conclusion and Future work

The project aims to solve a pertinent problem concerning visually impaired people by helping them to observe their immediate environment using object recognition techniques. It guides them by informing them about the immediate surroundings helping them avoid obstacles in path by alarming through a voice system integrated with the project. The face recognition module aims at helping the blind person recognise the person in camera field so as to improve overall user experience. Since the project is an Android application (and thus present on a mobile device) it ensures maximum portability. Further in the project we would like to integrate the application

with various other Android application which would help in carrying out daily chores easily for the visually impaired person.This integration can be in terms of ordering stuff online , booking cabs etc.

7. References

1. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.228.1007&rep=rep1&type=pdf
2. WHO: Visual impairment and blindness, Fact sheet nº 282, October 2011 (2011).
3. Google TensorFlow Documents https://github.com/tensorflow/tensorflow
4. Depth Estimation using Monocular Camera, A. Joglekar, D. Joshi, IJCSIT 2011
5. A comparison of low-cost monocular vision techniques for pothole distance estimation**,** A. Das, IEEE 2017
6. **http://www.immersivelimit.com/tutorials/create-coco-annotations-from-scratch**