Visually Impaired In-Door Assistant

Kareem Emad, Nouran Khaled, Shehab Mohsen, Sherif Akram

February 6, 2020

1 Introduction

1.1 Purpose of this document

The main purpose of this Software Requirements Specification document is to outline the requirements for Visually Impaired In-Door Assistant: detect and analyze objects in the surroundings to help user navigate in new rooms freely as well as find user's own objects and give directions to his intended object. This is done with the aid of sensors such as accelerometer and pedometer. This document will provide a detailed overview of our software product's parameters and goals and explain purpose and the features of Visually Impaired In-Door Assistant and describes its interfaces,hardware, software requirements and ex-plains what the system will do. This software requirements specication (SRS) document discusses how our stakeholder, team, and audience see the product and its functionality.

1.2 Scope of this document

This Software Requirements Specification (SRS) is the requirements work product that formally specifies Visually Impaired In-Door Assistant. This targets end users like people with partial or total impairment that would use Guide Me. Users would get audible directions to their destination as well as warnings if there is too near object that they should avoid in addition to that users shall have their own dataset to save their objects which our system will use to find a targeted object if it's asked for. It will also be beneficial and helpful for researchers and developers that may work on the visually impaired assistance application.

1.3 Overview

The proposed system uses mobile camera to act as the eyes of the blind person, it sends the captured stream to the main model which is object detection using YOLO library with a pretrained dataset of house items, the user then chooses between the system's two main functionalities using speech input by translating it through natural language processing either to safely navigate the room or to look for an item he seeks. If the user chooses safe navigation the objects in the frame are detected and distance to reach them is calculated and the user is notified by speech output if the object is too close to the other and is blocking their path and where he could move to avoid that obstacle. If the users chooses finding objects, he then is prompted to say the objects name and moves his phone to capture a stream with as many frames as possible and if the object is detected in the frame the mobile vibrates meaning that the object is in that direction, and the closer he gets to the object the more intense the vibration becomes, if the object is not found after a certain time period of searching frames the user is notified by speech that the object is not found.

