Smart Glasses for Visually Challenged

End-Sem Evaluation

Submitted By:

101703320 Manav Sharma

101703324 Manik Mangla

101883059 Priyanshu Tuli

101883072 Yashwika

BE Third Year, COE

CPG No. 109

Under the Mentorship of

Dr. Rajkumar Tekchandani

(Assistant Professor)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology, Patiala

Jan 2020 – Jan 2021

ABSTRACT

According to WHO (World Health Organization), the number of visually impaired people has been growing consistently for the past few decades. About 285 million people are estimated to be visually challenged, among whom 39 million are blind and 246 million have low vision, 90% of which live in low-income settings and 82% are aged 50 years and above. The irony is that even though having made so much technical advancement and reforms we do not have any cost-efficient assistive device for visually challenged people. So, a cost-effective product like ours has good opportunity as it has the potential to tap a large target market and fulfil their needs in a comprehensive manner.

We have developed a prototype of smart glasses which are comfortable, easy to use and assist visually impaired people in the following ways. It consists of ultrasonic sensors mounted in front of the frame of the glasses. These sensors will help in detecting obstacles in front of the blind person and alert him/ her about their proximity with the obstacles by informing them about their distance from it via an ear piece. These sensors will be able detect obstacles up to a distance of 3 meters from the blind person. Secondly, we will be having a face recognition system in place that will help the blind person to detect and recognize the known people in front of him/her and also provide safety from strangers. Furthermore, we will have an OCR technology in place, for the person to understand handwritten and printed text. The text would be extracted from the image captured by Pi camera, mounted on the glass itself, and would be converted to audio format using text to audio synthesis. Lastly, we will be having an image description feature that will vividly describe real world scenes to the blind person in the form of audio. For this purpose, we will use the API developed by Google. This feature will help the blind person to experience the real-life scenes by providing him/her an elaborate description of their surroundings.

DECLARATION

We hereby declare that the design principles and working prototype model of the project titled 'Smart Glasses for Visually Challenged' is an authentic record of our own work carried out in the Computer Science and Engineering Department, TIET, Patiala, under the guidance of Dr. Rajkumar Tekchandani as a part of curriculum during 6thsemester and 7th semester (2020 - 2021).

Date: 23rd December 2020

Roll No	**Name**	**Signature**
101703320	Manav Sharma	Manav Sharma
101703324	Manik Mangla	MMangla
101883059	Priyanshu Tuli	Tuli
101883072	Yashwika	Yashwika

Counter Signed By:

Faculty Mentor:
Dr. Rajkumar Tekchandani
Assistant Professor
Computer Science & Engineering Department,
TIET, Patiala

ACKNOWLEDGEMENT

We would like to express our thanks to our mentor Dr. Rajkumar Tekchandani. His mentorship and guidance have helped us immensely in completing this project. He has been an indispensable resource of technical knowledge for us in this project.

We are also thankful to Dr. Maninder Singh, Head of the Computer Science and Engineering Department, and the entire faculty and staff of the Computer Science and Engineering Department for the unconditional support and constant guidance.

We would also like to express our gratitude to our friends who have devoted their valuable time and supported us in all possible ways towards the successful completion of this project.

We are obliged to our families for their unyielding love and encouragement. They have always wanted the best for us and we admire their determination and sacrifice. Lastly, we thank all those who have contributed either directly or indirectly towards this project.

Date: 23rd December 2020

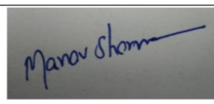
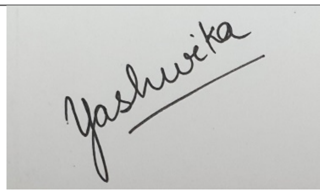
Roll No	Name	Signature
101703320	Manav Sharma	
101703324	Manik Mangla	
101883059	Priyanshu Tuli	
101883072	Yashwika	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

WHO	World Health Organization
GUI	Graphical user interface
OCR	Optical character recognition
NGO	Non-government organisation
SRS	Software Requirement Specification
API	Application Programming Interface
CV	Computer vision
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LBPH	Local Binary Pattern Histogram
GTTS	Google Text-to-Speech

INTRODUCTION

1.1 Project Overview

In this project, we are striving to make a prototype of comfortable and affordable smart glasses which will enable visually challenged persons in detecting obstacles in their way, recognizing the person in front of them along with a couple of other features to help them in their day-to-day life.

The operations performed by the glasses are as follows:

Obstacle detection

By using ultrasonic sensors (which act as a transmitter as well as a receiver), the person will be able to estimate how far the object is from them and prevent a collision. The sensor will send out a high-frequency sound pulse and then time how long it takes for the echo of the sound to reflect back. By using the formula, Distance = (Time x Speed of Sound)/2, it will calculate the distance between the person and the obstacle and warn them if they are too close.

Facial recognition

Once the CNN model has been trained for facial recognition, an image of the person is captured by a Pi camera. This image is then cropped and through feature extraction, software matches the images with its image database. After a match is found, the system tells the user that person's name whose image was captured initially and if the match is not found, it alerts the user as well. After this, the system is ready to capture the next image.

Text to speech conversion

This particular feature of the glasses will help the user in reading text. First it will confirm if the text area is correctly positioned and readable. Otherwise, it will ask the user to change the orientation of the material. After confirmation, the view is processed in real time and the image is sent to an Optical Character Recognition (OCR) software for text extraction which is subsequently forwarded to a text-to-speech synthesizer. The text is then read through the audio output port.

Image description

The image description feature of the glasses will help the user in receiving the description of the scene in front of him/ her. For this purpose, we will use a multi-model technique in which features will be extracted from the image using VGG16 model in TensorFlow, which uses Convolutional Neural Network (CNN) architecture. Captions would be generated using the Long Short-Term Memory (LSTM) model. This will be readout to the blind person in audio format.

1.1.1 Technical Terminology

- API: An Application Programming Interface (API) is a computing interface to define interactions between multiple software intermediaries. It describes different types of calls or requests that can be made, how to make them, the conventions to follow and the data formats that should be used.
- Raspberry Pi: It is a credit card sized, low cost computer that can be connected to a PC, TV or monitor and a mouse and standard keyboard can be used along with it. It can be used to program in languages like Scratch and Python and to explore and to learn computing.
- TensorFlow Object Detection: TensorFlow is an open source machine learning framework. It has a flexible and comprehensive ecosystem of libraries, tools, and community resources that can be utilized in various ML projects and thus using this, developers can easily deploy and build ML related applications.
- CNN: In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do image recognition, images classifications. Object detections and recognition faces are some of the areas where CNNs are widely used. CNN image classification takes an input image, processes it and classifies it under certain categories (eg. dog, cat, tiger, lion).
- RNN: Recurrent Neural Networks or RNN as they are called in short, are a very important variant of neural networks heavily used in Natural Language Processing. RNNs are called recurrent because they perform the same task for every element of a sequence, with the output being dependent on the previous computations.

1.1.2 Problem Statement

The main problem that is being solved by our prototype is the limited availability of assistive devices for visually challenged people specifically for those segments of the blind population who live in low-income settings as they are unable to afford such expensive assistive devices available in the market. This issue is being addressed by smart glasses we are developing which are capable of performing obstacle detection, facial recognition, text-to-speech conversion and image description and at the same time having a low manufacturing cost.

1.1.3 Goal

The goal of the project is to develop a prototype of smart glasses which not only have the capability to do multitasking, by performing various functions that help feel the user secure and give them a personalized experience, but are cost efficient also.

1.1.4 Solution

Make assistive smart glasses containing all features such as facial recognition image description, obstacle detection, conversion of text to audio and reading text to assist the visually challenged person in their daily routine and they can also experience their surroundings.

1.2 Need Analysis

Blindness and vision impairment affect around 2.2 billion people across the globe. Of those, 1 billion have a preventable vision impairment. Reduced or absent eyesight can have major and long-lasting effects on all aspects of life, including daily personal activities, interacting with the community, school and work opportunities and the ability to access public services.

Diseases like diabetes and trachoma, trauma to the eyes, age-related muscular degeneration and cataracts can contribute to reduced eyesight. Majority of people with vision impairment are over the age of 50 years. Blindness and vision loss are more prevalent among people in under developed and developing countries where accessibility to specific government services may be lacking. In these countries, the most common cause of vision impairment in children is congenital cataract.

Sight is an extraordinary gift which is vital for us to connect with the surroundings and keep us safe. Therefore, a cost-efficient method to assist the visually impaired is the need of the hour.

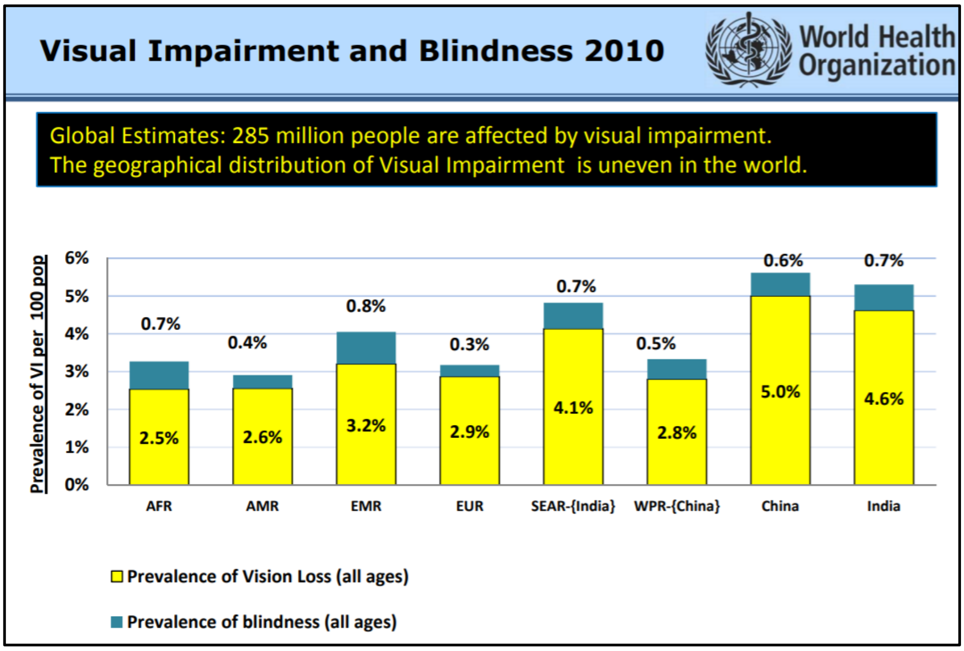


FIGURE 1: Global estimates of visual impairment: 2010, S.P. Mariotti, D. Pascolini, Br J Ophthalmol. 2012 May;96(5):614-8.

1.3 Research Gaps

From the preliminary research done on this project, it was found that extensive progress has not been made in this field and there were some gaps in the research as well.

- Image description feature is not available in any of the existing products for the visually impaired. This feature is a novelty in our product.
- All the available products only focus on solving one specific problem like reading text or assistance in walking. Our prototype contains all the necessary features required for the support in day-to-day life.
- Through the use of mode-shifter, users can use any of the available features required at any time efficiently. This facility is missing in almost all the products.
- Features like image description, facial recognition, obstacle detection and text recognition have been developed with higher efficiency, exclusively for this prototype.

1.4 Problem Definition and Scope

The problem statement for our project is that there is a need for a cost-efficient assistive device for visually challenged people. We aim to develop comfortable, easily wearable and affordable glasses that can detect obstacles in the user's way; tell them if a person standing in front is familiar or not; describe the surroundings; help them read by using the OCR technology; provide them a sense of security and assist them in their everyday tasks.

The scope of the project is to help blind people who live-in low-income settings and are unable to buy other assistive devices available in the market.

1.5 Assumptions and Constraints

TABLE 1: Assumptions and constraints

S. No.	Assumptions and Constraints
1	Our first constraint is that we can either do facial or text recognition at a time, via a push button to switch between different modes. But we will have the obstacle detection component at all times. This constraint is due to the computation complexity of facial and text recognition and the performance and processing capability of the hardware used.
2	The second assumption is that there should be proper lighting for face detection through the glasses, as we are using a comparatively less expensive Pi camera to keep the glasses affordable.
3	Our last constraint is that we can only recognize and identify a limited number of faces as we are using a micro SD card for storing the data.

1.6 Standards

- **IEEE Standard 830**

  IEEE standard 830 governs a set of rules for building the software requirement specifications. Since we are building the software requirement specifications document, hence we are incorporating this standard.

- **Industry 4.0**

  The industry 4.0 standard is also known as the fourth industrial revolution. It governs the current trend of automation and information exchange.

- **ISO/IEC 9241-11**

  Part 11 of this standard deals with the extent to which a product can be used by specified users to achieve particular goals with effectiveness, efficiency and satisfaction in the specified context of use. This standard proposes a framework that describes the usability components and the relationship between them. In this standard, the usability is considered in terms of user performance and satisfaction. According to ISO 9241-11, usability depends on context of use which is subject to contextual changes.

1.7 Approved Objectives

- Use machine learning model to recognize people

  Blind person can recognize people he/ she knows with the help of the face recognition module. For this, we have used the OpenCV library in which Haar Cascade detects faces and LBPH classifier identifies the face.

- Image Description

  This feature will describe the scene in front of the person in audio format. For this we have used a multi-model technique in which features will be extracted from the image using VGG16 model in TensorFlow, having CNN architecture, and captions would be generated using LSTM.

- Detect the obstacles using ultrasonic sensors

This feature will help the user to detect the distance from the obstacles.

- Convert text to speech and describe real-world images

  Through the use of this feature user will be able to enjoy books as well as handwritten text. It uses OCR software for reading text from the images.

1.8 Methodology Used

Analysis of the available resources and development of all the modules. These modules will be made and tested separately and integration will be carried out which will be followed by an integration testing.

We will start by gathering all the necessary equipment such as Raspberry Pi, ultrasonic sensors, earpiece, mode shifter and pi camera.

Thereafter design of the glasses will be made through which necessary algorithms for face recognition, text to speech and image description will be developed.

Requirement Analysis will follow next which will consist of designing use cases, data flow and collaboration diagrams along with the preparation of Software Requirement Analysis (S.R.S).

In the development phase accompanied by testing, multiple increments in making of the prototype will be carried out as follows:

TABLE 2: Methodology used

1	Training machine learning model for face recognition and testing
2	Developing text to speech synthesizer followed by its testing
3	Detecting the obstacles using ultrasonic sensors and testing
4	Integrating microcontroller to the glasses along with their testing

At the end, all these modules will be integrated and tested as whole. This will bring the project to its completion.

