VISVESVARAYA TECHNOLOGICAL UNIVERSITY
Jnana Sangama, Belagavi - 590 018

Technical Seminar Report
on

An AI-Based Visual Aid with Integrated Reading Assistant for the Completely Blind

Submitted in partial fulfillment for the award of degree of

Bachelor of Engineering
in
Electronics and Communication Engineering

Submitted by

Aditya Venkata Sheshu
1RN18EC167

Internal Guide
Dr.Nandini K.S
Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
(Accredited by NBA for the Academic Years 2018-19, 2019-20,2020-21 and 2021-2022)

RNS INSTITUTE OF TECHNOLOGY
(AICTE Approved, VTU Affiliated and NAAC 'A' accredited)
(UG Programs - CSE, ECE, ISE, EIE and EEE have been Accredited by NBA for the Academic Years 2018-19, 2019-20,2020-2021,2021-2022)
Channasandra,Dr.Vishnuvardhan Road,Bengaluru-560098

2021-22

VISVESVARAYA TECHNOLOGICAL UNIVERSITY
Jnana Sangama, Belagavi - 590 018

Technical Seminar Report
on

An AI-Based Visual Aid with Integrated Reading Assistant for the Completely Blind

Submitted in partial fulfillment for the award of degree of

Bachelor of Engineering
in
Electronics and Communication Engineering

Submitted by

Aditya Venkata Sheshu
1RN18EC167

Internal Guide

Dr.Nandini K.S
Assistant Professor

ESTD 2001

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
(Accredited by NBA for the Academic Years 2018-19, 2019-20 and 2020-21)

RNS INSTITUTE OF TECHNOLOGY
(AICTE Approved, VTU Affiliated and NAAC 'A' Accredited)
(UG Programs - CSE, ECE, ISE, EIE and EEE have been Accredited by NBA for the Academic Years 2018-19, 2019-20, 2020-2021 and 2021-2022)
Channasandra,Dr.Vishnuvardhan Road,Bengaluru-560098

2021-22

RNS INSTITUTE OF TECHNOLOGY
(AICTE Approved, VTU Affiliated and NAAC 'A' Accredited)
(UG Programs - CSE, ECE, ISE, EIE and EEE have been Accredited by NBA for the Academic Years 2018-19, 2019-20, 2020-2021 and 2021-2022)
Channasandra,Dr.Vishnuvardhan Road,Bengaluru-560098

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
(Accredited by NBA for the Academic Years 2018-19,2019-20,2020-21 and 2021-2022)

ESTD 2001

CERTIFICATE

This is to certify that the Technical Seminar presentation work entitled **"An AI-Based Visual Aid with Integrated Reading Assistant for the Completely Blind"** has been successfully carried out by **Aditya Venkata Sheshu** bearing the USN **1RN18EC167**, bonafide student of **RNS Institute of Technology** in partial fulfillment for the award of Bachelor of Engineering in **Electronics and Communication Engineering** from **Visvesvaraya Technological University, Belagavi**, during the year 2021-2022. It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The Technical Seminar report has been approved as it satisfies the academic requirements in aspect of the work prescribed for the award of degree of **Bachelor of Engineering**.

Dr. Nandini K.S	Dr. Vipula Singh
Assistant Professor	Head of Department

RNS INSTITUTE OF TECHNOLOGY
(AICTE Approved, VTU Affiliated and NAAC 'A' Accredited)
(UG Programs - CSE, ECE, ISE, EIE and EEE have been Accredited by NBA for the Academic Years 2018-19, 2019-20, 2020-2021 and 2021-2022)
Channasandra,Dr.Vishnuvardhan Road,Bengaluru-560098

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
(Accredited by NBA for the acadamic Years 2018-19,2019-20,2020-21 and 2021-2022)

ESTD 2001

DECLARATION

I, **Aditya Venkata Sheshu** bearing the **USN:1RN18EC167**, pursuing Bachelor of Engineering in Electronics & Communication, RNS Institute of Technology, Bangalore. I hereby declare that the technical seminar titled, **"An AI-Based Visual Aid with Integrated Reading Assistant for the Blind"** has been presented under the supervision and guidance of **Dr.Nandini K.S**. Submitted as a partial fulfilment for the award of Bachelor of Engineering degree in **Electronics and Communication Engineering** from **Visvesvaraya Technological University, Belagavi** during the academic year 2021-22. I also declare that the technical seminar paper has not been submitted previously for the award of any degree or diploma to any institution.

Aditya Venkata Sheshu

1RN18EC167

Abstract

Facing the challenges that arise due to health related consequences through science and technology is crucial and is a field that needs attention to detail and conceptual practicality. One of such issues is blindness, that prevents a person from gaining the knowledge of the surrounding environment and make navigation,obstacle avoidance and reading tasks dangerous and challenging. Integrating several computing concepts and experimenting it to achieve assistance in these circumstances for the blind is essentially the crux of this research. Because of its low cost, compact size,and ease-of-integration, Raspberry Pi 3 Model B+ a version of the computer raspberry pi, being a series of small single-board computers, has been used to demonstrate the functionality of the proposed prototype performing functions that aid the blind in many scenarios.The system includes an integrated reading assistant, in the form of the image-to-text converter, followed by an auditory feedback. The entire setup is lightweight and portable and can be mounted onto a regular pair of eyeglasses, without any additional cost and complexity.The proposed work has taken With different approaches to make the system work in detailed environments wherein experiments are carried out with 60 completely blind individuals to evaluate the performance of the proposed device with respect to the traditional white cane used the most commonly in the world today by the people with visual challenges and eyesight damage. Results show that the proposed device, as compared with the white cane, enables greater accessibility, comfort, and ease of navigation for the visually impaired.