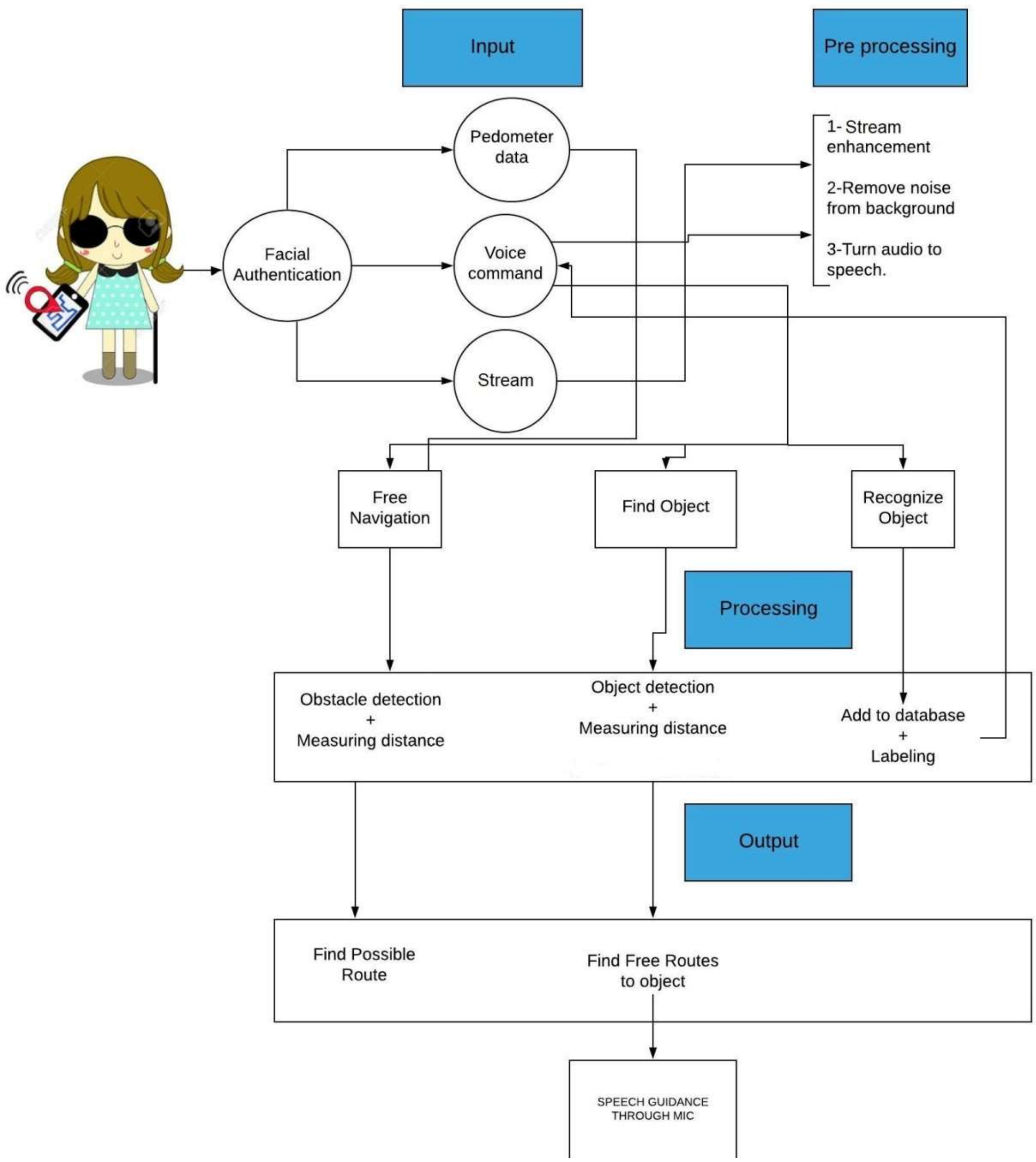

Figure 1: System Overview

1.4 Business Context

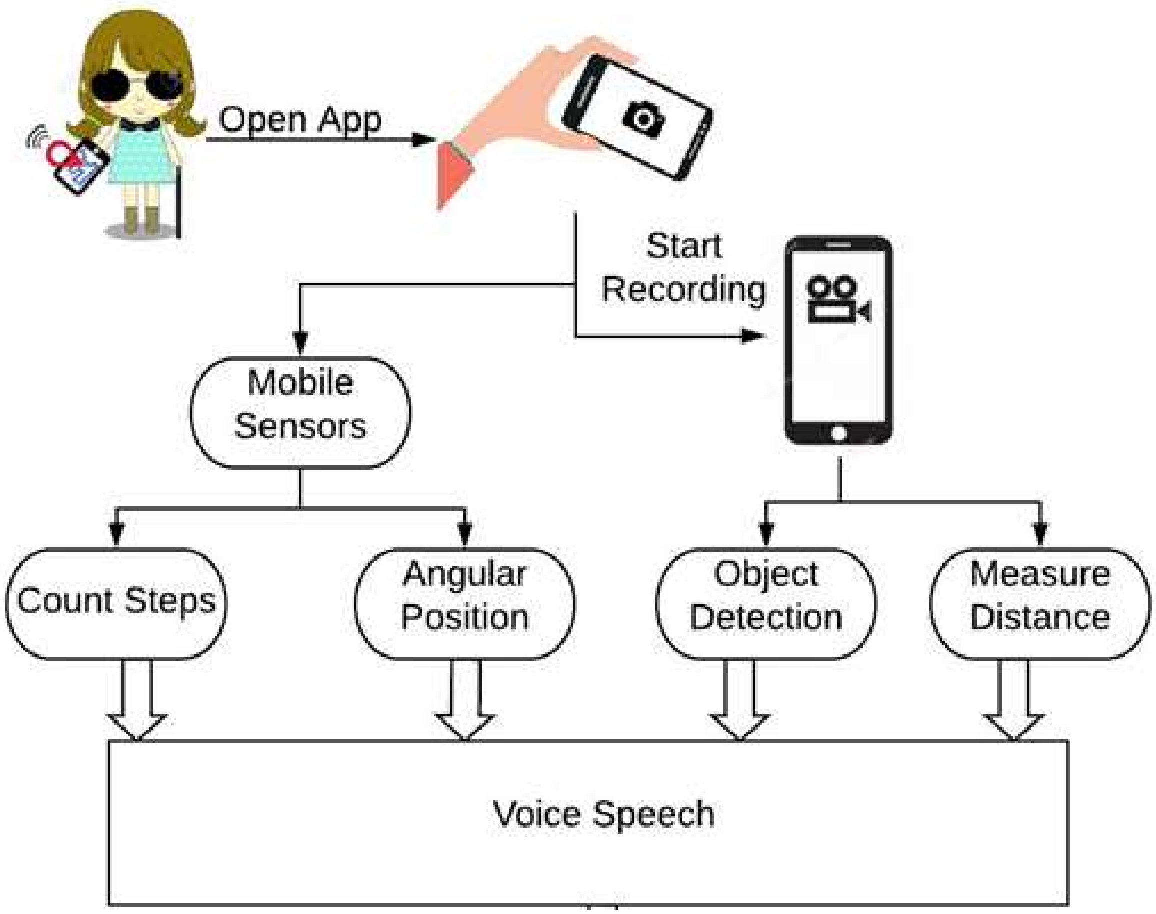

Figure 2 Business Context

Our work is motivated by both application domain and previous work, our system aims to provide an independent life for anyone with visual impairment as well as provide safe navigation which we aimed to cover in our three modules for example the most tackled problem for visually impaired people as mentioned by Mr.Salama (real life user of our system) is new places navigation. Our system solves this problem by processing the user’s captured stream of the room and finding objects in the surrounding area the system then safely navigate the user based on his chosen module whether it’s free roaming or it’s towards a targeted object.

2 General Description

2.1 Product Functions

1. Authenticate user using facial recognition
2. Speech recognition using natural language processing
3. Scanning frames captured by user for any hazards.
4. Search for required objects when asked to.
5. Measure distance between user and surrounding objects to avoid collision.
6. Alert user if there’s near objects that he’s about to hit.

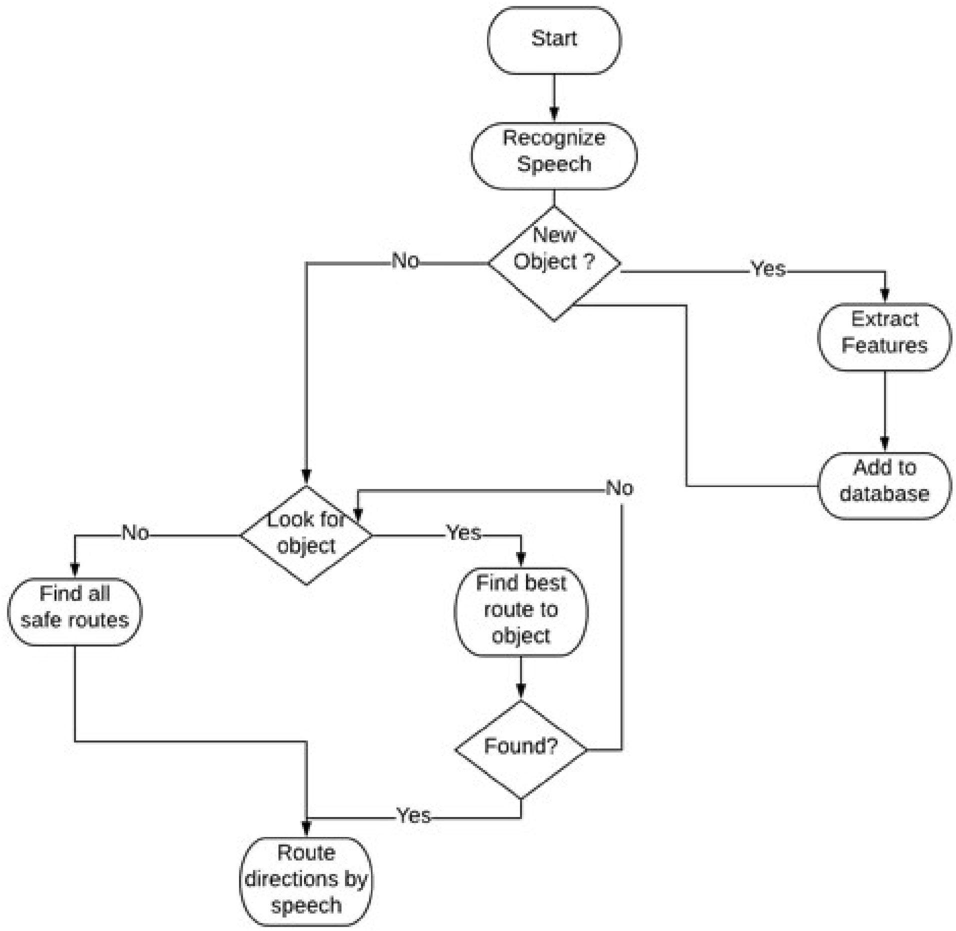

Figure 3 Flow Chart

2.2 Similar System Information

2.2.1 Intelligent Eye[2]: A Mobile Application for Assisting Blind People through four functions

1. Light detection: Which is done through the embedded light sensor in the phone to read light intensities

2. Color Detection: They obtain the image through the back camera and detect the color using OpenCV library, the RGB color of area touched by user color name is then spoken to the user using text to speech engine available in the smart phone.

3. Object recognition: Allows recognizing objects from images captured by the camera of a mobile device, they developed the activity using CNNdroid library the system then displays top ve results.