1.9 Project Outcomes and Deliverables

Following are the final outcome and deliverables of our project:

- Glasses that can detect obstacles in the user's way to avoid collision
- Intimate the user if a person in front of them is familiar or not
- Make the user aware of his/ her surroundings through audio cues
- Enable the user to understand text with the help of OCR technology, i.e., converting text to speech.

1.10 Novelty of Work

In this project we are using a machine learning algorithm for image description. We want our users to have experience of the visual world through audio. A visually challenged person cannot see an ocean or a fair. These glasses will describe any scene in front of the user. This shall be accomplished using auto ML vision API by Google.

REQUIREMENT ANALYSIS

2.1 Literature Survey

2.1.1 Theory Associated with Problem Area

Our product 'Smart Glasses for Blind' aims to cater to the needs of blind population in a comprehensive manner by providing easy to use, comfortable and cost-effective glasses. The statistics provided by the World Health Organization (WHO) depict an increasing trend in the number of visually impaired people since the past few decades. About 285 million people around the world are estimated to be visually challenged, among whom 39 million are blind and 246 million have low vision, 90% of which live in low-income settings and 82% are aged 50 years and above.

Seeing the current available products and technologies in the market, we identified a need for a cost effective and all-round product that could reach the masses in a more efficient manner. Currently there are a number of products available in the market that assist the blind people in some manner, but the drawbacks are that some are really expensive and out of the reach of the poor population while some aid only in reading or walking or a couple of tasks but is not an all-round assistive device. Our product aims to fill this gap and give the blind population a comprehensive device that they can rely on.

Our proposed smart eyewear design depends mainly on the processing unit, which is the Raspberry Pi 2. The main hardware is a Linux based ARM processor that accepts a micro SD card and thus allows us to increase the number of functions as per requirement. A Raspberry Pi camera is used for image acquisition. It is connected to the Raspberry Pi using a flex cable and is fixed on the top middle of the glasses for optimum image capturing. The Raspberry Pi has an audio port which connects to the earpiece. The Raspberry Pi GPIO port is configured to receive input from push button switches.

2.1.2 Existing Systems and Solutions

There are a few proposed models for smart glasses. The research papers are also published for them. Given below is a summary of a few research papers on it.

2.1.2.1 Low Cost Ultrasonic Glasses for Blind

This device includes a pair of glasses, an obstacle detection module fitted in the centre, a processing unit, an output device i.e. a beeping component and power supply. The obstacle detection module and output device are connected to the processing unit, which runs on power supply. This module basically consists of an ultrasonic sensor and a processing unit, consisting of a control module. The output unit consists of a buzzer to indicate the blind person about the nearness of the object to him/her. As the person comes nearer to the object the beeping intensity increases, making the person aware of how close he/she is to the object. The control unit in the object detection module controls the ultrasonic sensors and gets the information of the obstacle present in front of the user, which after being processed in the processing unit is sent as output through the buzzer.

This a portable device, easy to use, light weight, user friendly and cheap in price. It can easily guide the blind people and help them avoid obstacles.

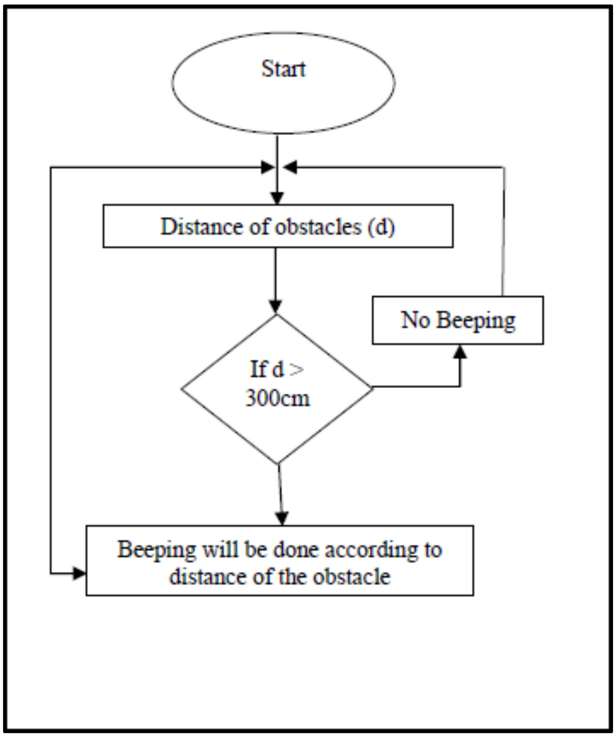


FIGURE 2: Explaining low cost ultrasonic device

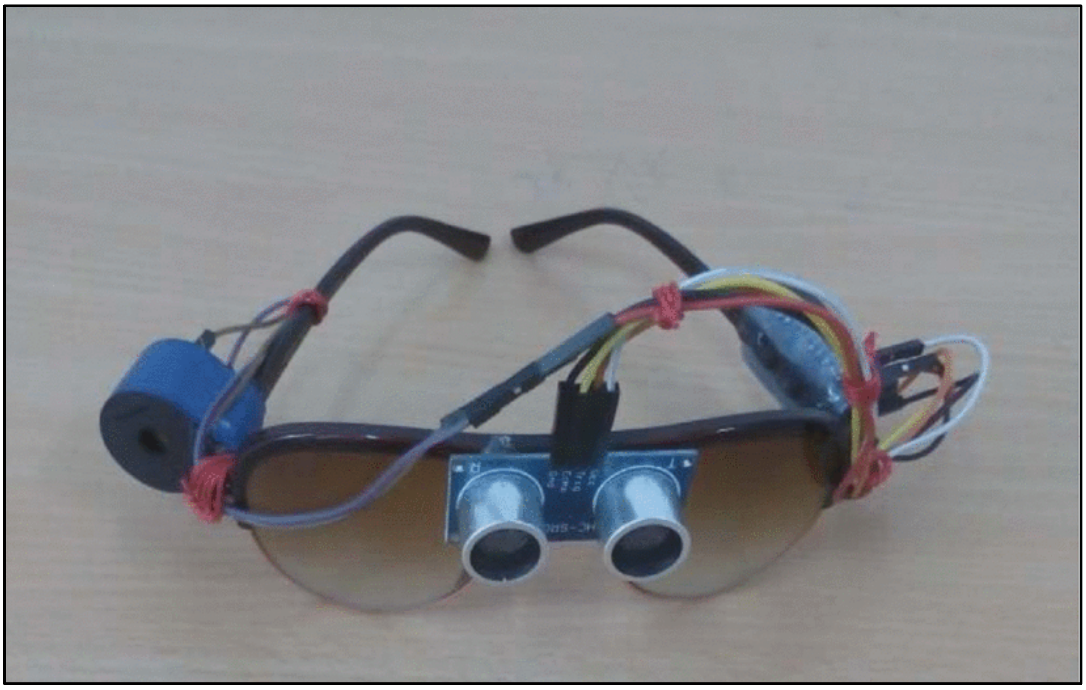


FIGURE 3: Sensor implementation

2.1.2.2 Smart Glasses Aid for Colour Blind

Epson Moverio BT200 is a prototype for helping colour blind people in distinguishing problematic colours in daily life. The prototype processes a live video stream from the mobile camera, remaps colours according to the user needs, and displays the augmented result. Colour transformation ensures high contrast between colours which are otherwise indistinguishable.

There are two processing modes to accommodate different display types. The first is the full image recolouring mode based on Jefferson's colour transformation method which increases colour contrast between problematic colours (such as red and green) and outputs a complete (recoloured) image; this approach is suitable for side-view displays, either opaque or transparent, as in the case of Google Glass. In contrast, see-through displays (like Epson Moverio) allow for a subtler second approach where high-contrast overlays are displayed only over problematic image areas. For example, a person with red-green colour vision deficiency (CVD) would directly see the world through transparent glasses, but any red objects would be superimposed with blue pixels in order to highlight them for the user and help distinguish them from green objects, other colours remain unaffected. This approach requires good visual alignment between the displayed area and the camera's field of view.

However, both the methods lead to ambiguity, as the user might not know whether any particular highlighting (such as blue) is the original colour or a result of colour transformation. To address this issue, this prototype displays processed video only for one eye; by comparing the colours seen by their eyes, the user can recognize an enhanced spectrum and distinguish otherwise problematic colours.

2.1.2.3 Smart Cane for Blind

This paper proposes a smart cane with a face recognition system to help the blind in recognizing human faces. This system detects and recognizes faces around the blind people. The result of the detection is informed to the blind person through a vibration pattern. The proposed system is equipped with a camera mounted on the glasses, a vibration motor attached to the cane and a mobile computer. The camera attached to the glasses sends the image to the mobile computer. The mobile computer extracts features from the image and then detects the face using Adaboost. It uses the modified census transform (MCT) descriptor for feature extraction. After face detection, the information regarding the detected face image is gathered. The compressed sensing with L2 Norm as a classifier is used. Cane is equipped with a Bluetooth module and receives a person's information from the mobile computer. The cane generates vibration patterns unique to each person as to inform a blind person about the identity of the detected person using the camera. Hence, the user can know the person standing in front of them.

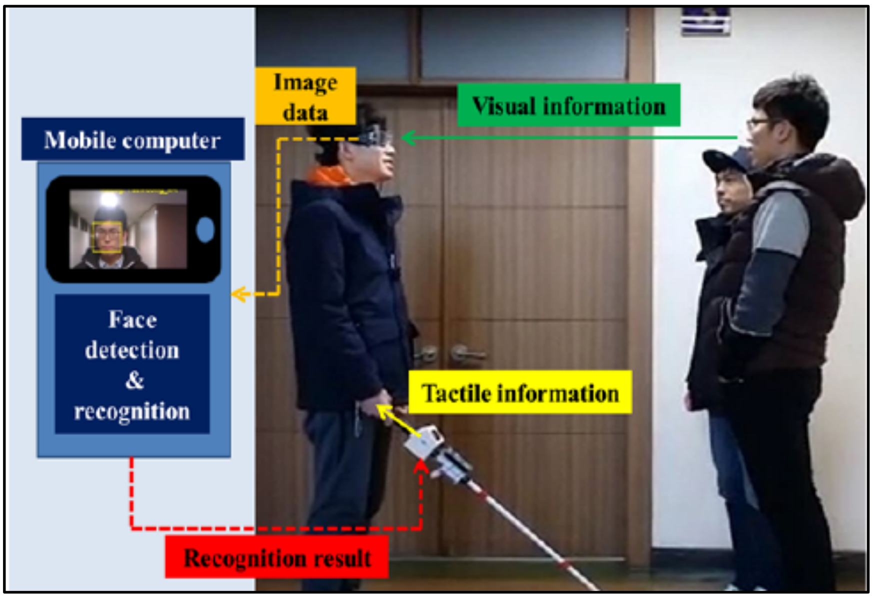


FIGURE 4: Smart cane for blind

This figure shows a situation where the blind person meets people in the corridor. A glass-type camera captures the front of view of the blind person. The mobile computer gets the front of view from the camera through WiFi link. The mobile computer detects and recognizes the person. The mobile computer sends the result to cane using Bluetooth communication. The cane informs blind person of the result of recognition.

2.1.2.4 Smart Glasses with Facial and Text Recognition

This paper presents a new design of assistive smart glasses for visually impaired people. The objective is to assist in multiple daily tasks using the advantage of wearable design format. This paper only illustrates one example application, i.e. text recognition technology that can help in reading from hardcopy materials. The building cost is kept low by using a single board computer Raspberry Pi 2 as the heart of processing and the Raspberry Pi 2 camera for image capturing. Experiment results demonstrate that the prototype is working as intended.

The general principle of operation for such glasses is by giving instructions via switches and listening to the output through an earpiece. Similarly, in this case the user starts the task mode by a push of the button. For text recognition mode, the glasses will first confirm if the text area is correctly positioned and readable. Otherwise, it will ask the user to change the orientation of the material. After confirmation, the view is processed in real time to get the image sent to an optical character recognition (OCR) software for text extraction and subsequently forwarded to a text-to-speech synthesizer. The text is then read through the audio output port. The main challenge that comes up is the image quality, text position and the orientation in the image. So, the first step would be detecting the red borders and the frame orientation and subsequently informing the user whether the image is skewed significantly or a part of it is cropped.

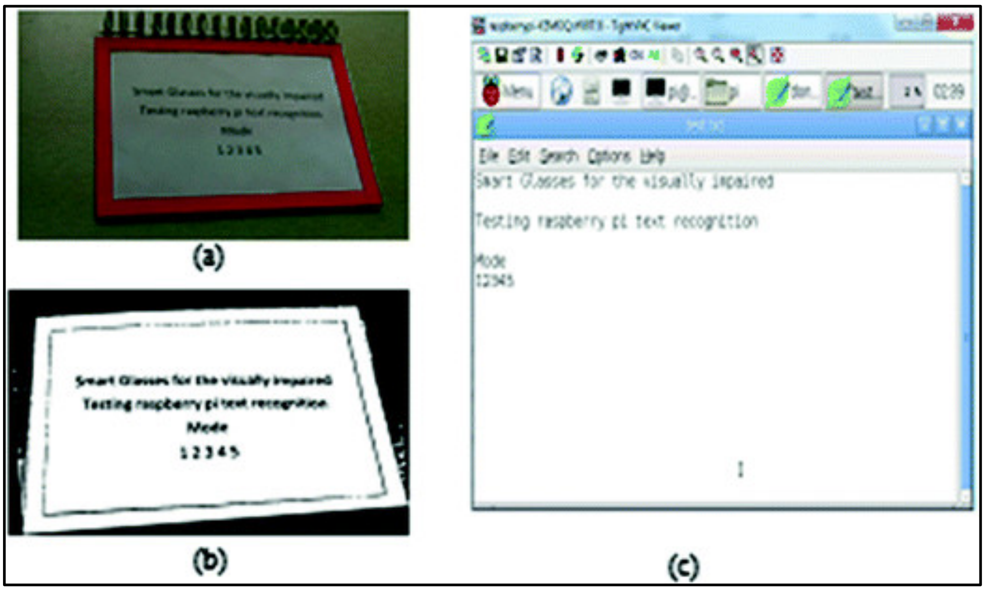


FIGURE 5: Text recognition: (a) original image, (b) image after enhancement, (c) Tesseract OCR result of reading material.