Acknowledgement

The joy and satisfaction that accompany the successful completion of any task would be incomplete without thanking those who made it possible. I consider myself proud to be a part of RNS Institute of Technology, the institution which molded me in all endeavors.

I express our gratitude to our Chairman late **Dr. R N Shetty** and **MD Satish R Shetty** for providing state of art facilities.

I would like to express our sincere thanks to **Dr. M K Venkatesha**, Principal and **Dr. Vipula Singh**, Professor and Head, Department of ECE, for their valuable guidance and encouragement throughout our program.

My sincere thanks and heartfelt gratitude to my internal guide **Dr.Nandini K.S**, Professor, for providing me an invaluable support throughout.

Aditya Venkata Sheshu

1RN18EC167

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

Loss of vision is one of the most common disabilities worldwide. With the number of people experiencing some kind of visual problems rising, it is crucial to understand the various difficulties it poses and find possible ways to tackle them. Several devices have been developed in order to aid visual illness and yield valuable assistance to the people experiencing the same. Apart from complete loss of vision or blindness, there also exists visual or vision impairment that is a reduction in vision, usually associated with age that cannot be corrected by prescription glasses, contact lenses, medicine or even surgery. It can range from low to severe. Visual impairment is a huge problem affecting a lot of people worldwide causing normal daily activities such as walking, reading or driving to become impossible or very challenging.

The domain of interest being a serious, health related one with high necessity for attention to detail, there exsits several areas of improvement in various devices and developing a balanced, efficient, low complex and affordable device is a wonderful challenge for any researcher developing computing devices in this domain. This paper addresses these issues and proposes a novel visual aid with valuable features that aid in the navigation, comfortability and safety of the user.

1.1 Challenges and factors affecting blind navigation

There is a significant amount of unnoticed or unnatended issues that blind navigation posesses which become a serious matter to conisder especially when designing elctronic equipment to aid these issues. There are many optical conditions that can cause people to go blind, and there are also many conditions that can reduce your vision to the point where you are unable to see effectively. Those who are unable to see well but can still make out objects and shapes are said to have partial blindness or low vision. People who are diagnosed with partial blindness have vision lower than 20/70. When someone reaches this point, there are few things that can be done to improve their vision back to what it once was.Partially blind people experience a cloudy vision, seeing only shadows, and suffer from poor night vision or tunnel vision. A completely blind person, on the other hand, has no vision at all.Recent statistics from the World

Health Organization estimate the number of visually impaired or blind people to be about 2.2 billion with statistics regarding the affection of visual issues continuing to grow.

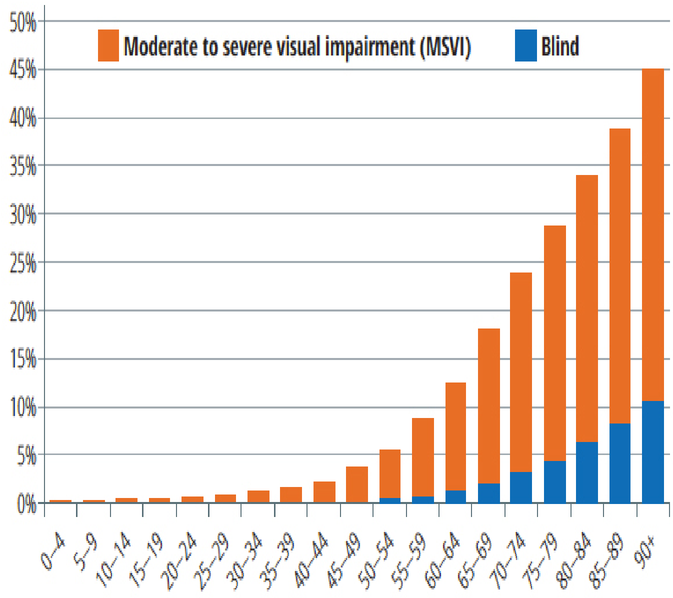

Figure 1.1: Impaired population

Figure 1.1 represents the percentage of population affected by severe visual disabilities as well as blindness as a function of their age. Whilst it is encouraging to see a continued decrease in the proportion of those suffering from visual impairment, there is still much work to be done to reduce the estimated 1 billion cases of visual impairment that could be prevented.Increasing life expectancy and a continued rise in the global population, together with poor access to health care in some low income countries, means the overall numbers of blind and visually impaired people continues to increase.