4. Banknotes recognition: Enables blind users to identify banknotes through CraftAR SDK for android.

Figure 4 Main screen of intelligent eye

2.2.2 This paper[6] is a proposed methodology for automatic identification of public surroundings. They divided their system into two objects First one is by creating their template of popular street signs(Pharmacy, special people..etc), Second step is template matching using SURF (Speeded Up Robust Features) which detects signs from captured images and relevant points from the template. Eventually they reached accuracy of 91.67% their problem was variation in color and illumination in captured image or in the template. Their future work includes increasing their accuracy as well as number of common signs and implementing same concept in the indoor navigation.

All of the previously mentioned functions are implemented to facilitate lives of visually impaired people and achieved using only mobile phone which is the same concept we are going to use in our system.

2.3 User Characteristics

Our system users are all people who have any type of visual impairment according to California optometric association [1] which are:

1. Loss of Central Vision: The loss of central vision creates a blur or blindspot, but side vision remains intact. People with this type of impairment don't need external help with our system.
2. Loss of Peripheral (Side) Vision: Central vision remains making it possible to see directly ahead. as a result they won't need external help as well.
3. Blurred Vision: Causes both near and far to appear to be out of focus. This type of impairment might need external assistance to de ne personalized items of the user.
4. Generalized Haze: Causes the sensation of a film or glare that may extend over the entire viewing eld. In this case it varies from one person to another.
5. Extreme Light Sensitivity: Washed out image and/or glare disability. this type of impairment need external assistance to capture personalized items of the user.
6. Night Blindness: Inability to see outside at night or in dimly lighted interior areas. In this case they'd need external help if they're setting up the application in insufficient light.

All previously mentioned impairments must have basic knowledge in using Android mobile devices as well as basic knowledge in understanding English navigation commands.

2.4 User Problem Statement

Detection and improvement of classification accuracy of surrounding objects as well as measuring distance between user and surrounding objects. So user needs to get a real-time information about the area he's navigating in.

2.5 User Objectives

By using Guide-Me, users with any type of impairment could now navigate freely and safely even in new places which is the most tackled issue by anyone with vision impairment as well as find their own objects which is something provided in our system using find my object module.

2.6 General Constraints

One of the main constraints of the system is the connection inside any building that the mobile could face. In addition, the position of the mobile must be not tilted in any direction to provide an accurate feedback of the detected objects as well as estimated distance.

3 Functional Requirements

3.1 Authentication

FR1	
Function	Register
Description	User can register a new account
Input	Video stream of the user’s face, Audio Stream for the Username
Source	Smartphone’s frontal camera and mic
Output	Home page with camera view
Action	Check audio for clear speech of username, Check video input for 10 frames containing the user’s face.
Pre-condition	None
Post-condition	The user’s data is sent and stored into the database for further processing
Dependencies	None

FR2	
Function	Login
Description	The user can login into his account from any Device
Input	Video stream of the user’s face
Source	Smartphone frontal camera
Output	Home page with camera view
Action	Check video input for at least one face, use facial recognition model to identify the user.
Pre-condition	The user has an account in the database
Post-condition	The Frame captured is processed and encoded on the server and classi ed accordingly using facial recognition
Dependencies	F1

FR3	
Function	Facial features extraction
Description	The system converts facial features to data.
Input	Image of a face
Source	Smartphone frontal camera
Output	Encodings
Action	Check the image for a face, Facial features are extracted and converted to data which can be used for comparisons.
Pre-condition	None
Post-condition	Encodings are saved on the server along with the the user id so they can be used for recognition later.
Dependencies	None

FR4	
Function	Facial Recognition
Description	This model allows uses Aritifcal Intelligence techniques in order to identify the user existing in the frame
Input	Image containing user’s face
Source	Smartphone frontal camera
Output	User info if the user exists in the database
Action	Check video input for at least one face, encode the facial features and compare them to existing ones in the database
Pre-condition	The user has an account with facial images available in the database
Post-condition	If the user is valid, he is allowed to access the application and his own customized objects.
Dependencies	F1,F2.f3

3.2 All Object Detetcion Model

FR5	
Function	Live stream
Description	User open a stream for detection
Input	live video from mobile camera
Source	Smartphone camera
Output	Home page
Action	None
Pre-condition	User is logged in ,camera isn’t blocked
Post-condition	Objects are detected
Dependencies	F1,F2

FR6	
Function	Object Detection
Description	Detect and track objects in the stream
Input	Live Stream
Source	None
Output	Labels detected and coordinates
Action	Check frame by frame for objects and recognize them and track their x,y coordinates.
Pre-condition	Live stream contains objects.
Post-condition	Coordinates are used to keep track of where an object resides and sent to the navigation model
Dependencies	F1,F2,F5

3.3 Customized Model

FR7	
Function	capture images
Description	users capture 10 images for the object he want to add.
Input	10 images of the object
Source	smartphone camera
Output	Homepage
Action	None
Pre-condition	The user is logged in,camera isnt block
Post-condition	the users data is sent and stored into the database for further processing.
Dependencies	F1,F2

FR8	
Function	transform images
Description	transfrom them to a lower scale so the training process is faster.
Input	Images
Source	Database
Output	images with lower scale
Action	Transform image with speci c width and Height
Pre-condition	images sent from database
Post-condition	images with lower scales used to extract features from it .
Dependencies	F1,F2,F7

FR9	
Function	Create bounding box
Description	Detect objects in the photo and draw a boundary box.
Input	Images
Source	Server
Output	Dataset with coordinates of the object
Action	Check images contain at least one object and get coordinates from it.
Pre-condition	Images is sent
Post-condition	Dataset are saved so they can be used to generate tf record.
Dependencies	F1,F2,F7,f8

FR10	
Function	Generate dataset records
Description	dataset records(Tensorow) that can be served as input data for training of the object detector.
Input	Dataset
Source	Server
Output	dataset record(Tensorow) les
Action	Check if rows and columns are valid.
Pre-condition	Images is sent
Post-condition	Generate a train.record and a test.record le which can be used to train our object detector.
Dependencies	F1,F2,F7,F8,F9

FR11	
Function	Train model
Description	Object detection model with high accuracy .
Input	dataset records les(Tensorow records format)
Source	Server
Output	Model with high accuracy
Action	Check if tfrecord le is exist.
Pre-condition	TFrecords le is sent
Post-condition	Model are saved with the user id so they can be used for customized object detection later.
Dependencies	F1,F2,F7,F8,F9,f10

FR12

Function	User Positioning
Description	The system tracks the user's location compared to the object
Input	Pedometer data.
Source	Smartphone Sensor.
Output	X,Y Coordinates
Action	Pedometer data is used to specify the user's Position.
Pre-condition	Phone's pedometer is available.
Post-condition	The system awaits the user's speech for the Needed object.
Dependencies	F1,F2

FR13	
Function	Video validation
Description	The system checks that the entire surroundings are captured in the video
Input	Video Stream, accelerometer data
Source	Smartphone camera, accelerometer sensor
Output	Alert if the video is invalid, alert on video Completion
Action	Check that the phone has rotated around itself
Pre-condition	The user has chosen a menu item, and is asked to capture his surroundings
Post-condition	The video is sent to the server for further processing and feature extraction
Dependencies	F1,F2,F5