2.1.3 Research Findings for Existing Literature

TABLE 3: Research findings for existing literature

S.No.	Roll No.	Name	Paper Title	Tools/ Technology	Findings	Citations
1	101883072	Yashwika	Tactile spatial resolution in blind Braille readers	Rounded pins to display text which can be comprehended through hand movement.	Electronic device to read the text displayed on a computer monitor. The computer sends the text to the output device, where it is converted to Braille and "displayed" by raising rounded pins through a flat surface on the machine. It is priced in the range of $1000 to $3000 and supports only tactile materials.	[1]
2	101883072	Yashwika	An AudioBook Project for Blind Students at the Open Education System of Anadolu University	A device to store and listen to audio books.	Recording of books. This is available only for certain books and costs around $25 per month on subscription basis.	[2]

3	101703320	Manav Sharma	Method and apparatus for interacting with a visually displayed document on a screen reader	A speech synthesizer which speaks out the changes that occur on the computer screen.	These are software programs that allow blind or visually impaired users to read the text displayed on the computer screen with a speech synthesizer or braille display. It is the interface between the computer's operating system, its applications, and the user. The user sends commands by pressing different combinations of keys on the computer keyboard or braille display to instruct the speech synthesizer what to say and to speak automatically when changes occur on the computer screen. A command can instruct the	[3]

					synthesizer to read or spell a word, read a line or full screen of text, find a string of text on the screen, announce the location of the computer's cursor or focused item, and so on. Its shortcoming is that it is only applicable for digital content.	
4	101703324	Manik Mangla	Use wearable visual assistance system	Smart Camera, Machine Learning, Artificial Vision	It is an intuitive portable device with a smart camera mounted on the frame of any pair of eyeglasses. The device uses the power of artificial vision to assist people who are living with vision loss. It gives audio feedback to relay visual information to the user. It can read texts in real time, recognize faces, identify products, barcodes, colours and	[4]

					more. This device is not within reach of people with low income as it costs around $2,500 per piece.	
5	101703324	Manik Mangla	New Facial Recognition Smart Glasses for Visually Challenge-d Person	OLED screens, patented video processing software.	eSight uses a sophisticated high-speed camera, patented video processing software, a computer processor and the highest quality video OLED screens to project a real-time image that allows the legally blind to actually see. This device is also at a non-affordable price of $15,000 and is meant only for people with low vision and for people with total blindness.	[5]
6	101883059	Priyanshu Tuli	Low cost ultrasonic smart glasses for blind	ultrasonic sensors, buzzer	The concept of obstacle detection by SONAR sensor has been used here. As soon as the obstacle is	[6]

					detected by the sensor, its distance is sent to the Arduino. We convert the distance into centimetres from milliseconds and check whether the distance of obstacle is less than 3m, if yes then we send the output through a buzzer. The beeping frequency of the buzzer is indirectly proportional to the distance of the obstacle from humans.	
7	101883059	Priyanshu Tuli	Colorizer: Smart Glasses Aid for the Colorblind	Colour enhancement	It highlights the colour deficiency areas for the colourblind person for them to easily distinguish the problematic colours.	[7]
8	101883059	Priyanshu Tuli	Smart Cane: Face Recognition System for Blind	Mobile computer, Bluetooth, microcontr-oller, Machine Learning	Smart cane is connected via Bluetooth to the mobile computer. The mobile computer after face recognition	[8]

					sends the compressed, processed data to the microcontroller of the cane which generates different vibration patterns according to the data received.	

2.1.4 Problem Definition

The major problems identified in the above listed existing devices are twofold.

1. The cost of most of the existing solutions is not affordable by the population in the low-income group, which is where a majority segment of the blind population lies. Thus, it is critical to develop a cost-efficient solution which is affordable by the blind population below the poverty line.
2. The existing solutions cater to assist the blind person in one specific task such as reading text, walking or recognising faces. Thus, most of the devices provide a limited functionality and are not all-round devices which can be used for multiple purposes. Hence, there is a need for a cost-effective, comprehensive, assistive device to fulfill all the requirements of the blind person.

2.1.5 Survey of Tools and Technologies Used

At the heart of our smart glasses prototype is a Linux based ARM processor that accepts a micro SD card for storing data and the machine learning algorithms used. A Raspberry Pi camera is used for capturing images. It is connected to the Raspberry Pi using a flex cable, and is fixed on the top middle of the glasses for taking optimum images. The Raspberry Pi Processor has an audio output port which connects to earpiece. The Raspberry Pi GPIO port is configured to receive input from push button switches to switch between different ports.

The technologies used in our prototype are threefold.

1. Obstacle detection and Face Recognition

   The ultrasonic sensors mounted on the front frame of the glasses will be used for obstacle detection. These sensors act as both transmitter and receiver and help the blind person in intimating about their nearness to obstacles ahead. Moreover, the face recognition feature will help the blind person in recognizing familiar people in front of him/ her. In this feature, an image of the view in front is captured by the Pi camera and fed to a trained CNN model for feature extraction and image identification. The model matches the face features with its image database stored in the micro SD card. After checking in the database if a match is found, this feature tells the user that person's name and if there is no match, it alerts the user. Thereafter it is ready to capture the next image.

2. Text Recognition

   For text recognition mode, the glasses will first confirm if the text area is correctly positioned and readable. If not then it will ask the user to change the orientation of the material. After confirmation, the view is processed in real time to get the image sent to an optical character recognition (OCR) software for text extraction and subsequently forwarded to a text-to-speech synthesizer. The text is then read out to the user through the audio output port.

3. Image Description

   In this mode the surroundings of the blind person are described in an audio form to enable them to experience the scenic beauty of a place or an occasion as we people do.

2.1.6 Summary

In our prototype we have incorporated the various features into one assistive device so that the blind people have a comprehensive device to rely on for everyday tasks. Our work aims to cumulate the various features of the existing devices into one single device and at the same time keep the cost of the glasses quite cheap and affordable so that it is within reach of the low-income blind population as well. Also, our prototype has a novelty feature of image description using which the blind person will be able to experience their surroundings. This feature will enable the user to know the situations around them through audio cues.

We aim to replace the use of other assistive devices like walking stick, braille readers through our smart glasses so that the blind person only needs to rely on a single assistive device for help in day-to-day tasks.

2.2 Standards

Following are the standards are used in the project:

- **ISO- 9001**: The ISO 9001 standard describes various aspects of the quality process and defines the organizational standards along with procedures that a company should define and follow during product development.
- **IEEE 1118.1-1990**: IEEE standard for microcontroller system serial control bus defines a protocol for the interconnection of independent devices. It ensures that the devices can reliably communicate with one another.
- **IEEE/ANSI Standard 829**: IEEE standard 829 for software and system test documentation. The test processes determine whether the developed products of a given activity conform to the requirements and also whether the system and/or software satisfies its intended use and the user needs.
- **IEEE Standard 830**: IEEE standard 830 governs a set of rules for building the software requirement specifications (SRS)and also offers a tutorial/walk through for making a good SRS. Since we are building the software requirement specifications document for this project, hence we are incorporating this standard.
- **Industry 4.0**: The industry 4.0 standard is also known as the fourth industrial revolution. It governs the current trend of automation and information exchange.

2.3 Software Requirements Specification

This section describes the software requirement specifications for our prototype. It comes under the IEEE standard 830 which governs the rules for building the document.

2.3.1 Introduction

2.3.1.1 Purpose

The purpose of the SRS document is to provide a detailed explanation of the software product, its parameters and goals. This document describes the project's target audience, its user interface and hardware and software requirements. It defines how the project's client, team and audience see the product and its functionality. This document is an in-depth description of the working prototype and the circumstances in which the product can possibly fail. This will be the first release of this document.

2.3.1.2 Intended Audience and Reading Suggestions

Following are a few points regarding the intended audience and reading suggestions for this document:

- This document is intended for users and testers. The document represents a detailed description of the idea being implemented in this project.
- Developers who are able to review this project's capabilities, scope and limitations and can easily identify the areas and scope of improvement.
- Project testers can use this SRS document as a basis for their testing strategy to find bugs and glitches. This way testing becomes more structured, organized and easy to conduct.
- Social welfare organizations such as NGOs or even other companies like Lenskart, which we will be partnering with for the production of smart glasses can go through this document for a better understanding of the product and detailed explanation about the deliverables.
- Moreover, this document will also be available for the government officials and authorities to read since we are planning to associate with them as well for providing a subsidy on the cost price of the smart glasses.

2.3.1.3 Project Scope

- The scope of this project is to provide the blind people with comprehensive, comfortable, easy to use, affordable smart glasses to fulfill their requirements.
- This project aims to cater to the need of the blind population in an effective manner by providing them with a user-friendly, affordable product.

- Through this project we aim to address the needs of the visually impaired people and how we as a community can develop a solution which is cheap yet effective to help them lead a better life, not being dependent for assistance.
- We are also planning to tie up with NGOs, blind schools and other social welfare organizations to provide a wider reach for our product.

2.3.2 Overall Description

In this section we will discuss about the overall description of the entire product under following sub headings:

2.3.2.1 Product Perspective

This product aims to cater to the need of the blind population in a cost-efficient manner by providing a user-friendly, comfortable assistive device. This is a complete product for the blind population as it assists them in a number of everyday tasks they face difficulty in performing. Moreover, companies like Lenskart which we are planning to tie up with for production, will have the opportunity to increase their customer base and enter a new segment of eyeglasses backed by the trust of their brand name. Also, NGOs will help us in the social cause and increase our customer reach.

2.3.2.2 Product Features

Our product has namely four features.

1. Obstacle Detection

   By using ultrasonic sensors, acting as a transmitter and receiver, the person will get to know how far the object is from them and prevent collision. The sensor will send out a high-frequency sound pulse and then calculates the time elapsed between transmission and reception of the echo. By using the formula, distance = (time x speed of sound)/2, the distance between the person and the obstacle is calculated and subsequently the user is alerted if they are too close.

2. Facial recognition

   After training the model, an image of the person is captured by a Pi camera, it is cropped and through feature extraction, the model matches the images with its image database. After a match is found, the system tells the user that person's name and if the match is not found, it alerts the user as well. After this, the system is ready to capture the next image.

3. Text Recognition

   This feature uses the OCR technology to convert text to speech which is then transmitted using the ear piece. Through this feature the blind person can read both handwritten and printed text. The only thing that needs to be ensured before starting text recognition is correct alignment and readability of the text area so that text is neither cropped nor distorted. This is critical for text recognition mode to correctly convert text to speech.

4. Image Description

   Image Description is the novelty feature of our product. This unique feature will enable the blind person to experience a scenic beauty or an occasion. Such a feature is incorporated to give the blind person close to real experience of the surroundings.

2.3.3 External Interface Requirements

Following list presents the external interface requirements:

- The product requires a Raspberry Pi 2 microcontroller, which is the heart of all the processing in our prototype.
- A micro SD card is also needed for storage purposes. This micro SD card will store the image database of the known people and allow for multitasking as well.
- An earpiece, ultrasonic sensors, flex cables, switch, Pi camera and power bank for a constant supply to the microcontroller are among the other hardware requirements for the product.

2.3.3.1 User Interface

The blind person interacts with the smart glasses by switching between various modes for using the various functions provided. Also, the user can adjust the volume of the audio output coming through the earpiece as per their requirement.

2.3.3.2 Hardware Interface

The hardware interface mainly builds upon the Raspberry Pi 2 microcontroller. It has a Pi camera attached to it for image acquisition; ultrasonic sensors for object detection and the earpiece for audio output. The micro SD card is also inserted into the microcontroller for storage purposes. The power bank is used to provide a constant voltage supply to the microcontroller.

2.3.3.3 Software Interface

The software interface includes the Google API, pretrained machine learning models and frameworks such as TensorFlow, Keras and OpenCV using which these models are built.

2.3.4 Other Non-functional Requirements

2.3.4.1 Performance Requirements

A. Optimum performance time of face recognition, text recognition and image description algorithms.
B. Proper alignment and readability of the text area for text recognition.
C. Image capturing speed and clarity of Pi camera in all kinds of weather conditions and external lighting.
D. Time elapsed in converting text to speech should be minimum.
E. The images captured from the camera should be processed quickly and efficiently for extracting the facial features using the face recognition mode.
F. Computation power of Raspberry pi microcontroller should be appropriate enough to process various functions side by side and should be able to adjust with frequent switching between modes.

2.3.4.2 Safety Requirements

- High scale testing should be done before deploying the machine learning models.
- Hardware components should have adequate power supply and voltage.
- The smart glasses should be light and comfortable so as to not harm the users face due to its heaviness.

- The power bank attached for voltage supply to the microcontroller should be of good quality so as to not catch fire or provide electric shock in any kind of weather condition.

2.3.4.3 Security Requirements

- The user will be authenticated by logging into the system. Only after logging in the user will be able to use the features of the smart glasses.

2.4 Cost Analysis

2.4.1 Software Requirement

TABLE 4: Software requirement

S. No.	Product	Cost
1	Optical Character Recognition (OCR)	₹0
2	TensorFlow	₹0
3	Keras	₹0
4	OpenCV	₹0
5	Haar Cascade	₹0
6	Google API GTTS	₹0
Total Software Cost		**₹0**

2.4.2 Hardware Requirement

TABLE 5: Hardware requirement

S. No.	Product	Cost per piece	Total Cost
1	Ultrasonic Sensor pair (2 pieces)	₹165	₹330
2	Raspberry Pi (1 piece)	₹2500	₹2500
3	Pi camera (1 piece)	₹2000	₹2000
4	SD card (1 piece)	₹	₹1300
5	Power Bank (1 piece)	₹500	₹500
6	Connecting wires and resistors (10 pieces)	₹15	₹150
7	Earpiece (2 piece)	₹150	₹300
Total Hardware Cost			**₹7080**

Net Cost = Total Hardware Cost + Total Software Cost

= ₹7080 + ₹0.0

= ₹ 7080

2.5 Risk Analysis

- The major risk while using the prototype is switching between different modes very quickly which might cause the system to hang due to limited processing capability of the Raspberry Pi 2 microcontroller.
- A relatively low priority risk is running out of memory space. Although, the micro SD card provides sufficient storage capacity to store machine learning models and images of familiar faces.
- One very low priority risk is that hardware devices in the prototype could burn out due to short-circuit or any other technical glitch. Also, the camera lens might have scratches over a period of time, leading to low quality images.
- One important risk is that smart glasses might not function properly during rain or other extreme weather conditions, giving faulty results and warnings.