There are many ocular diseases which can reduce your ability to see normally. The most common causes of partial or complete blindness are-

- An injury to the eye- If someone is hit directly in the eye or has chemicals splashed in them, then they can develop partial blindness.
- Glaucoma- This is a condition that is characterized by increased ocular pressure and results in damage to the optic nerves. It will eventually lead to complete blindness, but most people have it treated before it gets to that point.
- Macular degeneration- This is a condition where the macula, or center portion of the retina, begins to deteriorate. It results in blind spots or partial blindness.

- Cataracts- When a cataract develops over the lens of someone's eye, it may cause them to experience partial blindness. This condition is treatable with surgical correction, though.

Considering blind people almost always travel independently, another important point to note is that perception unlocks the mind for other actions and in many cases visual perception comes first. Since eyes and the brain act in cohesion with each other there is inevitably going to be a significant delay in response of the individual when he or she encounters anything in the environment making responsivity vastly different from that of a normal person. This is an unnoticed , dangerous issue since it poses a great threat in cases of emergencies and even challenges normal well being.Blindness not only poses a difficulty in perceiving the environment but also a challenge to understand assisting mechanisms since it is significantly different and challenging to assist a person with these visual difficulties than it is for someone without it. By comparison, sighted people solve these problems visually in a more automatic, less cognitively demanding way. In other words, vision-based navigation is more of a perceptual process, whereas blind navigation is more of an effortful endeavor requiring the use of cognitive and attentional resources.

Vision also affords access to many orienting cues in the environment. For instance, use of local landmarks such as street signs or colorful murals and global landmarks such as tall buildings or mountain ranges can aid spatial updating and determination of location. Since access to this type of environmental information is difficult from non-visual modalities, blind wayfinders must rely on other cues for orientation which are often ambiguous and unreliable Most sighted people have never considered how they avoid obstacles, walk a straight line, or recognize landmarks. It is not something they consciously learned; it's just something they do. By contrast, the majority of blind people who are competent, independent travelers have had specific training to acquire these skills. This is called orientation and mobility (OM) training.

Electronic devices such as phones would not come in handy taking the task of assistance to another level of difficulty .Two of the biggest challenges to independence for blind individuals are difficulties in accessing printed material and the stressors associated with safe and efficient navigation.Access to printed documents has been greatly improved by the development and proliferation of adaptive technologies such as screen-reading programs, optical character recognition software, text-to-speech engines, and electronic Braille displays. By contrast, difficulty accessing room numbers, street signs, store names, bus numbers, maps, and other printed information related to navigation remains a major challenge for blind travel. Therefore it is essential to develop technology that addresses these crucial constraints and covers a broad spectrum of countering challenges thereby assists the people in need in the most efficient and best possible manner.

1.2 Some standard visual aids and their drawbacks

A white cane is used traditionally by the blind people to help them navigate their surroundings, although use of the white cane does not provide information for moving obstacles that are approaching from a distance. Moreover, white canes are unable to detect raised obstacles that are above the knee level. Trained guide dogs are another option that can assist the blind. However, trained dogs are expensive and not readily available. Recent studies have proposed several types of wearable or hand-held electronic travel aids (ETAs). Most of these devices integrate various sensors to map the surroundings and provide voice or sound alarms through headphones. The quality of the auditory signal, delivered in real-time, affects the reliability of these gadgets. Many ETAs, currently available in the market, do not include a real-time reading assistant and suffer from a poor user interface, high cost, limited portability, and lack of hands-free access. These devices are, therefore, not widely popular among the blind and require further improvement in design, performance, and reliability for use in both indoor and outdoor settings. A growing number of special devices are available for use by blind persons. They vary in cost from only a few dollars to $30,000 for a single device. Seldom does a device by itself make the difference in whether or not a blind person can do a job. Devices do, however, provide added independence and flexibility to blind persons in numerous positions. All the systems, services, devices and appliances that are used by disabled people to help in their daily lives, make their activities easier, and provide a safe mobility are included under one umbrella term: assistive technology . In the 1960s, assistive technology was introduced to solve the daily problems which are related to information transmission navigation and orientation aids which are related to mobility assistance.

Most special techniques of the blind require no special devices, and sometimes a simple, homemade device is as good as an expensive one. Nevertheless, blind job applicants and prospective employers of blind persons can benefit from knowledge of the devices available. This article is not meant to be an extensive list of all specialized equipment for the blind. Rather, it is intended as a sampling.

1.3 Proposed work

In this article and research work a novel visual aid system for completely blind individuals is proposed. The unique features, which define the novelty of the proposed design, include the following.

- Hands free, wearable, low power, and compact design mountable on a pair of eyeglasses, for the indoor and outdoor navigation with an integrated reading assistant
- Complex algorithm processing with a low-end configuration.
- Real-time, camera-based, accurate distance measurement, which simplifies the design and lowers the cost by reducing the number of required sensors.

The proposed setup, in its current form, can detect both stationary and moving objects in real time and provide auditory feedback to the blind. In addition, the device comes with an in-built reading assistant that is capable of reading text from any document.