3.6 Audio and Voice Assistance

FR14	
Function	Audio Menu
Description	The system presents menu options using speech.
Input	None
Source	Smartphone speaker
Output	Audio
Action	Convert text to audible speech
Pre-condition	The user is logged in
Post-condition	The system awaits audio command from the user.
Dependencies	F1,F2

FR15	
Function	Voice commands
Description	The system converts audio speech to text
Input	Audio
Source	Smartphone Mic
Output	Text
Action	Check audio for speech, look for desired item in menu.
Pre-condition	The system has presented the voice menu.
Post-condition	The system executes the option the user has selected.
Dependencies	F1,F2

FR16	
Function	Voice Directions
Description	The system converts the path into audible directions.
Input	Path
Source	Navigation module.
Output	Audio
Action	The path is prcoessed and turn into directions where the user can move to reach his item.
Pre-condition	Path is generated
Post-condition	None
Dependencies	F1,F2,F3,F4,F5,F6,F7,F8,F9,F10,F11,F12

4 Interface Requirements

4.1 User Interfaces

4.1.1 GUI

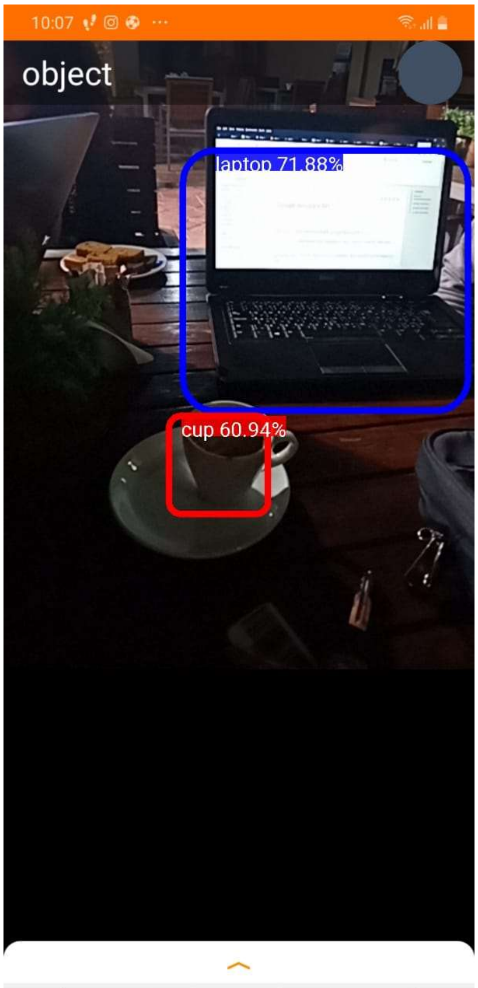

Figure 5 General Object Detection GUI

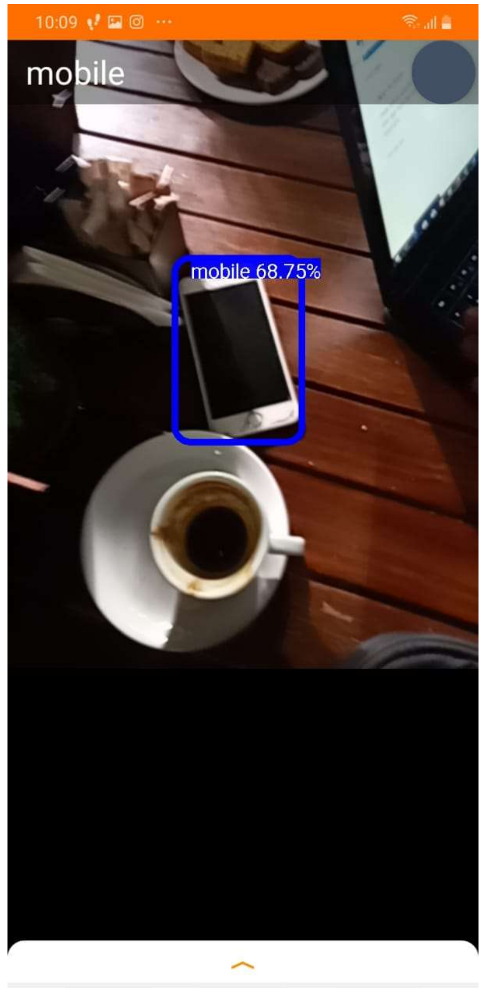

Figure 6 Specified Object Detection GUI

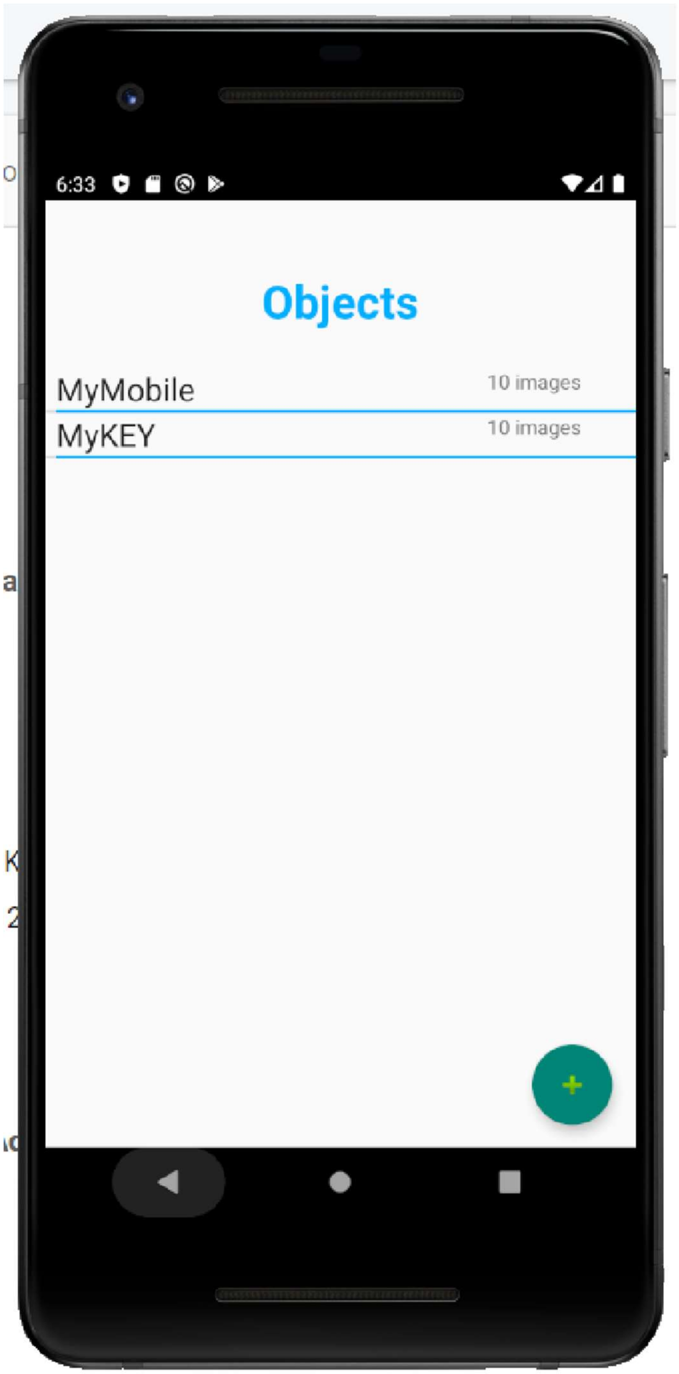

Figure 7 Adding Objects

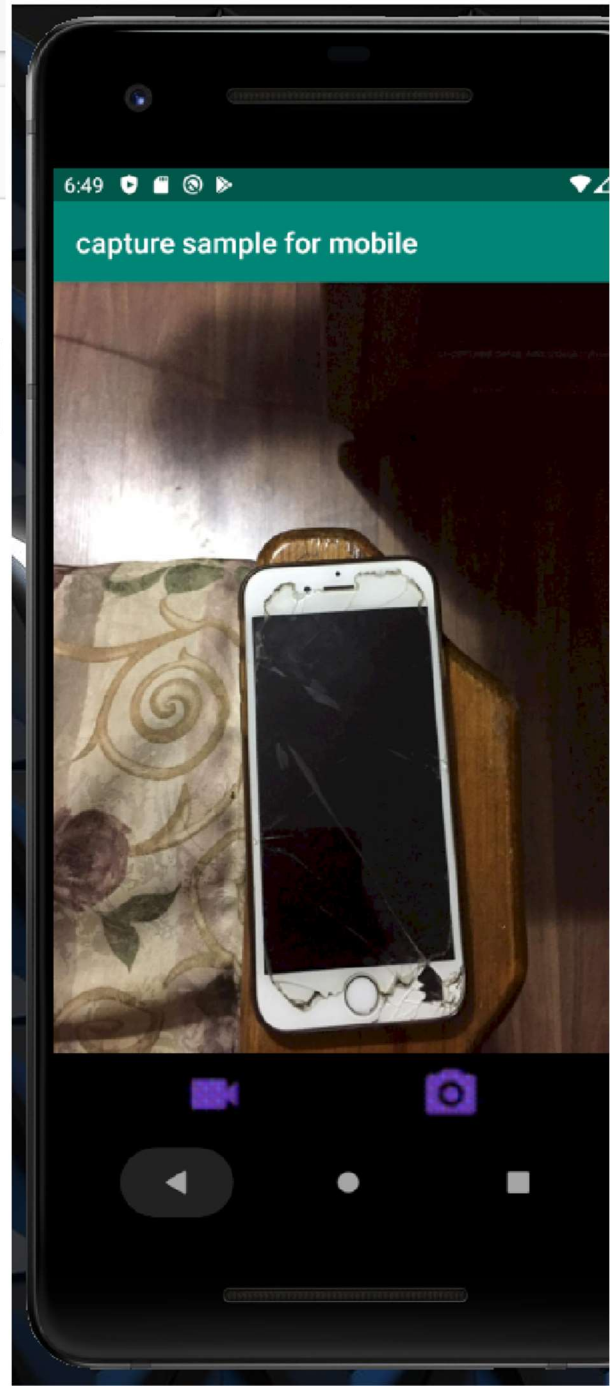

Figure 8 Capturing Object

4.1 API

1. Tensorflow Lite[3]: used for machine learning applications such as neural networks, will be used in our object detection models.
2. Luxand Facial Recognition[5]: used for processing and manipulating faces for user authentication.
3. Google Assistant[4]: used for capturing the user's speech and converting it to text, and converting text to speech.

4.2 Communications Interfaces

The communication interface is one of the most important requirements of our software as it will need a connection to the internet or a local host connection.

5 Performance Requirements

Requires real time server response with ability to process large numbers of frames, For model training, the system must be able to handle large training datasets to ensure model accuracy.