METHODOLOGY ADOPTED

3.1 Investigative Techniques

In our project titled 'Smart Glasses for Visually Challenged', experimental investigative techniques have been used keeping in mind the idea of contributing and adding value to the society through our engineering skills. Investigative techniques involve ideating, constructing the required procedure for execution of the idea and finally coming up with the hypothesis for the project. Our idea is primarily to modify the existing smart glasses by adding more features and making a comprehensive assistive device for the blind while keeping the cost low and affordable. We have also incorporated the image description feature in our prototype which enables the blind person to experience their surroundings. The procedure followed by us is divided into the following 3 steps:

- Face recognition and obstacle detection for recognizing familiar people and measuring the distance from obstacles respectively.
- Text to speech for converting handwritten and printed text to audio format.

- Image description for the blind person to experience their surroundings.

Independent variables are those variables which are not dependent on other factors for a change in their value and modify the value of dependent variables.
Following are the independent variables in our prototype:

- Ultrasonic sensor for obstacle detection
- Pi camera for capturing images
- Mode shifter for shifting modes

Dependent variables are those variables whose value depends on other factors i.e. independent variables determine the value of dependent variables. These variables act as the motivation for this prototype. Following are the dependent variables in our project:

- Text to speech depends on the orientation of text in the image captured by the Pi camera
- Face recognition depends on the image clarity of the image captured by the Pi camera
- Every feature depends on the mode selected by the mode shifter

Constant variables are variables whose value does not change during the experiment and remains constant.

Following are the constants in our project:

- Constant power supply to the microcontroller using power bank
- Pi camera position on the smart glasses
- Ultrasonic sensor position on the smart glasses

Hypothesis

Using the input image taken from the pi camera and the mode selected from mode shifter our project follows the hypothesis given below.

- Once a picture is captured, mode shifter will switch to perform the next feature automatically
- If the first mode is selected then object detection will take place using the Ultrasonic

sensors mounted on the front of the smart glasses.

- If the second mode is chosen using the mode shifter then face recognition will take place for recognizing familiar people in front of the blind person.
- On selecting the third mode, the text to speech feature will function converting handwritten and printed text to audio format.
- On choosing the fourth mode, the image description feature will enable the blind person to experience their surroundings through audio cues.

Discussion and Conclusion

The object detection feature in our project will be able to detect objects in front of the user up to a distance of 3 meters using ultrasonic sensors and thereafter intimate the user about their proximity to the obstacle via beeping intensity of the buzzer. We Also have the text to speech synthesizer in our prototype which uses the Google API GTTS to convert text present in image captured to audio cues.

The machine learning model used for the image description feature will explain the surroundings to the blind person. In this feature image captured by the Pi camera will be described in audio format to the user. It uses a multi-model technique in which features are extracted from the image iteratively using the pretrained VGG16 model in TensorFlow having CNN architecture and thereafter captions are generated for each corresponding iteration using LSTM. For the face recognition feature, we have used OpenCV library in which Haar Cascade is used to detect faces from the images and LBPH classifier identifies the face among the ones stored in the database.

Hence, on the basis of the above discussion we have found that our proposed hypothesis, once a picture is captured, the mode shifter will switch to perform the next feature automatically, holds true.

3.2 Proposed Solution

Our idea is to develop smart glasses which are comfortable, easy to use, affordable and assist the blind people in a number of ways as follow:

- Using ultrasonic sensors, mounted on the front frame of the smart glasses, for obstacle detection. These sensors will detect obstacles in front of the blind person and intimate them about their proximity with the obstacles ahead through audio cues transmitted via the ear piece. These sensors will only be able detect obstacles up to a distance of 3 metres from the blind person.

- Secondly, we have the face recognition feature that will enable the blind person to recognize the familiar people in front of them and also provide safety from strangers.
- Furthermore, we also have the OCR technology in our prototype which allows the user to understand printed as well as handwritten text. The text will be extracted from the image captured by Pi camera which is mounted on the smart glasses itself and thereafter it will be converted to audio form using text to audio synthesizer.
- Lastly, we have the image description feature which will describe the everyday scenarios to the blind person in the form of audio cues. For this purpose, we will use the API provided by Google. This feature will help the blind person to have an experience of the real-life situations by having a description of various objects and movements around them as audio cues.

3.3 Work Breakdown Structure

The project modules are divided into software, hardware and integration part to make management and development easier. Following is the work breakdown structure for our project:

TABLE 6: Work breakdown structure

Level	WBS	Task Description
1	1	Initiation
2	1.1	Evaluations and recommendations
2	1.2	Develop project charter
2	1.3	Deliverable: submit project charter
2	1.4	Project charter: signed/approved
1	2	Planning
2	2.1	Create preliminary scope statement
2	2.2	Determine the project team
2	2.3	Project team meeting
2	2.4	Develop project plan
2	2.5	Submit project plan
2	2.6	Milestone: project plan approval
1	3	Execution
2	3.1	Train machine learning model for face recognition
3	3.1.1	Develop text to speech synthesizer
3	3.1.2	Integrate OCR technology

2	3.2	Develop image description feature
3	3.2.1	Integrate image description into the model
3	3.2.2	Integrate microcontroller with the smart glasses
1	4	Testing
2	4.1	Testing ultrasonic sensor
2	4.2	Testing face recognition feature
2	4.3	Testing mode shifter
2	4.4	Testing the entire functioning of the smart glasses
2	4.5	Full system testing
1	5	Control
2	5.1	Project management
2	5.2	Project status management
2	5.3	Risk management
2	5.4	Update project management plan
1	6	Closeout
2	6.1	Document lesson learned
2	6.2	Update files and records
2	6.3	Gain formal acceptance
2	6.4	Archive files/documents

3.4 Tools and Technologies Used

TABLE 7: Tools and technologies used

Software	**Hardware**
Python OpenCV OCR Google API (GTTS) TensorFlow LBVH Haar Cascade LSTM	Raspberry Pi Pi camera Ultrasonic sensors Power bank Earpiece Connecting wires and resistors Mode shifter

This section discusses the design specifications of the product which comprises the system architecture and the logical representation of the system.

4.1 System Architecture

The system architecture for smart glasses are shown in following section

4.1.1 Component Diagram

Component diagrams are used to express the relationship between different components of the system. It is usually made when the system involves both software and hardware.

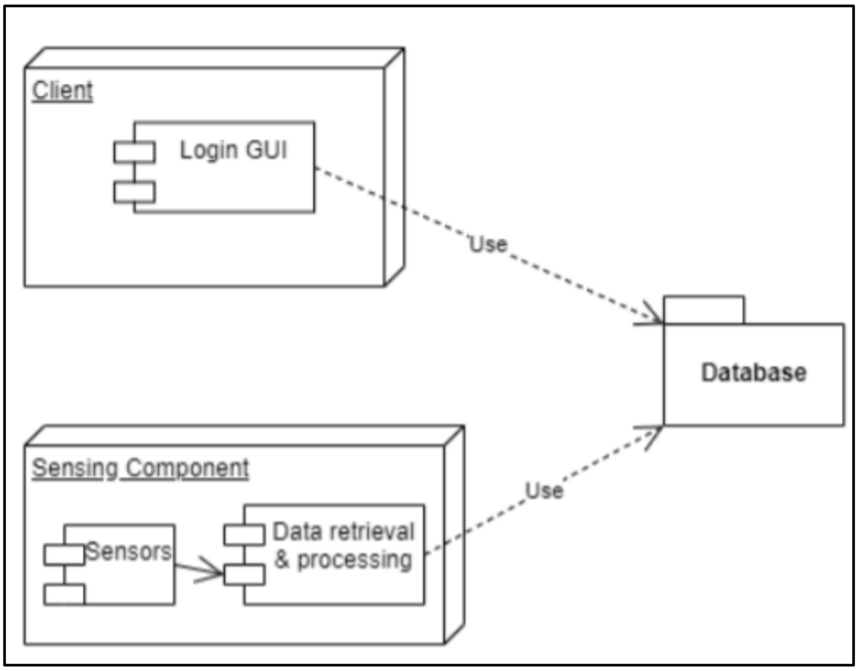


FIGURE 6: Component diagram

4.1.2 Tier Architecture Diagram

The tier architecture diagram depicts the server-client software architecture pattern. Users are authenticated in the application tier and further information is fetched from the data tier.

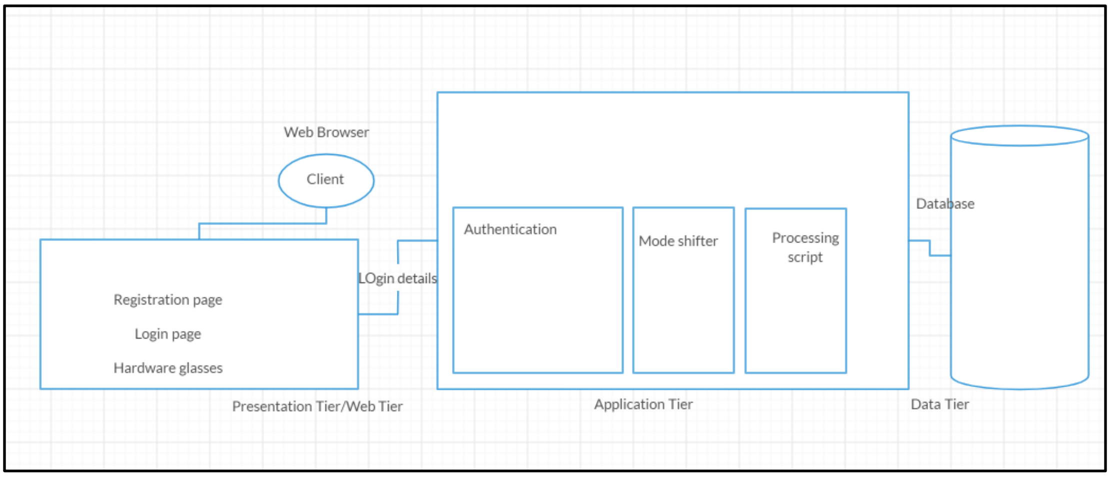


FIGURE 7: Tier architecture diagram

4.2 Design Level Diagrams

4.2.1 Class Diagram

Different classes of our project are login system, database, mode shifter, blind person and smart glasses having relationship and attributes is shown in figure below.

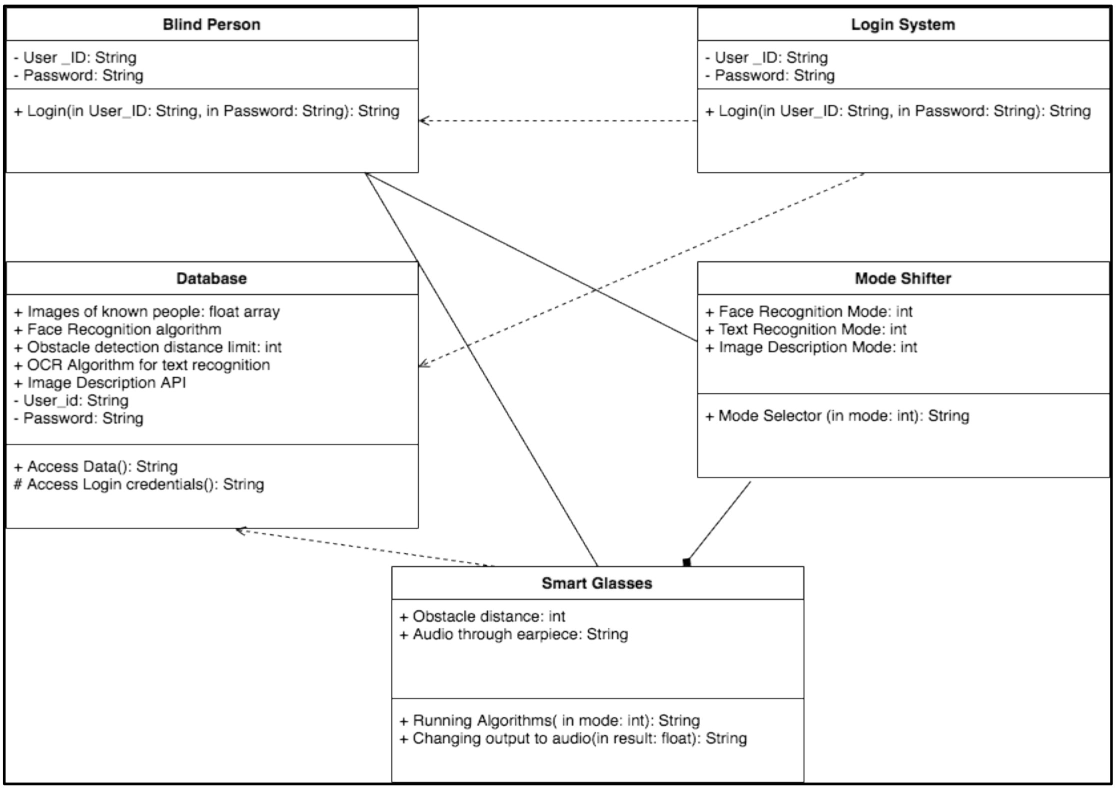


FIGURE 8: Class diagram

4.2.2 Activity Diagram

Activity diagram shows the steps the user will follow to perform the basic use case of the prototype, starting from logging in to using the various features.

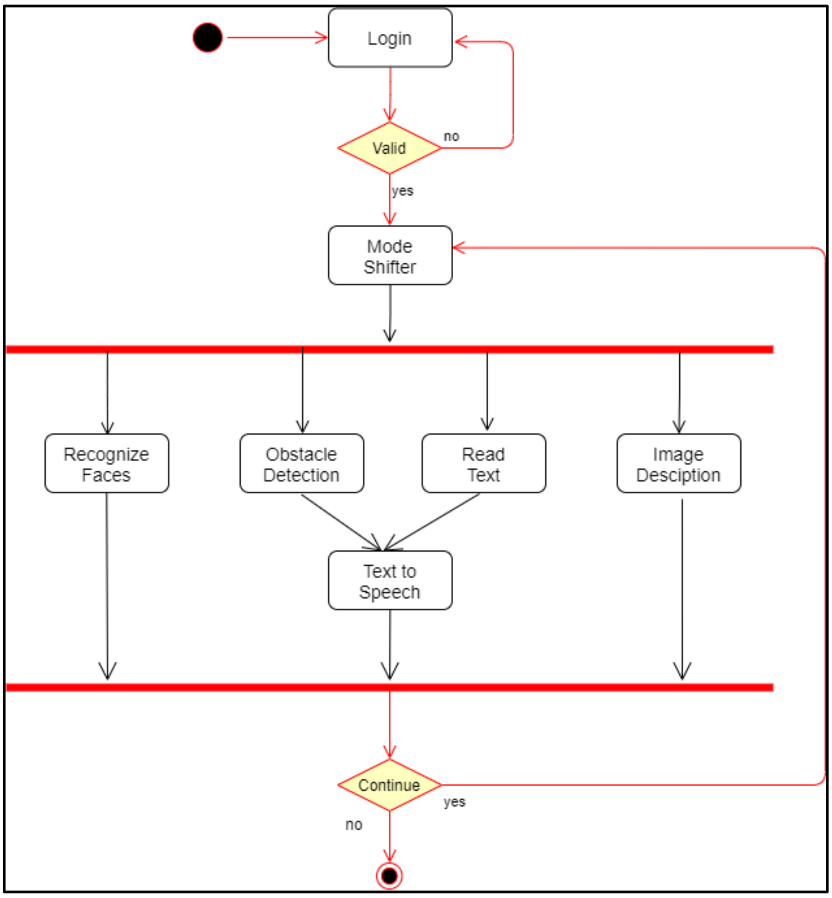


FIGURE 9: Activity diagram

4.2.3 Sequence Diagram

The sequence diagram depicts the time span of different objects in the system and the procedure of each function taking place.