Chapter 2

Relevant Work

2.1 Types of Visual Aids

The electronic aids for the visually impaired can be categorized into three different subcategories, ETAs, electronic orientation aids, and positional locator devices. ETAs provide object detection, warning, and avoidance for safe navigation [1],[2]. ETAs work in few steps; sensors are used to collect data from the environment, which are then processed through a computing device to detect an obstacle or object and give the user a feedback corresponding to the identified object. The ultrasonic sensors can detect an object within 300 cm by generating a 40 kHz signal and receiving reflected echo from the object in front of it. The distance is calculated based on the pulse count and time-of-flight (TOF). Smart glasses [2], [3] and boots [4], mounted with ultrasonic sensors, have already been proposed as an aid to the visually impaired. A new approach by Katzschmann et al.[5] uses an array of infrared TOF distance sensors facing in different directions. Villanueva and Farcy [6] combine a white cane with near-IR LED and a photodiode to emit and detect the IR pulses reflected from obstacles, respectively. Cameras [7],[8] and binocular vision sensors [9] have also been used to capture the visual data for the blind.

2.2 Types of Algorithms and different technical approaches

Different devices and techniques are used for processing the collected data. Raspberry Pi 3 Model B+, with open computer vision (OpenCV) software, has been used to process the images captured from the camera [10]. Platforms such as Google tango have also been used. A field-programmable gate array is also another option to process the gathered data [11]. The preprocessing of captured images is done to reduce noise and distortion. Images are manually processed by using the Gaussian filter, gray scale conversion, binary image conversion, edge detection, and cropping [12]. The processed image is then fed to the Tesseract optical character recognition (OCR) engine to extract the text from it [13]. The stereo image quality assessment [14] employs a novel technique to select the best image, out of many. The best image is then fed to a convolutional neural network (CNN), which is trained on big data and runs on a cloud device. The audio feedback in most devices described above is provided through a headset or a speaker. The audio is either a synthetic voice from the text-to-speech synthesis system [15] or a voice user interface [16] generating a beep sound. Vibrations and tactile feedback are also used in some systems.Andò et al. [17] introduced a haptic device, similar to the white cane, with an embedded smart sensing strategy and an active handle, which detects an obstacle and produces vibration mimicking a real sensation on the cane handle. Another traditional white cane like system, guide cane, rolls on wheels and has steering servo motors to guide the wheels by sensing the obstacles from ultrasonic sensors. The backdrop of this system is that the user must always hold the device by their hand, whereas, many systems, which provide a hands-free experience, are readily available. NavGuide and NavCane [18] are assistive devices that use multiple sensors to detect obstacles up to the knee level. Both NavGuide and NavCane are equipped with wet floor sensors. NavCane can be integrated into the white cane systems and offers a global positioning system (GPS) with a mobile communication module.

A context-aware navigation framework is demonstrated by Xiao et al, which provides visual cues and distance sensing along with location-context information, using GPS. The platform can also access geographic information systems, transportation databases, and social media with the help of Wi-Fi communication through the Internet. Lan et al. [19] proposed a smart glass system, which can detect and recognize road signs, such as public toilets, restaurants, and bus stops, in the cities in real time. This system is lightweight, portable, and flexible. However, reading out the road signage alone may not carry enough information for a blind user to be comfortable in an outdoor environment. Since public signs can be different in different cities, therefore, if a sign is not registered in the database of the system, the system will not be able to recognize it. Hoang et al. designed an assistive system using mobile Kinect and a matrix of electrodes for obstacle detection and warning. However, the system has a complex configuration and an uncomfortable setup because the sensors are always placed inside the mouth during navigation. Furthermore, it is expensive and has less portability. Islam et al. [20] presented a comprehensive review of sensorbased walking assistants for the visually impaired. The authors identified key features that are essential for an ideal walking assistant. These include low cost, simple, and lightweight design with a reliable indoor and outdoor coverage. Based on the feedback from several blind user groups, software developers, and engineers, Dakopoulos and Bourbakis also identified 14 structural and operational features that describe an ideal ETA for the blind.

2.3 Key drawbacks and areas of improvement

Many navigational technologies have been developed throughout the years, but few are still in existence. Part of the reason may be due to a disconnect between engineering factors and a device's perceptual and functional utility; that is, a device may work well in theory but be too difficult or cumbersome in practice to be adopted by the intended user. Four important factors should be considered when discussing the design and implementation of technology for blind navigation.[21]

Therefore,Despite numerous efforts, many existing systems do not incorporate all features to the same satisfactory level and are often limited by cost and complexity. The main contribution of the presented work was to build a simple, low cost, portable, and hands-free ETA prototype for the blind, with text-to-speech conversion capabilities for basic, everyday indoor and outdoor use. While, the proposed system, in its present form, lacks advanced features, such as the detection of wet floors and ascending staircases, reading of road signs, use of GPS, or mobile communication module, the flexible design presents opportunities for future improvements and enhancements.

Chapter 3

Proposed Design

The proposed design with the implemented hardware and software aspects is illustrated in this chapter. The setup is mounted on a pair of eyeglasses and can provide real-time auditory feedback to the user through a headphone. Camera and sensors are used for distance measurement between the obstacle and the user. The schematic view in Figure 3.1 presents the hardware setup of the proposed device.