6 Design Constraints

The room where the user is located must be well lit. The size of the room that can be captured is directly proportional to the quality of the smartphone's camera.

6.1 Standards Compliance

6.2 Hardware Limitations

- Mobile must have access to a camera and is compatible with camera API version 2.
- Minimum android SDK 19.
- Surrounding room must be well lit.
- Quality of results is directly proportional to camera quality and memory.
- Phone must be connected to the internet.

7 Other non-functional attributes

7.1 Performance and Speed

The Guide-Me system must be interactive and the delays involved must be reduced. So in every action response of the Guide-Me system. Detection and classi cation must have no delays.

7.2 Reliability

The Guide-Me is reliable. It must be make sure that the system is reliable in its operations. This would be mainly focusing on the detection and classi cation. As sensors readings should be accurate and error free. When room is scanned all objects detected are being classi ed, it is very important to identify the type of the object as well as personalized objects correctly with no errors and the user should be able to trust the Guide-Me system fail rate is almost 0 percent.

7.3 Maintainability

Guide-Me system could be improved by di erent developers so its maintainability should be easy by documenting the code and the design as we implemented them using design patterns as MVC design pattern to manage the code to model, view and controller and at the the end link with the cloud, single tone for database connection and observer for noti cations .

7.4 Usability

The user should use our application and interact with it using speech, and also be able to reach their object easily by guiding him through good implemented design that guide him the easiest way to process. Proportion of functionalities or tasks implemented does not need time to be learned. Also, this system is easy to be memorized due to the small number of tasks the user will do.