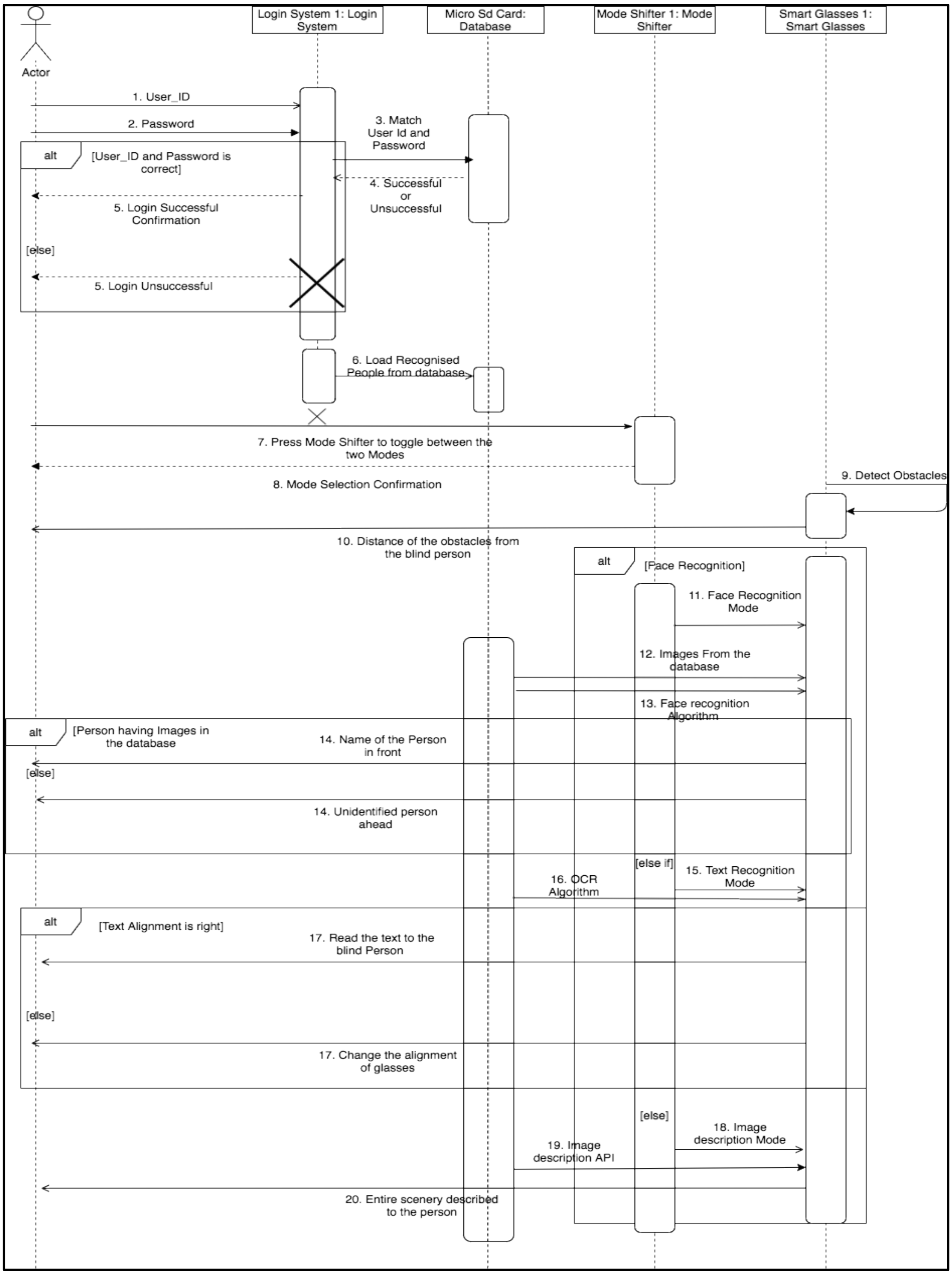


FIGURE 10: Sequence diagram

4.3 User Interface Diagrams

This includes the use case diagram which describes how different users interact with the software starting from login to the various features of the smart glasses.

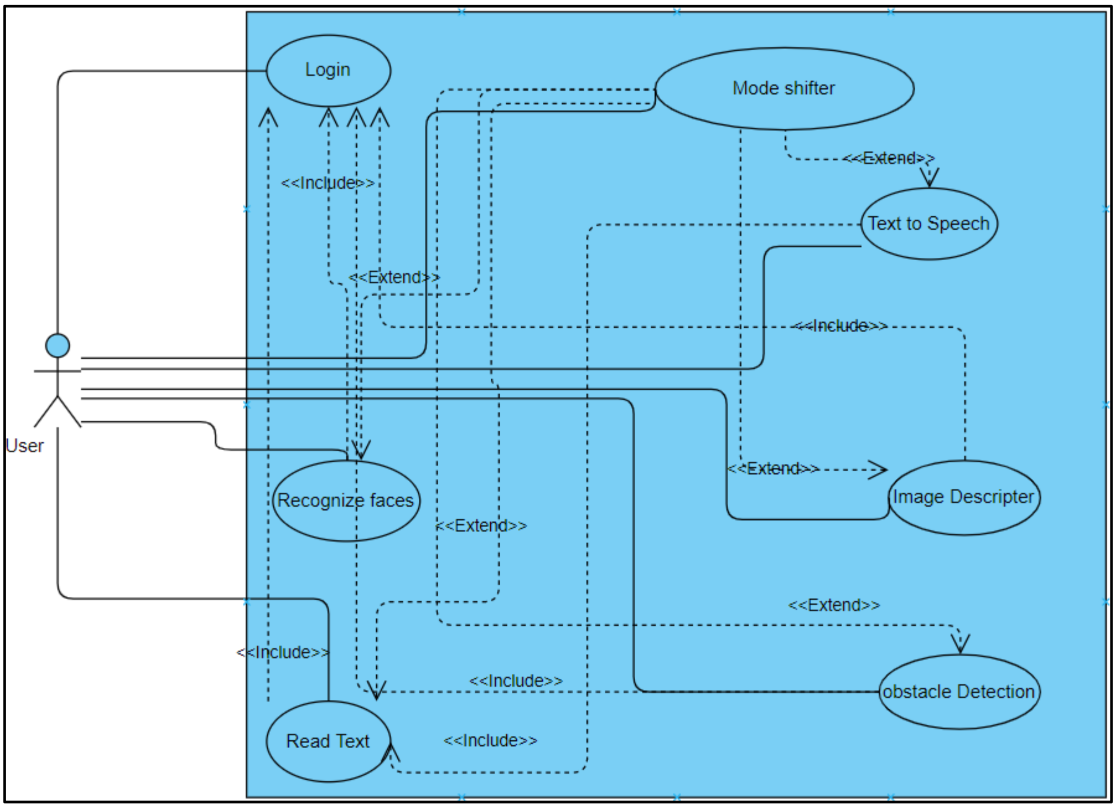


FIGURE 11: User interface diagram

IMPLEMENTATION AND EXPERIMENTAL RESULTS

We have tested the project on various corner cases. Majorly, we have focused on unit testing of the project, which begins with selecting the method to be tested and then taking the input parameters which allow or disallow the user to perform certain actions.

5.1 Experimental Setup

To examine and validate the functioning of our smart glasses we have used an experimental setup consisting of the following modules:

1. Ultrasonic sensors are mounted on the front frame of glasses. These sensors will help in detecting obstacles in front of the blind person and caution them about their proximity with the obstacles by informing them about their distance through an earpiece.
2. OpenCV library in which the Haar Cascade algorithm is used to detect faces in the images captured and LBPH classifier identifies the face by finding a match in the image database.
3. For image description feature we have used a multi-model technique in which features are extracted from the captured image using VGG16 model built using TensorFlow framework, having CNN architecture and subsequently captions are generated iteratively using LSTM model.
4. OCR software extracts text from the image and thereafter converts it into audio output using the text to speech synthesizer GTTS.

5.2 Experimental Analysis

In the experimental analysis we will perform specific tests on the smart glasses prototype to test their performance and accuracy. Testing of all the four modes is done using the data collected from various datasets available online for testing accuracy of machine learning models and also the data gathered from various daily life scenarios. In this section we will discuss two major components. First the description of data used for training the various machine learning models incorporated in our prototype. Second the performance parameters like model architecture, values of various hyperparameters and utility of different libraries used for training the machine learning models.

5.2.1 Data

The data was required primarily to train the deep learning models used for the two features, image description and face recognition. For the image description feature the dataset required to train the model consisted of images and their corresponding captions taken from Kaggle website. The dataset consisted of around 8000 images. Data cleaning was then done to remove the images with missing captions from the dataset. Thereafter noises like orientation and brightness were removed from the dataset images using the max pooling layer of the CNN architecture. The concept of transfer learning has been used here as the VGG-16 model is a very deep layered model, pre-trained on thousands of different images. Transfer learning refers to the idea of overcoming the isolated learning paradigm and utilizing knowledge acquired for one task to solve related ones.

For the facial recognition feature we trained the model by creating our own dataset of around 30 images of the face profile of all the group members. In the Haar Cascade model for face recognition, the model is pre-trained on a number of images since a dataset of 30 images is quite small to train a deep learning model and get a substantial prediction accuracy.

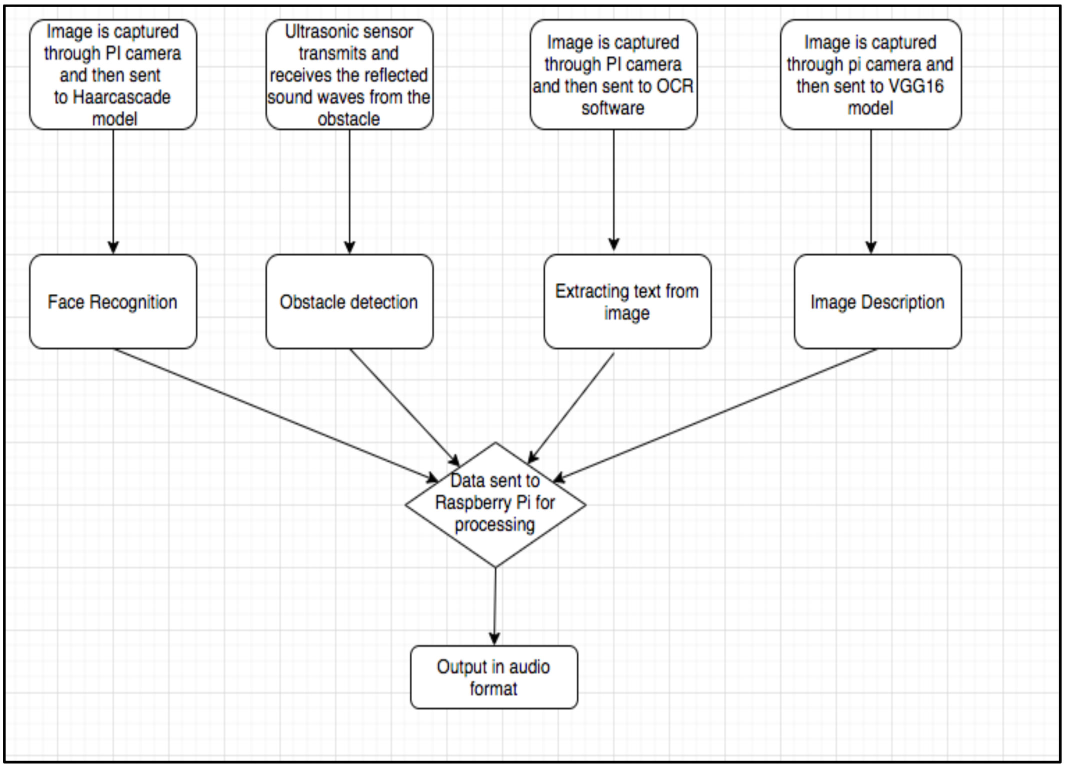


FIGURE 12: Flow of data

The above diagram shows the flow of data in the system and the steps followed for further processing to get the in-depth analysis.

5.2.2 Performance Parameters

After a lengthy discussion with our mentor and amongst the group members, we finalized following basic performance parameters to judge the performance of our prototype through this experiment. The parameters are as follow:

1. Facial recognition feature should be able to detect the person in front of the smart glasses and identify the familiar person from the image database. Even if there are more than one person in the frame, this feature should be able to recognize all the people in the image captured.

2. Text to speech feature should be able extract text from the image captured from a suitable distance and thereafter synthesize the extracted text clearly into audio cues

3. Image description feature should be able to explain most objects present in the image captured by pi camera and interaction with each other. The audio cue explanation of the scenario should not be much different when compared to that given by a person seeing their surroundings

4. The obstacle detection feature should be able to correctly identify the obstacles up to a distance of 3 metres from the blind person and thereafter intimate their correct distance from the user via audio cues.

5. The response time of all these features for producing the desired output should be low and there should not be much time taken to switch from one feature to another using the mode shifter.

5.3 Working of the Project

This section explains the step by step workflow of our project and the detailed working of the various algorithms used.

5.3.1 Procedural Workflow

On wearing the smart glasses, blind person will firstly have to login into the system using the credentials provided. This is done so as to authenticate the user and then correctly load the images of familiar people specific to that user into the database for the facial recognition feature. Next the blind person will select the function to be performed using the mode shifter. Features are selected using the mode shifter in a sequential manner starting from the first feature i.e. obstacle detection.