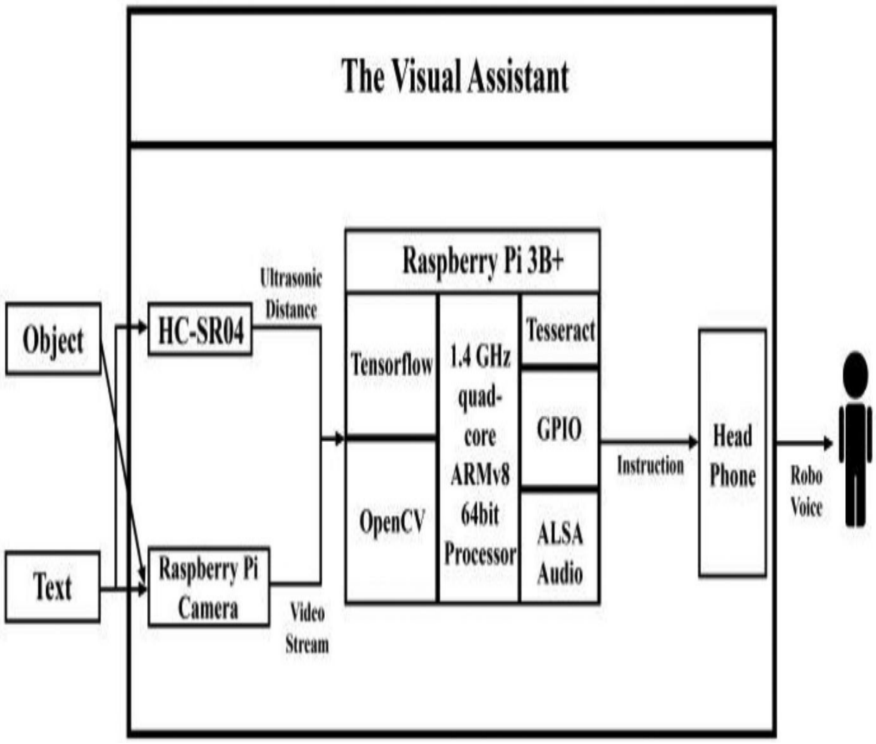

Figure 3.1: Hardware configuration of the proposed system

Raspberry Pi 3 Model B+ which can be seen in the above figure, was chosen as the functional device owing to its low cost and high portability. Also, unlike many existing systems, it offers a multiprocessing capability. It also offers a 1.2 GHz quad-core ARM Cortex A53 processor that can output a video at a full 1080p resolution with desired details and accuracy.

As illustrated in Figure 3.1 the visual assistant takes the image as inputs,processes it through the Raspberry pi processor and gives the audio feedback through a headphone.

3.1 Implemented techniques

For the key features of this device such as object detection and auditory feedback , multiple techniques have been adopted that take into consideration the fundamental concepts implemented in this device such as optical character recognition, machine learning through CNN and computer vision, text to speech etc. The techniques and their specifications implemented are as follows -

- TensorFlow object detection application programming interface (API), frameworks, and libraries, such as OpenCV and Haar cascade classifier, are used for detecting faces and eyes and implement distance measurement.
- Tesseract, which is a free OCR engine, for various operating systems, is used to extract text from an image
- eSpeak, which is a compact open-source speech synthesizer (text-to-speech), is used for auditory feedback for object type and distance between the object and the user
- For obstacles within 40–45 inches of the user, the ultrasonic transducer (HC-SR04) sets off a voice alarm, while the eSpeak speech synthesizer uses audio feedback to inform the user about his or her distance from the obstacle, thereby, alerting the blind person and avoiding any potential accident

The proposed prototype is illustrated in fig 3.2 where the setup is mounted on a pair of conventional eyeglasses incolving a front facing camera module and sensors interfaced with the Raspberry Pi

3.1.1 TensorFlow

Tensorflow is an end to end open source platform developped by google primarily for deeplearning as well as machine learning applications . It provides a collection of workflows to develop and train models using Python or JavaScript, and to easily deploy in the cloud, on-prem, in the browser, or on-device no matter what language you use. The tf. data API enables you to build complex input pipelines from simple, reusable pieces. TensorFlow accepts data in the form of multi-dimensional arrays of higher dimensions called tensors. Multi-dimensional arrays are very handy in handling large amounts of data.

Figure 3.2: Proposed prototype

TensorFlow works on the basis of data flow graphs that have nodes and edges. As the execution mechanism is in the form of graphs, it is much easier to execute TensorFlow code in a distributed manner across a cluster of computers while using GPUs.

3.1.2 Open CV and Haar cascade classifier

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms.

These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc.

Haar cascade first introduced by Viola and Jones in their seminal 2001 publication, Rapid Object Detection using a Boosted Cascade of Simple Features, are arguably OpenCV's most popular object detection algorithm.One of the primary benefits of Haar cascades is that they are just so fast — it's hard to beat their speed.

3.1.3 Tesseract

Tesseract — is an optical character recognition engine with open-source code, this is the most popular and qualitative OCR-library.

OCR uses artificial intelligence for text search and its recognition on images.