8 Preliminary Object-Oriented Domain Analy-sis

8.1 Class Diagram

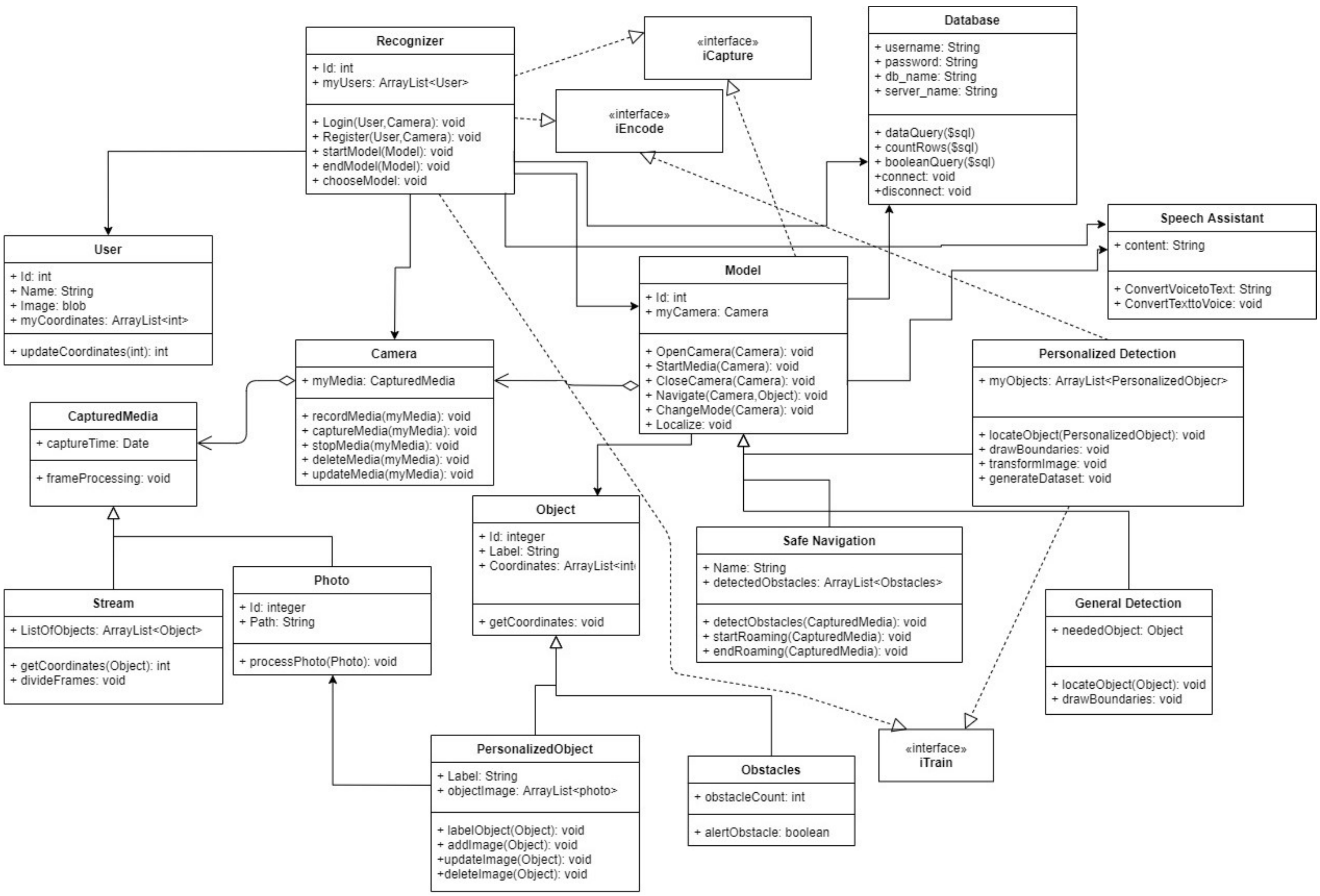

Figure 9 Class Diagram

8.2 Database Diagram

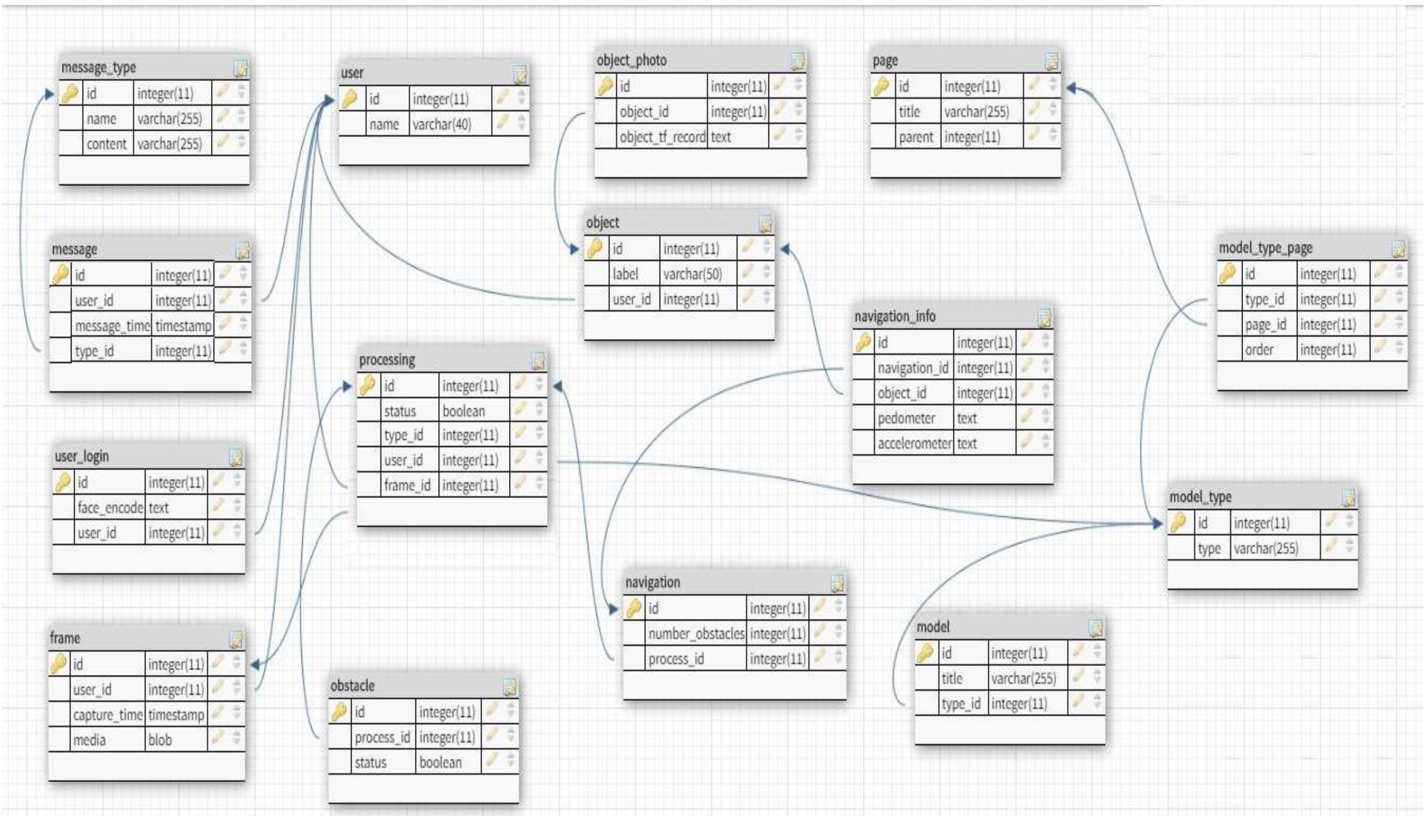

Figure 10 Database Diagram

8.3 Context Diagram

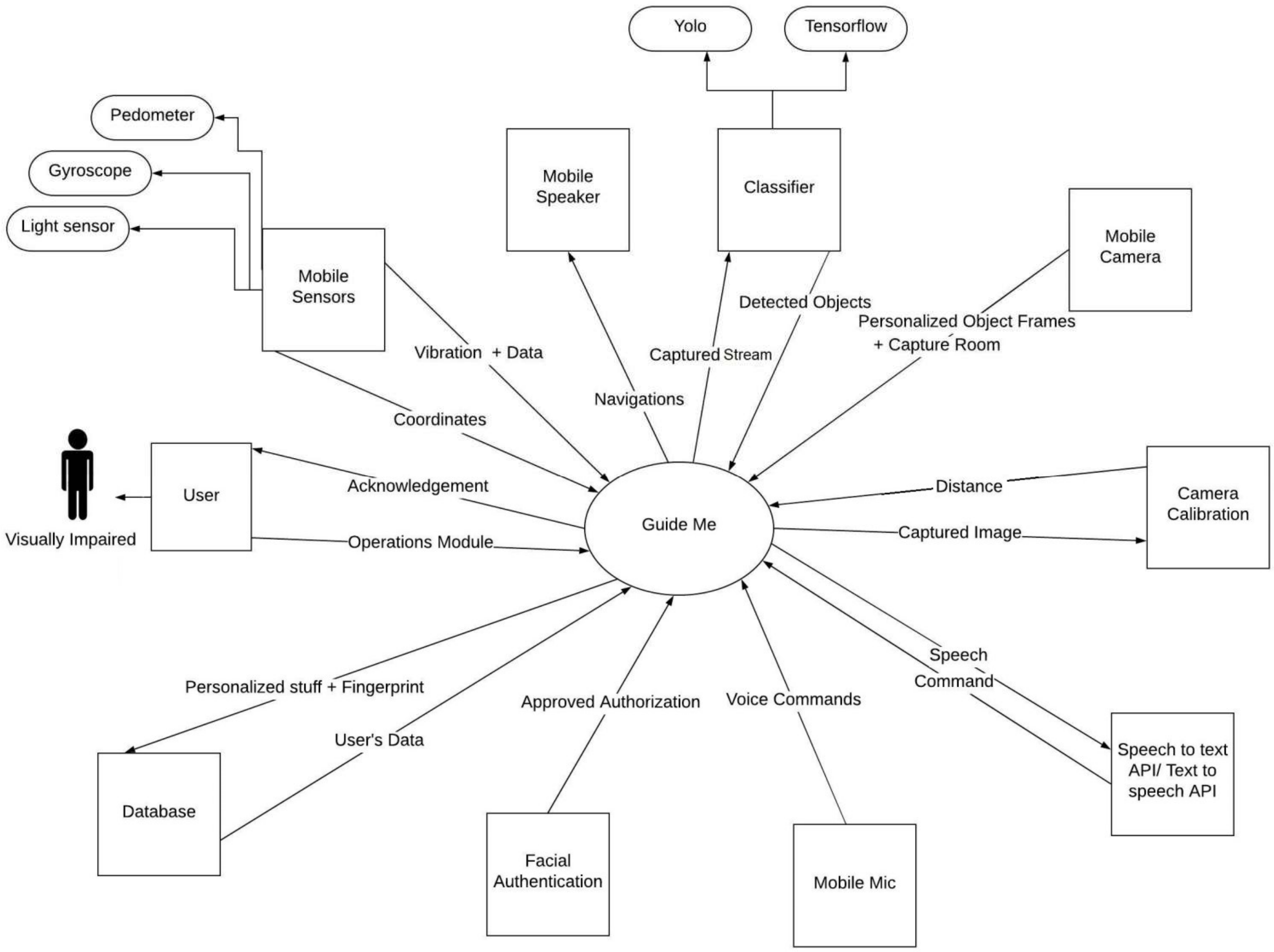

Figure 11 Context Diagram

8.4 Block Diagram

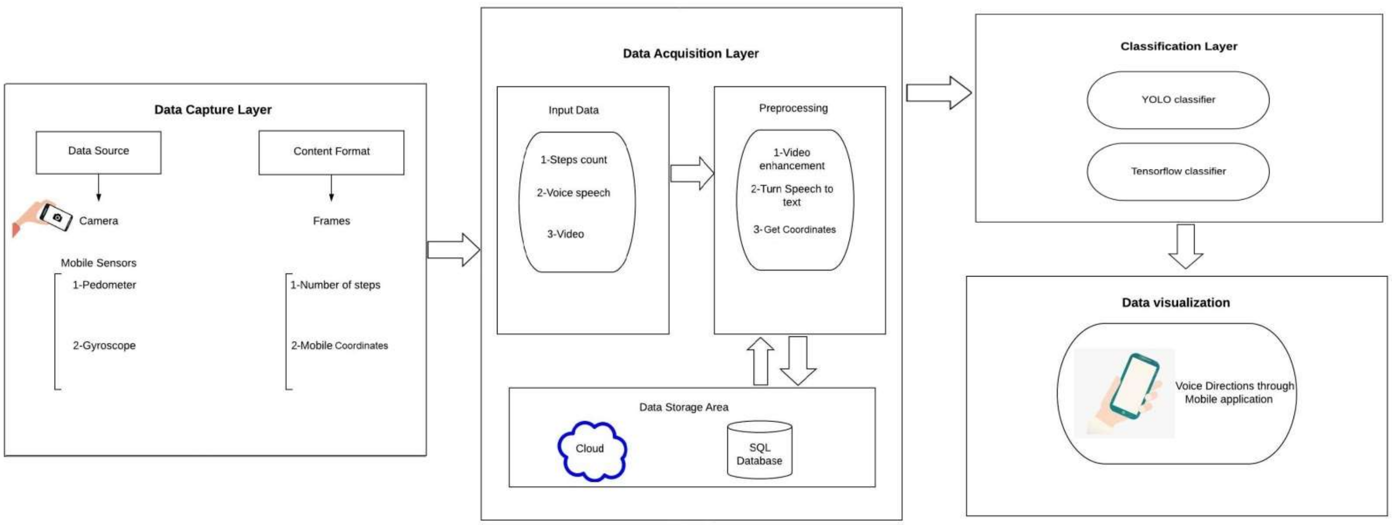

Figure 12 Block Diagram

9 Operational Scenarios

9.1 Scenarios

9.1.1 Scenario 1: Capture room process

The visually impaired person starts the application and whether his module is free roaming or nd a speci c object he captures the whole room and the application starts to send data to the system and analyze data to detect the room objects as well as calculate distance between user and these objects.

9.1.2 Scenario 2: Customized dataset handling

The visually impaired person is the one who manipulates the customized dataset which includes

1. Add object to personalized dataset.
2. Delete object from personalized dataset.
3. Edit object from personalized dataset.