In the obstacle detection feature, the ultrasonic sensors which act as both transmitter and receiver send out high frequency ultrasonic waves to measure the distance of the obstacle from the user. These sensors can measure distance up to 3 metres from the user. Once the receiver receives the reflected ultrasonic waves, the time elapsed between transmission and reflection gives the distance of the obstacle from the user. This distance is then communicated to the blind person in audio format via the ear piece. The obstacle detection feature functions at all times even when a different feature is selected by the user. This is so that the blind person is warned every time an obstacle comes in front even while using different features.

On choosing the next feature i.e. the face recognition feature, the blind person will be able to know the person in front of them. In this feature, firstly the image is captured using the pi camera of the person in front. Then the Haar Cascade model in OpenCV library detects the face in the image using the facial landmarks and the sliding window technique. It makes a bounding box around the face detected and then the LBPH classifier identifies the cropped image from the image database. If the person is an acquaintance having their image in the database then the name of the person is communicated to the user otherwise it is intimated that it is a stranger.

In the third feature, text is converted to speech to enable the blind person to understand printed as well as handwritten text. In this feature, again the pi camera first captures the view in front. If the text is not properly visible or is cropped from the image then the glasses tell the user to change the orientation of the material. After confirmation that the text area is correctly positioned and readable using the red border outline, the image is processed in real time by the optical character recognition (OCR) software for text extraction and subsequently forwarded to the text-to-speech synthesizer. The text is then read to the user through the audio output port.

On selecting the last feature i.e., the image description feature, the blind person will be able to experience their surroundings through audio cues. This feature is a novelty of our project. The image description feature uses the multi-model technique for describing objects in the image iteratively. The VGG-16 model built using TensorFlow framework is a very deep convolutional neural network consisting of 16 layers pre-trained on thousands of images. This model uses the concept of transfer learning by utilizing the knowledge learnt from previous tasks to improve the generalization of related ones. After detecting objects in the image, captions are generated using the one-to-many sequence model, LSTM. Theses image captions are then converted to audio format to be transmitted to the user.

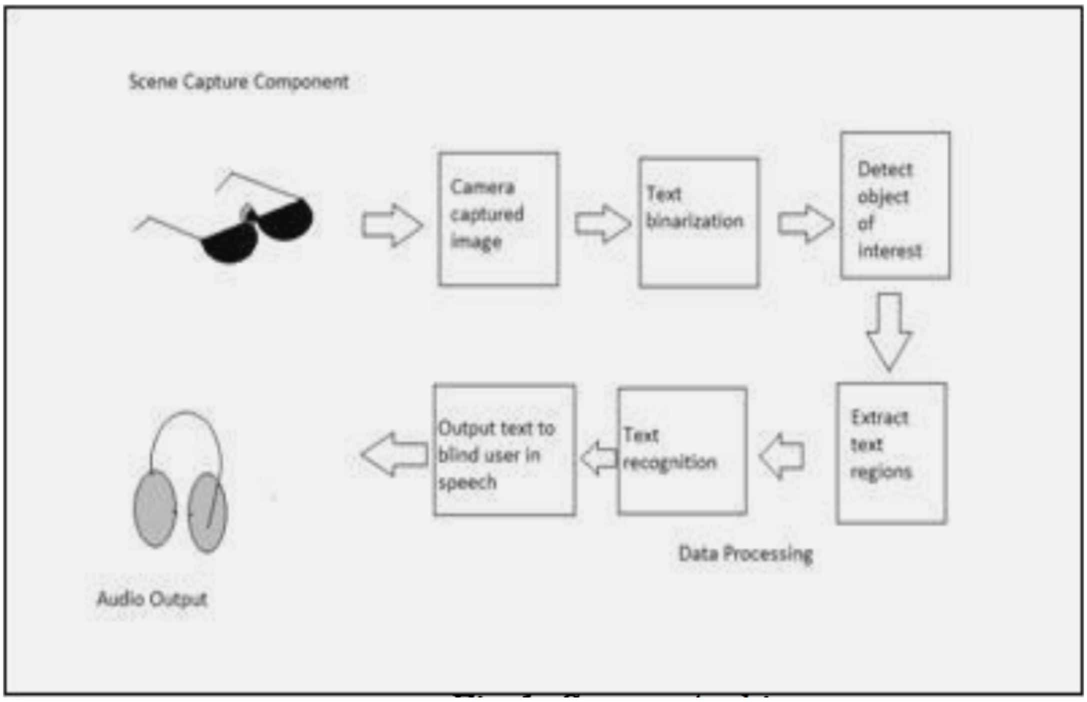


FIGURE 13: Procedural workflow

5.3.2 Algorithms Approaches Used

Following is the explanation of the 2 algorithms used in our project.

5.3.2.1 Haar Cascade Algorithm for Face Recognition

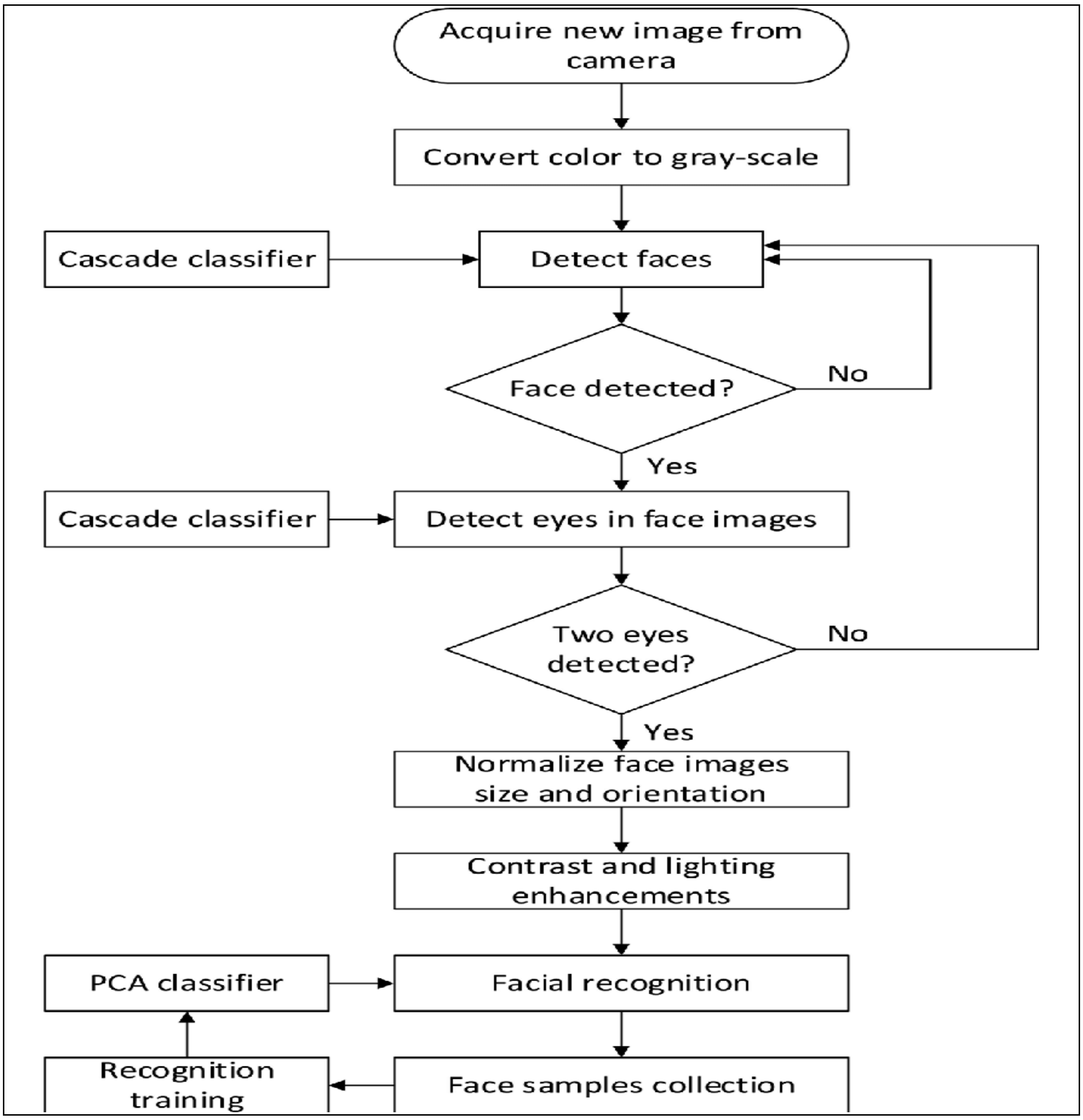


FIGURE 14: Flow Diagram for Haar Cascade Algorithm

Haar Cascade is a machine learning object detection algorithm used to identify objects in an image or video and based on the concept of features proposed by Paul Viola and Michael Jones in their paper "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001.

It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images.

The algorithm has four stages:

1. Haar Feature Selection
2. Creating Integral Images
3. Adaboost Training

4. Cascading Classifiers

Following are the steps to use Haar Cascade for face Recognition.

Step 1

Scale down the images by resizing them for better output.

Step 2

Load the image and convert it into gray-scale. Images that are captured using PI camera are in the form of RGB (Red, Green, Blue) channel. So, when OpenCV reads the RGB image, it usually stores the image in the BGR (Blue, Green, Red) channel. For the purposes of image recognition, we need to convert this BGR channel to gray channel. The reason for this is that the gray channel is easy to process and is computationally less intensive as it contains only 1-channel of black-white.

Step 3

After converting the image to gray- scale, locate the exact features in the face.

Step 4

Finally draw a rectangle or bounding box around the detected face.

```
import cv2
import numpy as np


face_classifier =
cv2.CascadeClassifier('/haarcascade_frontalface_default.xml')


gray = cv2.cvtColor(resized, cv2.COLOR_BGR2GRAY)


''' Our classifier returns the ROI of the detected face as a tuple,
It stores the top left coordinate and the bottom right
coordiantes'''


faces = face_classifier.detectMultiScale(gray, 1.0485258, 6)


'''When no faces detected, face_classifier returns and empty
tuple'''
if faces is ():
    print("No faces found")


'''We iterate through our faces array and draw a rectangle over each
face in faces'''
for (x,y,w,h) in faces:
    cv2.rectangle(resized, (x,y), (x+w,y+h), (127,0,255), 2)
    cv2.imshow('Face Detection', resized)
    cv2.waitKey(0)

cv2.destroyAllWindows()
```

FIGURE15: Pseudocode for Haar Cascade algorithm

5.3.2.2 VGG-16 Model for Image Description

VGG-16 is a convolution neural net architecture which was used to win ILSVR(ImageNet) competition in 2014. The most unique thing about VGG16 is that instead of having a large number of hyper-parameters they focused on having convolution layers of 3x3 filter with a stride 1 and always used same padding and maxpool layer of 2x2 filter of stride 2. It follows this arrangement of convolution and max pool layers consistently throughout the whole architecture. In the end it has 2 FC (fully connected layers) followed by a softmax for output. The 16 in VGG-16 refers to it has 16 layers that have weights. This network is a pretty large network and it has about 138 million (approx.) parameters.

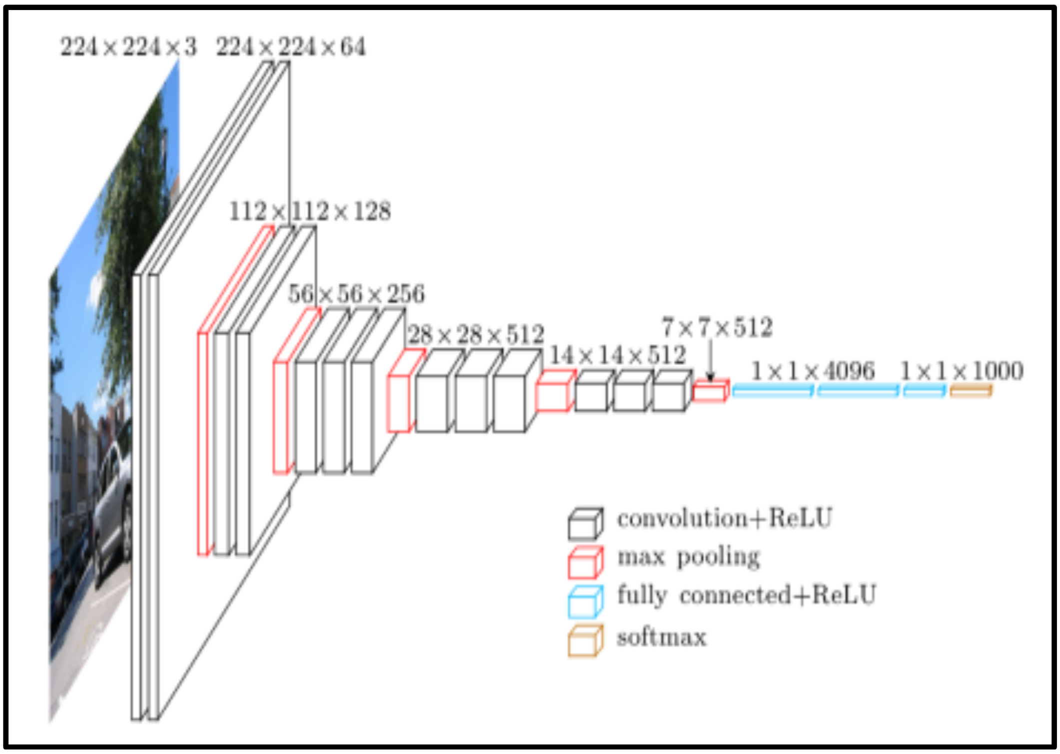


FIGURE 16: Architecture of VGG-16

5.3.3 Project Deployment

Component Diagram (figure 6) depicts the different components of the system and their connections between each other. In smart glasses for blind, login credentials are stored in a server database and upon successful login retrieval of corresponding images of the known people is done from the micro SD card for the face recognition feature.

Deployment diagram below shows the hardware components namely the ultrasonic sensor, Pi camera, mode shifter and the micro SD card connected to the Raspberry Pi microprocessor. For each module of the project a python file has been created containing the code related to that module and are imported in the main.py python file which finally runs on Raspberry Pi.