Tesseract is finding templates in pixels, letters, words and sentences. It uses two-step approach that calls adaptive recognition. It requires one data stage for character recognition, then the second stage to fulfil any letters, it wasn't insured in, by letters that can match the word or sentence context.Tesseract is based on traditional computer vision algorithms.

3.1.4 eSpeak

Espeak is a free text to speech synthesizer. It was developed by Jonathan Duddington and its development has stopped in 2015. In 2015 Reece Dunn has taken a copy of espeak and together with a group of developers they maintain and actualize their version of espeak which they call ”eSpeak NG”. eSpeak NG uses formant synthesis. Currently it supports 130 languages with varying quality of the voices.

3.2 System Workflow

3.2.1 Data Acquisition

Figure 3.3 shows how the Raspberry Pi 3Model B+is connected to other components in the system. Data are acquired in two ways. Information that have red, green, and blue (RGB) data were acquired using the Raspberry Pi camera module V2, which has a high quality, 8-megapixel Sony IMX219 image sensor.

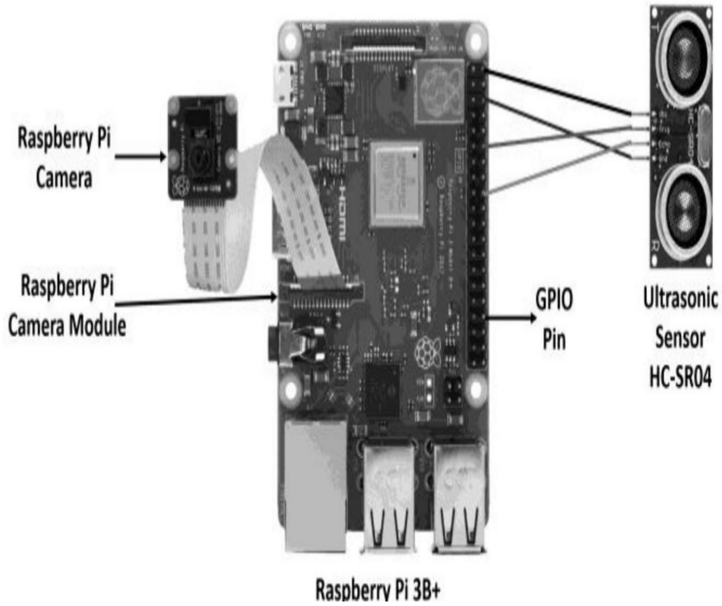

Figure 3.3: Basic hardware setup with camera and ultrasonic sensors

The camera sensor, featuring a fixed focus lens, has been custom designed to fit onboard into Raspberry Pi. It can capture 3280 pixels $\times$ 2464 pixels static images and supports 1080p, 720p, and 640 pixels $\times$ 480 pixels video. It is attached to the Pi module through small sockets, using the dedicated camera serial interface. The RGB data are retrieved by our program, in real time, and can recognize objects from every video frame that is already known to the system. To acquire data from the ultrasonic rangefinder, HC-SR04 was mounted below the camera, as shown in Figure 3.3. The working of a standard ultrasonic sensor can be summarized as follows -

Ultrasonic sensors work by sending out a sound wave at a frequency above the range of human hearing. The transducer of the sensor acts as a microphone to receive and send the ultrasonic sound. The sensor head emits an ultrasonic wave and receives the wave reflected back from the target. Ultrasonic Sensors measure the distance to the target by measuring the time between the emission and reception.

There are four pins on the ultrasound module that were connected to the Raspberry Pi's GPIO ports. VCC was connected to pin 2 (VCC), GND to pin 6 (GND), TRIG to pin 12 (GPIO18), and the ECHO to pin 18 (GPIO24). The ultrasonic sensor output (ECHO) will always give output LOW (0 V), unless it has been triggered, in which case, it will give output HIGH (5 V). Therefore, one GPIO pin was set as an output to trigger the sensor and one as an input to detect the ECHO voltage change. However, this HC-SR04 sensor requires a short 10 s pulse to trigger the module. This causes the sensor to start generating eight ultrasound bursts, at 4 kHz, to obtain an echo response. So, to create the trigger pulse, the trigger pin is set HIGH for 10 s and then set to LOW again. The sensor sets ECHO to HIGH for the time it takes for the pulse to travel the distance and the reflected signal to travel back. Once a signal is received, the value changes from LOW (0) to HIGH (1) and remains HIGH for the duration of the echo pulse. From the difference between the two recorded time stamps, the distance between the ultrasound source and the reflecting object can be calculated. The speed of sound depends on the medium it is traveling through and the temperature of that medium. In our proposed system, 343 m/s, which is the speed of sound at sea level, has been used.

3.2.2 Feature Extraction

3.2.2.1 Tensorflow and its object detection API

The TensorFlow object detection API is used to extract features (objects) from images captured from the live video stream. The TensorFlow object detection API is an open-source framework, built on the top of TensorFlow, which is easy to integrate, train, and create models that perform well in different scenarios. TensorFlow represents deep learning networks as the core of the object detection computations. The foundation of TensorFlow is the graph object, which contains a network of nodes. GraphDef objects can be created by the ProtoBuf library to save the network. For the proposed design, a pretrained model, called single-shot detection (SSD)Lite-MobileNet, from the TensorFlow detection model zoo, has been used. The model zoo is Google's collection of pretrained object detection models trained on different datasets, such as the common objects in context (COCO) dataset. This model was particularly chosen for the proposed prototype because it does not require high-end processing capabilities, making it compatible with the low processing power of the Raspberry Pi. To recognize objects from the live video stream, no further training is required since the models have already been trained on different types of objects.