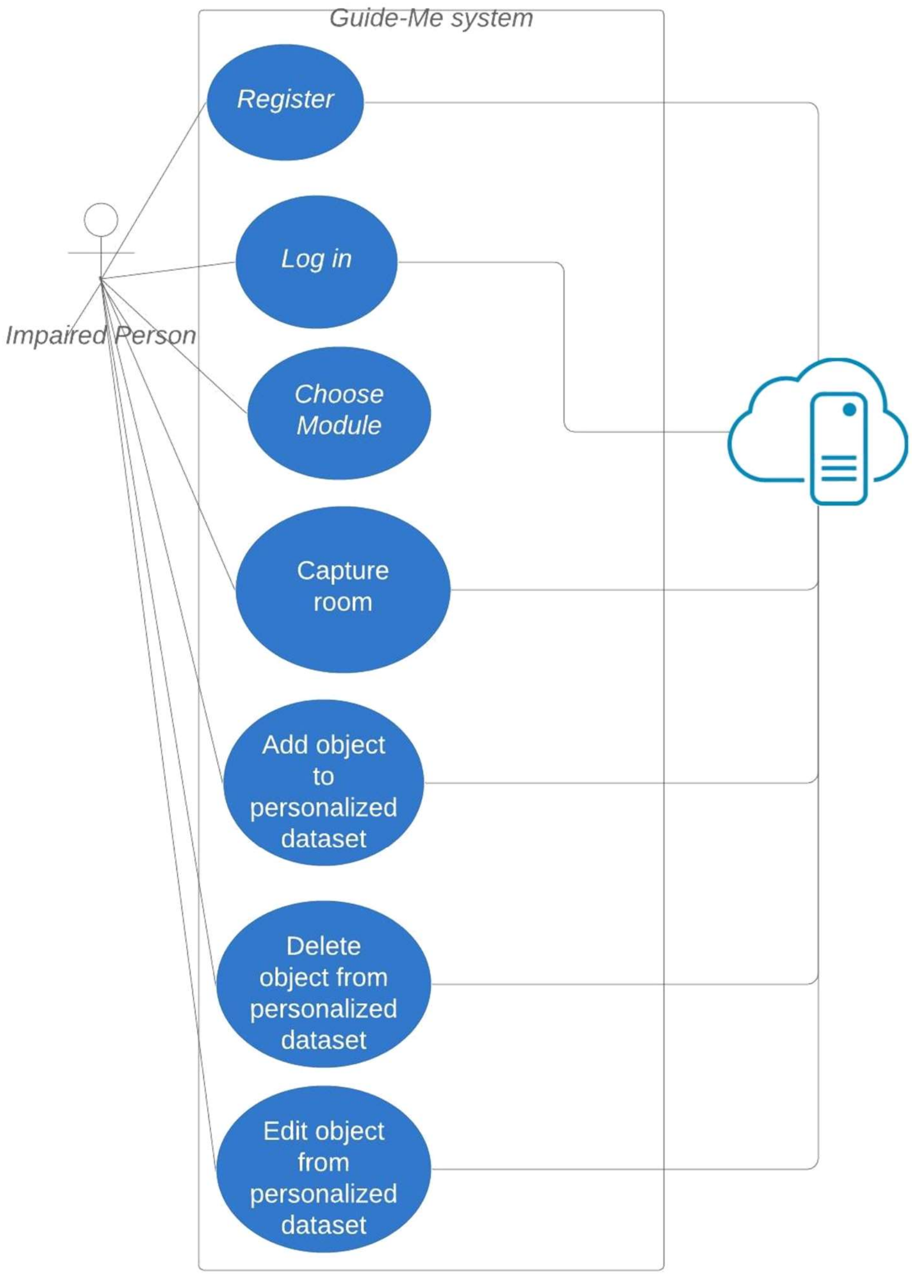

Figure 13 Use Case Diagram

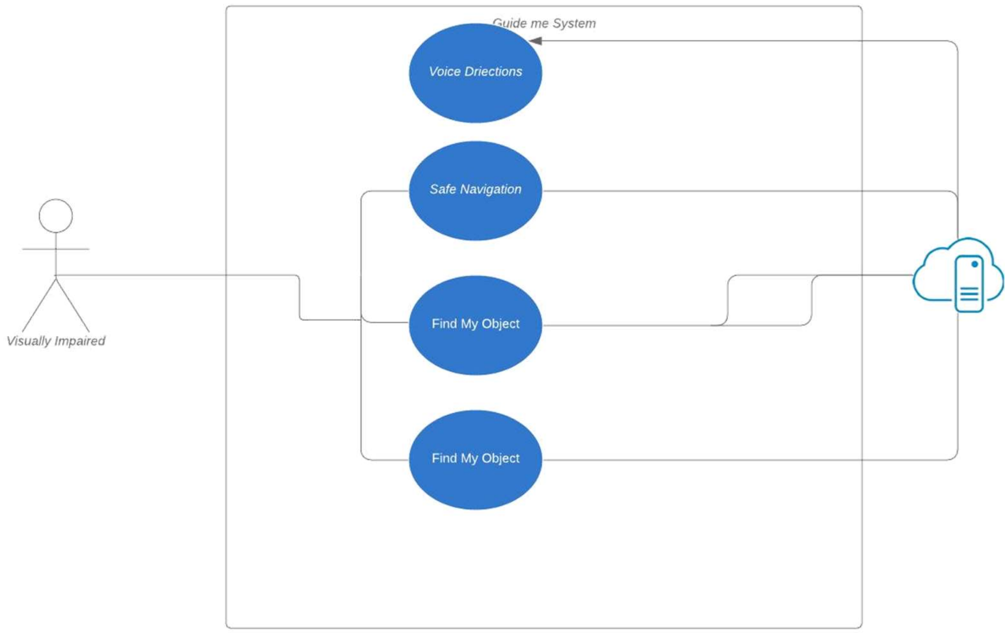

Figure 14 Use Case Diagram

10 Preliminary Schedule Adjusted

Task	Start Date	End Date
Idea discussion	25-7-2019	30-8-2019
Idea research	1-9-2019	25-9-2019
Proposal presentation	26-9-2019	10-10-2019
Implementing prototype	26-9-2019	10-10-2019
Design Application	11-10-2019	1-11-2019
Dataset collecting	11-10-2019	1-11-2019
Dataset classification	1-11-2019	11-11-2019
SRS Writing	12-11-2019	12-12-2019
SRS presentation	12-11-2019	12-12-2019
Implementing application	13-12-2019	20-1-2019
SDD writing	25-1-2019	21-2-2019
SDD presentation	25-1-2019	25-2-2019
Final prototype implementation	26-2-2019	20-3-2019
Writing paper	20-3-2019	1-4-2019
Deliver paper	2-4-2019	2-4-2019
Writing thesis	2-4-2019	20-5-2019
Final presentation	24-6-2019	24-6-2019

Figure 15 Budget Schedule

11 Preliminary Budget Adjusted

1. Luxand Facial Recognition and it’s cost is 19$ per month.

12 References

References

[1] Common types of visual impairment. [Online]. Available: https://www.coavision.org/m/pages.cfm?pageid=3625

[2] E. Y. E. K. J. E. H. M. M. Milios Awad, Tarek Mahmoud, \Intelligent eye: A mobile application for assisting blind people," 2018 9th IEEE An-nual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), p. 6, 2018.

[3] https://www.tensorflow.org/api_docs

[4] https://cloud.google.com/speech-to-text/docs/

[5] https://rapidapi.com/aboykov/api/luxand-cloud-face-recognition

[6] M. K. D. K. P. K. Dhruv Dahiya, Ashish Issac, "Computer vision technique for scene captioning to provide assistance to visually impaired,"2018 41st International Conference on Telecommunications and Signal Processing (TSP), p. 4, 2018