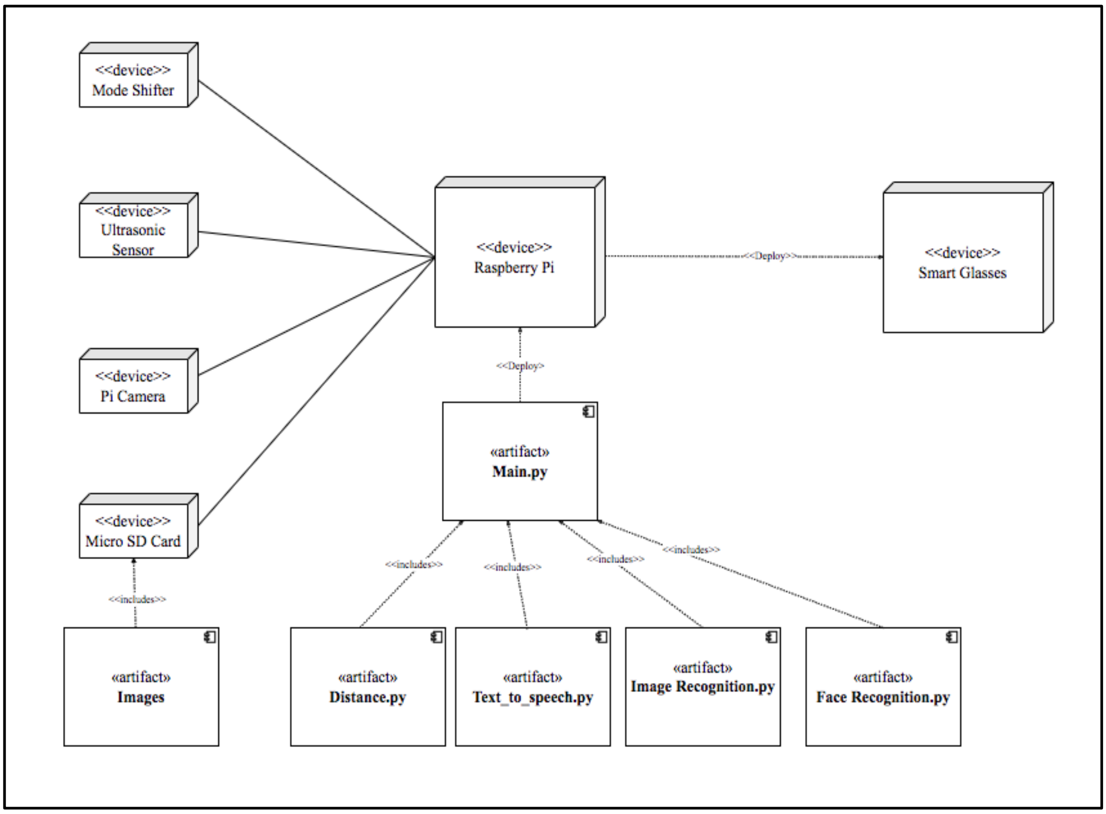


FIGURE 17: Deployment diagram

5.3.4 System Screenshots

5.3.4.1 System Prototype

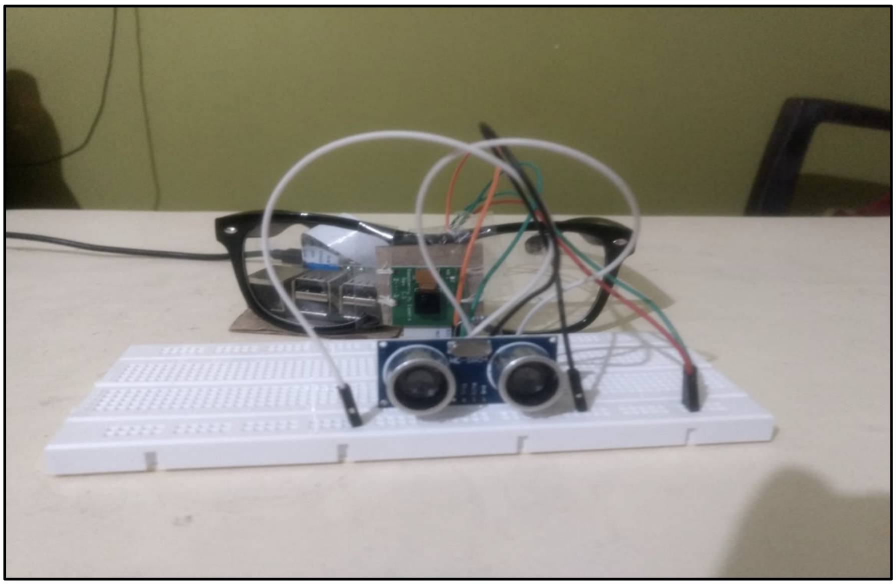


FIGURE 18: Our smart glasses prototype integrated with all the components

5.3.4.2 Face Recognition Feature

FIGURE 19: Facial recognition taking place using the smart glasses

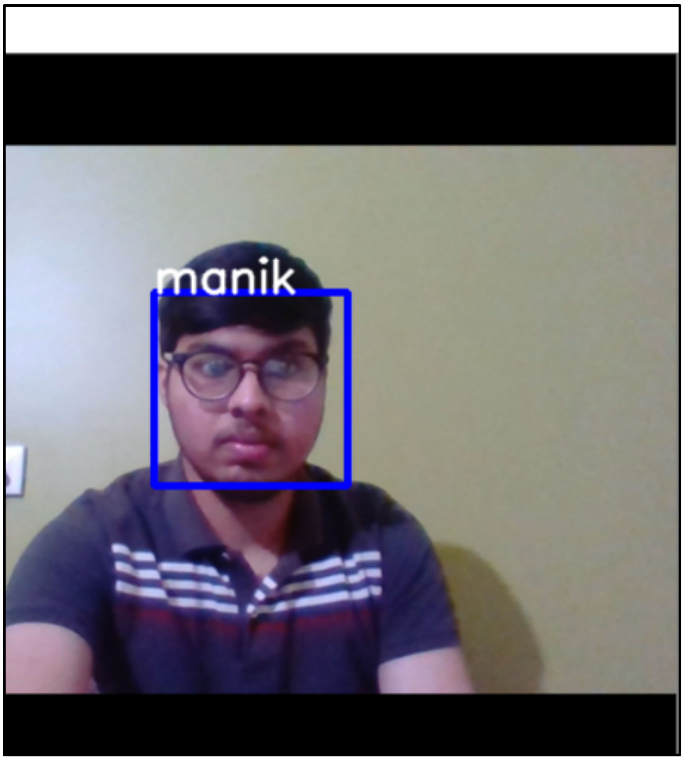


FIGURE 20: Bounding box around the detected face and name of the person identified

5.3.4.3 Reading Text using Smart Glasses

FIGURE 21: User reading text using the smart glasses

FIGURE 22: Image of the text captured by the Pi camera

5.3.4.4 Image Description Feature

FIGURE 23: Image captured for image description feature

```
WARNING:tensorflow:From D:\image_caption\virtual\lib\site-packages\tensorflow\python\keras\backend.py:3794: add_dispatch_support.<locals>.wrapper (from tensorflow.python.ops.array_ops) is deprecated and will be removed in a future versio
n.
Instructions for updating:
Use tf.where in 2.0, which has the same broadcast rule as np.where
man in red shirt is sitting on bench
```

FIGURE 24: Output on the command prompt of the caption generated of the above image

5.3.4.5 Obstacle Detection Feature

FIGURE 25: Obstacle in front detected by the smart glasses

5.4 Testing Process

This section includes the test plan, the different various test cases and their results, and the strategies undertaken to ensure the successful completion of this project.

5.4.1 Test Plan

The test plan was decided in a way so as to ensure we stayed in alignment with our objectives and performance parameters. We all put our ideas forth and decided that the best way forward would be to establish small progressive goals and test the project in a stepwise manner we proceeded to the next feature.

5.4.2 Features to be Tested

This project consists of 4 different modules. In the distance module working of the ultrasonic sensor will be tested whether it is giving the correct distance of any obstacle as soon as it comes in front of the glasses. In the face recognition module, we will test whether this feature identifies the person in front correctly and also if it is able to detect the faces in the image captured by making a bounding box around them. For the image description module, we will test whether the feature is able to explain the objects in the image captured to certain extent and how close it is to the human explanation of the scenario. In the last feature, we will test if the OCR software is correctly able to extract text from the image and also inform the user to correct the orientation of the material if the text is cropped or not properly visible. We will also test the text synthesizer to see whether it is correctly able to convert text to speech and transmit the same to the user.

5.4.3 Test Strategy

We would separately test the individual features of our project and finally test the integrated module. The strategy used for different modules to achieve high accuracy is to repeat testing again and again using different test cases and along with this testing the corner cases as well.

5.4.4 Test Techniques

We have used a variety of test techniques for testing our project as explained above in the test strategy. Blackbox testing is used to test whether the correct output is obtained for a given input without looking in the system functionalities whereas Whitebox testing does the same while improving the system functionalities for increasing functionality. We performed manual testing for our project. Classifying Whitebox testing further, we also performed control flow and data flow testing. Control flow testing involves testing all the condition cases and data flow testing involves testing all data variable functionalities. We also performed unit, integration and system testing.

5.4.5 Test Cases

TABLE 8: Test cases and their results

Test Case ID	Scenario	Test step	Expected outcome	Actual outcome
T001	Check whether the ultrasonic sensor is measuring distance correctly	Ran the code and verified the printed distance manually using scale	Print the distance accurately in real time	Print the distance as actual
T002	Check the functioning of image description feature	Ran the code and verified that this feature correctly identifies the objects in the image and explains nearby things.	Explaining the image in real time accurately	Explained the image in real time accurately
T003	Check the OCR software to extract text from image	Fetched the required data from OCR software	Get the requested data in real time	Got the data required
T004	Check the working of function of facial recognition	Ran the code and verified that it should correctly recognize the faces	Correctly determining the face of person whose data is stored	Correctly determined the face of person whose data is stored
T005	Check the working of Pi camera	To capture images	See the feed from the camera	Saw the video captured and image captured
T006	Check the working of Google API GTTS	Convert text extracted	Converting the text to speech immediately	Converted the text to speech immediately

		from the image to speech		

5.4.6 Test Results

The test results came out to be all positive. This was not so the first time – thus we had to make some adjustments and amendments to the project which only bettered it in the long run. Also, the abovementioned test cases are only the core, i.e. the major ones. There were other sub-functionalities too that were tested but they were somehow associated with the major test cases themselves, hence it was the major test cases, the successful completion of which meant the successful execution of the project. Hence, seven major test cases have been mentioned.

5.5 Results and Discussions

The overall experiment proved the merit of our prototype. Thus, demonstrating the efficiency and accuracy of our project. Some specific results and scope for improvement are as follow:

1. The system we deployed is successfully describing any image captured. Input image given to the image description feature is captured from the real time video stream. The video is divided into frames and then an image is taken from that feed and sent for processing to Raspberry Pi microcontroller. After processing the image, the corresponding caption is generated using the sequence model LSTM. Thereafter it is converted to audio output using the text to speech synthesizer.
2. We have tried to optimized our prototype as much as possible by reducing the response time of the Pi camera as well as decreasing the processing time of machine learning algorithms on the microcontroller. Optimization is a very critical component of our prototype and every second increase in the processing time makes a huge difference in the outcome. Also, the real time usage of our prototype requires the output to be generated instantly otherwise the assistive device will not be of much usage.
3. For facial recognition, the image captured by the Pi camera is sent to Raspberry Pi for processing and finally converted to audio output which is subsequently transmitted to the user.

4. Though the response time of the entire system has been mitigated to quite some extent, it still requires hardware upgradation for efficient usage in everyday tasks. For e.g. a more robust and higher picture quality camera with night vision is required to capture images in all lighting conditions; a faster processor for high end processing to reduce the response time of the output; more RAM for more storage space and multitasking capability.
5. Smart glasses with a sleek design, easily wearable and usable with a faster processor and better camera for image capturing would be a great areas to improve upon, but it looked quite difficult for us considering the resources we had, both in terms of money and time.

5.6 Inferences Drawn

Following are the major inferences drawn from this experiment:

1. Response time: The entire prototype needs to be robust, comfortable and user friendly while keeping the response time of the system quiet low for it to be effectively used as an assistive device by the blind people. The Pi camera response time plays an important role in capturing good quality, focused images from the real-time video stream and the hardware module should have a good processing capacity to fasten the processing time of the machine learning algorithms.

2. Successful implementation of algorithms: In order to make our prototype even more robust we have to ensure that the camera lens is not damaged in a scenario due to its continuous usage. Thus, we have to ensure the usage of the best quality camera having a long-lasting life for it to give maximum throughput. Another constraint which needs to be kept in mind is the processing time of the images. Hence, we are using extensive algorithms like VGG-16 for image description but they can further be improved upon with better processing power of the hardware.

3. Google API GTTS is the one of the best algorithms for converting text to speech.

5.7 Validation of Objectives

TABLE 9: Validation of objectives

S. No.	Objective	Status
1	To use machine learning models to recognize people. Blind person is able to recognize people they know with the help of the face recognition module. For this, we have used the OpenCV library in which Haar Cascade detects faces and LBPH classifier identifies the face. .	Successfully integrated all the necessary modules and code with the hardware. Verified the proper working of this feature.
2	Describe the scenario to the person in audio cues. For this we have used a multi-model technique in which features will be extracted from the image using VGG16 model in TensorFlow, having CNN architecture, and captions would be generated using LSTM.	Successfully developed this feature and verified that it is working correctly.
3	Detect the obstacles using ultrasonic sensors. This feature will help the user by intimating the distance of the obstacles in front.	Successfully built this feature and it correctly tells the distance of the obstacle in front of the user.
4	Convert text to speech and describe real-world images. Through the use of this feature the user will be able to enjoy books as well as handwritten text. It uses OCR software for reading text from the images and GTTS API to convert text to speech and then subsequently transmit it to the user.	Successfully made this feature and it is converting text to speech efficiently.

6.1 Conclusions

The present work emphasizes on the development of smart glasses using a number of algorithms and technologies which will help the blind people to afford an comprehensive, assistive device at a low cost. These glasses will assist the blind person in their daily routine tasks and they will not require another person's assistance for accomplishing their everyday tasks.

The obstacle detection will allow them to know about the distance of obstacles in front and they will be able to understand printed as well handwritten text. These smart glasses will enable them to have a sense of security because using the facial recognition feature they can know whether the person in front is familiar or not. So, this project will help the visually impaired in their day-to-day tasks and provide them with a convenient, easy to use and comfortable assistive device.