3.2.3 Fundamentals of Convolutional Neural Networks for object recognition

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision.

The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image Video recognition, Image Analysis Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, primarily over one particular algorithm — a Convolutional Neural Network.

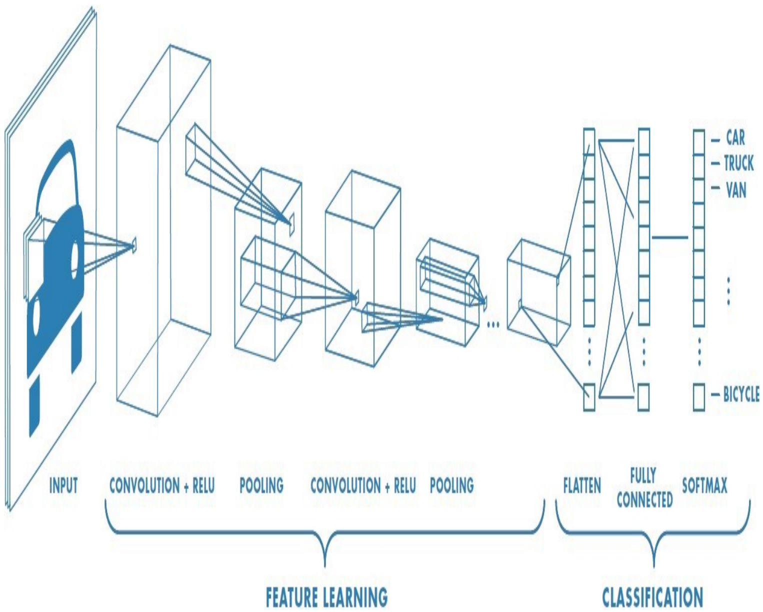

Figure 3.4: CNN layers

Figure 3.4 depicts a convolutional neural network (ConvNet/CNN) which is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

3.2.3.1 Single Shot Detection - SSD

SSD has two components: a backbone model and SSD head. Backbone model usually is a pre-trained image classification network as a feature extractor. This is typically a network like ResNet trained on ImageNet from which the final fully connected classification layer has been removed. We are thus left with a deep neural network that is able to extract semantic meaning from the input image while preserving the spatial structure of the image albeit at a lower resolution. For ResNet34, the backbone results in a 256 7x7 feature maps for an input image.The SSD head is just one or more convolutional layers added to this backbone and the outputs are interpreted as the bounding boxes and classes of objects in the spatial location of the final layers activations.

3.2.3.2 Why SSD?

The Single shot detection (SSD) approach uses a small convolutional filter to predict object categories and these filters are used to multiply feature maps to perform detection at multiple scales. This results in high-accuracy detection even in low-resolution images. In order to achieve high detection accuracy, the SSD model produces predictions at different scales from the feature maps of different scales and explicitly separates predictions by aspect ratio. These techniques result in simple end-to-end training and high accuracy even on input images of low resolutions.

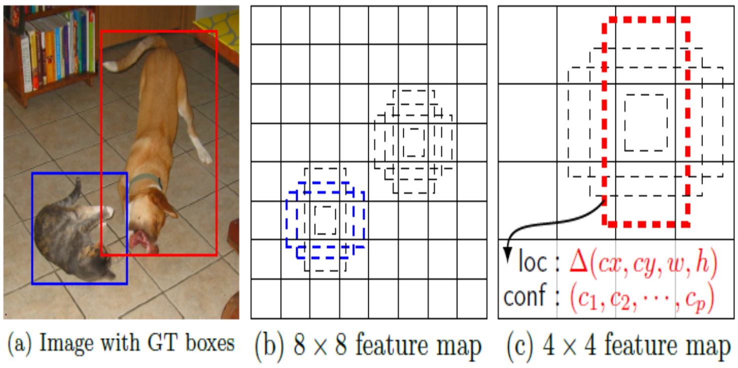

Figure 3.5: Single shot detection

Figure 3.5 depics gow an object is detecte using SSD. The bounding boxes are formed as shown in figure 3.5(a) based on the graphical location of the objects learned by the CNN where in a feature map is employed as shown in figure 3.5(b). A furhter dimensionally reduced feature map is then created as shown in figure 3.5(c) to infer mathematical features such as the intersection over union.

The SSD is based on the use of convolutional networks that produce multiple bounding boxes of various fixed sizes and scores the presence of the object class instance in those boxes, followed by a non-maximum suppression step to produce the final detections. The SSD model works as follows, each input image is divided into grids of various sizes and at each grid, the detection is performed for different classes and different aspect ratios. And a score is assigned to each of these grids that says how well an object matches in that particular grid. And non maximum supression is applied to get the final detection from the set of overlapping detections. This is the basic idea behind the SSD model.