This capstone project has been a learning curve for all the members in the group with everyone learning from every other in the group, thereby enhancing our knowledge and understanding of the subject as well as team work. This project has been instrumental both in increasing our knowledge on software development and in telling us about the various processes and documentation associated with it. It has given us deeper insights on our course subjects which have been used here, thereby giving us the opportunity to learn the practical implementations of the same. Further it has helped us develop leadership skills and also has taught us how to work in a team and produce the best output.

6.2 Environmental, Economic and Societal Benefits

6.2.1 Environmental Benefits

With the help of these glasses blind people can read and understand printed and handwritten books. Thus, there is no need of making separate braille reading cardboards. This will save a lot of cardboard and paper and hence save trees and contribute to sustainable development.

6.2.2 Economic Benefits

The smart glasses being built have a low-cost price but at the same time it is ensured that the product quality is at the same level as those existing in the market, if not more. Moreover, the prototype is being built to ensure that it is an all-round assistive device for the blind.

6.2.3 Societal Benefits

These smart glasses are built keeping in mind the lives of visually impaired people and the difficulties they face in doing their chores. Hence these glasses provide them an easier way to function and accomplish their everyday tasks. It makes the lives of the blind easier by providing them with an all-round assistive device.

6.3 Reflections

This project started with the idea of helping blind people by providing them with to experience their surroundings. To give them virtual eyes so that they do not require assistance and will be able to know what’s going on in their surroundings. We have come so far, by understanding the real-life challenges and difficulties faced through this project and this project truly helped us to grow as an individual not only by learning new skills but also learning how to deal with different people and get the work done. Various skills were enhanced by working on this project such as team work, time management, decision making, multitasking and many more. All these reflections from the project helped us to develop personally as well as on academic grounds.

6.4 Future Work

- Camera optimization for improving fast capturing.
- Improve the prototype and improve the accuracy in the Image description model.
- Better glasses design for comfort in wearing and in ease of use.

7.1 Challenges Faced

We faced the following challenges while working on our project:

- Environmental conditions: The camera gave problems in situations of different light
  conditions other than the default one. Also, the microphone was recording a lot of noise
  along with the desired sound.
- Bluetooth connectivity issues: Bluetooth connectivity was needed for the voice module which was a bit difficult the first time since there were some connection issues with the Raspberry Pi.
- Integration issues: The modules were running properly independently however when they were put together and deployed on the microcontroller there were some issues shifting between the features.
- Connection issues: The connections with the raspberry pi had to be checked almost continuously as they got loose fast.

7.2 Relevant Subjects

Table 10: Subject code and subject name

Subject code	Subject name	Description
UCS503	Software Engineering	Assisted in UML diagrams required for documentation, designing of test cases and system design.
UML501	Machine Learning	Training and testing of image dataset, working of python, well-tuned model.
UCS742	Deep Learning	Assisted knowledge of CNN and RNN, working on various activation functions and OCR techniques.
UEC001	Electronics	Helped in understanding how to configure hardware devices in our project like raspberry pi 2, ultrasonic sensor etc.
UCS615	Digital Image Processing	Helped in pre-processing of images and error analysis.

7.3 Interdisciplinary Knowledge Sharing

During the course of the project, members had various group sessions to discuss the feasibility and utility of the idea. The sharing of knowledge between human psychology and utilitarian thoughts helped develop the software as it is today. The principles management as well ancient rules of stoic helped us stick together as a team. Sharing knowledge between computational image processing and machine learning made image detection a lot better. The joint strides in image processing and deep learning helped bring life to the image description and helped solve the problem of performance bottleneck.

7.4 Peer Assessment Matrix

TABLE 11: Peer assessment matrix

		Evaluation of			
		Manav Sharma	Manik Mangla	Priyanshu Tuli	Yashwika
Evaluation by	Manav Sharma	4	5	5	5
	Manik Mangla	5	4	5	5
	Priyanshu Tuli	5	5	4	5
	Yashwika	5	5	5	4

7.5 Role Playing and Work Schedule

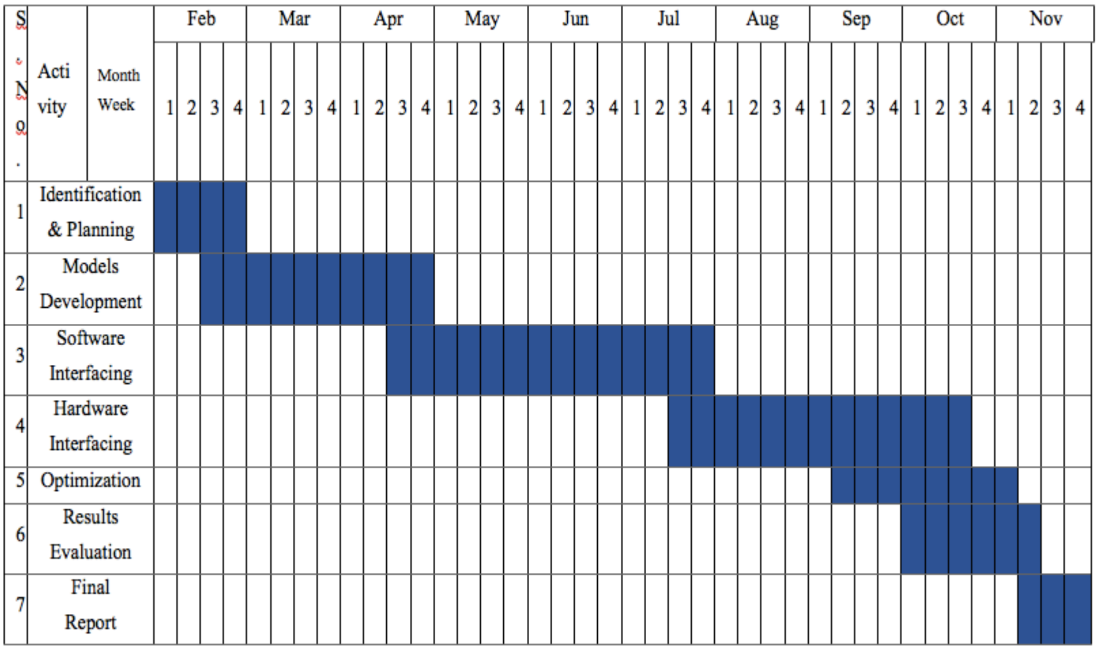
Manav Sharma: Creation and integration of text-to-speech module, hardware connections, Raspberry Pi linking, documentation

Manik Mangla: Creation and integration of image description module, dataset creation and collection, Raspberry Pi linking, hardware connections, documentation

Priyanshu Tuli: Creation and integration of facial recognition module, hardware connections, testing, documentation, UML diagrams

Yashwika: Creation and integration of text-to-speech module, documentation, testing, UML diagrams

TABLE 12: Gantt chart



7.6 Student Outcomes Description and Performance Indicators (A-K Mapping)

TABLE 13: Student outcomes

S. No.	Description	Outcome
A1	Applying mathematical concepts to obtain analytical and numerical solutions.	
A2	Applying basic principles of science towards solving engineering problems.	Applied basic principles of science for solving engineering problems.
A3	Applying engineering techniques for solving computing problems.	Used the concept of neural networks for image recognition.
B1	Identify the constraints, assumptions and models for the problems.	Stated the constraints, assumptions and models.
B2	Use appropriate methods, tools and techniques for data collection.	Collected data using appropriate techniques.
B3	Analyse and interpret results with respect to assumptions, constraints and theory.	Analysed and interpreted the results with respect to assumptions, constraints and theory.
C1	Design software system to address desired needs in different problem domains.	Designed the specified software system to fit the needs
C2	Can understand scope and constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability.	Designed the system such that all the constraints were taken care of.
D1	Fulfil assigned responsibility in multidisciplinary teams.	Built the desire modules in separate teams
D2	Can play different roles as a team player.	Played different roles as a team player.

E1	Identify engineering problems.	Clearly identified the engineering problems such as hand gesture and voice command recognition and interfacing.
E2	Develop appropriate models to formulate solutions.	Formulated practical solutions for the desired problems
E3	Use analytical and computational methods to obtain solutions.	Yes, we were able to obtain solution by using experimental investigative techniques and analysis.
F1	Showcase professional responsibility while interacting with peers and professional communities.	Showed professional responsibility while interacting with others.
F2	Able to evaluate the ethical dimensions of a problem.	Yes, our problem identified was ethical as it helped those who need our utmost care and support.
G1	Produce a variety of documents such as laboratory or project reports using appropriate formats.	Produced the required documentation
G2	Deliver well-organized and effective oral presentation.	Delivered effective oral presentations
H1	Aware of environmental and societal impact of engineering solutions.	Yes, our project was developed by considering both the aspects.
H2	Examine economic trade-offs in computing systems.	
I1	Able to explore and utilize resources to enhance self-learning.	Yes, we were able to achieve it by team coordination and interdisciplinary knowledge gathering.
I2	Recognize the importance of life-long learning.	Yes, it is our major learning from the capstone project.
J1	Comprehend the importance of contemporary issues.	Yes, we reviewed a variety of issues for our project development and it made us more aware of such issues.
K1	Write code in different programming languages.	Wrote codes in different languages and platforms
K2	Apply different data structures and algorithmic techniques.	Applied different techniques studies in our course
K3	Use software tools necessary for computer engineering domain	Applied and used the different software tools necessary

7.7 Brief Analytical Assessment

Q1. What sources of information did your team explore to arrive at the list of possible project problems?

Ans: The group was aware of the understanding of the capstone requirement and some of the problems that were required to be explored. We explored the literature, mostly journals and research papers from various institutions. Some articles were picked form the internet as well and played a crucial part in finalising at the list of possible project problems.

Q2. What analytical, computational and/or experimental methods did your project team use to obtain solutions to the problems in the project?

Ans: Our project was divided into five parts consisting of building the face recognition module; the image description module; the voice recognition module; the obstacle detection module and integration with hardware. We used CNN and RNN for image description and Google API for voice command module. For obstacle detection, we used ultrasonic sensors and for microcontroller we used Raspberry Pi 2.

Q3. Did the project demand demonstration of knowledge of fundamentals, scientific and/engineering principles? If yes, how did you apply?

Ans: We used quite a lot of engineering subjects. Concepts from subjects Deep Learning, Machine Learning and Image Processing were used in building the image description and voice recognition module. Natural Language Processing was used for building the facial recognition module. Engineering Design II and Operating Systems were used for integration and working with the microcontroller and the ultrasonic sensor.

Q4. How did your team share responsibility and communicate the information of schedule with others in the team to coordinate design and manufacturing dependencies?

Ans: Our team consisted of four members. We divided the project into subtasks, each individual carrying out specific tasks and in turn helping each other out. Information was communicated via mails, online meetings and Whatsapp messages as well.

Q5. What resources did you use to learn new materials not taught in class for the course of the project?

Ans: We took online tutorials and online courses. Moreover, we read various blogs and articles on websites link Kaggle, Towards Data Science and Medium to learn new concepts.

Q6. Does the project make you appreciate the need to solve problems in real life using engineering and could the project development make you proficient with software development tools and environments?

Ans: Our project was built for disabled people to make their lives easier by providing them with an all-round assistive device. Working on this project made us appreciate the need to solve real life problems using engineering. This project taught us a lot about new technologies and about software development life cycles, thereby making us proficient in the same.

REFERENCES

[1] RW Van Boven, RH Hamilton, T Kauffman, JP Keenan… - Neurology, 2000 AAN Enterprises.

Tactile spatial resolution in blind Braille readers Available: https://n.neurology.org/content/54/12/2230.short

[2] AZ Ozgur, HI Gurcan - Turkish Online Journal of Educational Technology …, 2004 – ERIC

An Audio-Book Project for Blind Students at the Open Education System of Anadolu University. Available: https://eric.ed.gov/?id=EJ1101899

[3] B Slotznick, S Sheetz - US Patent App. 11/642,247, 2007 Method and apparatus for interacting with a visually displayed document on a screen reader Available:https://patents.google.com/patent/US20070211071A1/en

[4] Na'aman, A Shashua, Y Wexler - US Patent App. 13/397,919, 2012 - Google Patents User wearable visual assistance systemE Available: https://patents.google.com/patent/US20120212593A1/en

[5] Ashwini Kumar Sinha (2019) New Facial Recognition Smart Glasses for Visually Challenged Person [Online]. Available: https://www.electronicsforu.com/electronics-projects/new-facial-recognitionsmart-glasses-for-visuallychallenged-person

[6] Rohit Agarwal, Nikhil Ladha, Mohit Agarwal, Kuntal Kr. Majee (2017) Low cost ultrasonic smart glasses for blind [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8117194

[7] Andrei Popleteev, Nicolas Louveton, Roderick McCall (2015) Colorizer: Smart Glasses Aid for the Colorblind [Online]. Available: https://doi.org/10.1145/2753509.2753516

[8] Yongsik Jin, Jonghong Kim, Bumhwi Kim, Rammohan Mallipeddi, Minho Lee (2015) Smart Cane: Face Recognition System for Blind [Online]. Available: https://doi.org/10.1145/2814940.2814952

urkund

Document Information

Analyzed document	Capstone_Report_Final (2).pdf (D90523576)
Submitted	12/23/2020 6:10:00 AM
Submitted by	Rajkumar Tekchandani
Submitter email	rtekchandani@thapar.edu
Similarity	4%
Analysis address	rtekchandani.thapar@analysis.urkund.com

Sources included in the report

W	URL: https://www.cse.iitd.ac.in/~chetan/papers/mavi-vlsid2017.pdf Fetched: 2/18/2020 7:12:44 AM	1
W	URL: https://www.ijser.in/archives/v7i4/IJSER18769.pdf Fetched: 2/2/2020 2:30:42 AM	10
W	URL: https://jespublication.com/upload/2020-110437.pdf Fetched: 7/16/2020 2:19:18 PM	1
W	URL: https://www.irjet.net/archives/V7/i8/Velammal/NCRACES-33.pdf Fetched: 11/18/2020 3:58:15 AM	6
W	URL: https://eric.ed.gov/?id=EJ1101899 Fetched: 12/23/2020 6:11:00 AM	2
W	URL: https://www.researchgate.net/publication/342887708_Visually_Impaired_System Fetched: 12/2/2020 4:27:02 PM	1
W	URL: https://www.researchgate.net/publication/271855592_Design_of_a_Mobile_Face_Recogni ... Fetched: 10/16/2019 7:19:12 AM	1
W	URL: https://www.researchgate.net/publication/339509454_Multi-Purpose_Assist_Device_for ... Fetched: 8/22/2020 2:50:45 PM	1