The SSD model is therefore one of the fastest and efficient object detection models for multiple categories. And it has also opened new doors in the domain of object detection.

3.2.4 Object detection

The human brain focuses on the region of interests and salient objects, recognizing the most important and informative parts of the image [29]. By extracting these visual attributes [30], the deep learning techniques can mimic human brains and can detect salient objects from images, video frames, and even from optical remote sensing. A pixelwise and nonparametric moving object detection method can extract from the spatial and temporal features and detect moving objects with intricate background from the video frame. Many other techniques for object detection and tracking, from the video frame, such as the object-level RGB-D video segmentation, are also commonly used. For object detection, every object must be localized within a bounding box, in each frame of a video input. A "region proposal system" or Regions + CNN (R-CNN) can be used, where, after the final convolutional layers, a regression layer is added to get a number that consists of four variables x0, y0, width, and height of the image. This process must train the support vector machine for each class, to classify between the object and background, while proposing the region in each image. In addition, a linear regression classifier needs to be trained, which will output some correction factor. To eliminate the unnecessary bounding boxes from each class, the intersection over union method must be applied to filter out the actual location of an object in each image. Methods used in faster R-CNN dedicatedly provide region proposals, followed by a high-quality classifier to classify these proposals . These methods are very accurate but come at a big computational cost. Furthermore, because of the low frame rate, these methods are not fit to be used on embedded devices.In SSDLite, MobileNetv2 was used as the backbone and has depthwise separable convolutions for the SSD layers. The SSDLite models make predictions on a fixed-sized grid. Each cell in this grid is responsible for detecting objects, in a location, from the original input image and produces two tensors as the outputs that contain the bounding box predictions for different classes. SSDLite has several different grids ranging in size from 19 × 19 to 1 × 1 cells. The number of bounding boxes per grid cell is 3, for the largest grid, and 6, for the others, making a total of 19 × 17 boxes.For object detection,MobileNetv2 is used as the base network, along with SSD since it is desirable to know both high-level as well as low-level features by reading the previous layers. as depicted in figure 3.6. Since object detection is more complicated than the classification, SSD adds many additional convolution layers on the top of the base network. To detect objects in live feeds, we used a Pi camera. Basically, our script sets paths to the model and label maps, loads the model into memory, initializes the Pi camera, and then begins performing object detection on each video frame from the Pi camera. Once the script initializes, which can take up to a maximum of 30 s, a live video stream will begin and common objects inside the view of the user will be identified. Next, a rectangle is drawn around the objects. With the SSDLite model and the Raspberry Pi 3 Model B+, a frame rate higher than 1 fps can be achieved, which is fast enough for most real-time object detection applications.

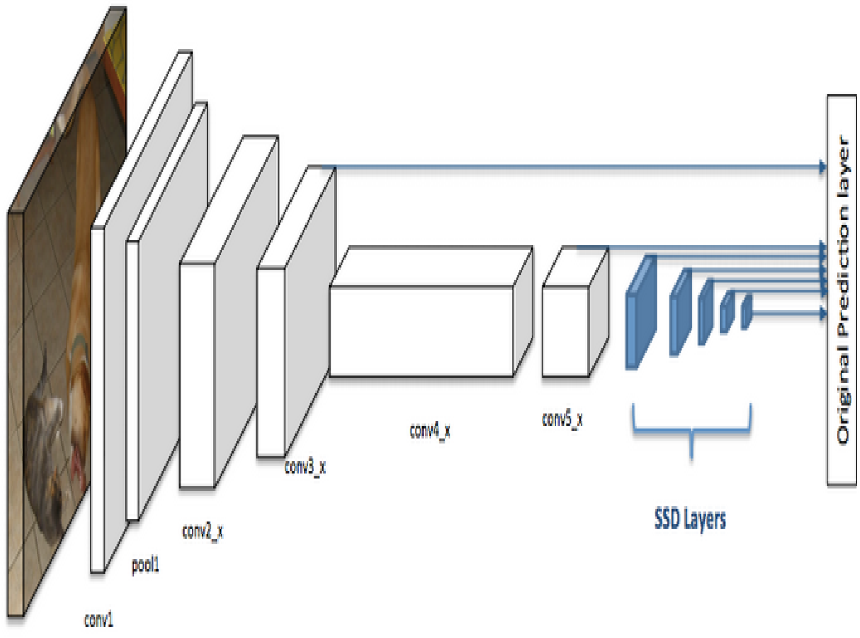

Figure 3.6: Single shot detection

3.2.5 Reading Assistant

The proposed system integrates an intelligent reader that will allow the user to read text from any document. An open-source library, Tesseract version-4, which includes a highly accurate deep learning-based model for text recognition, is used for the reader incorporating Optical character recognition (OCR) An OCR program extracts and repurposes data from scanned documents, camera images and image-only pdfs. OCR software singles out letters on the image, puts them into words and then puts the words into sentences, thus enabling access to and editing of the original content. An OCR system is made up of both hardware and software and works in three main steps as follows -

- **Image preprocessing** - In the first stage, the technology converts the document's physical shape into a picture, such as a record picture. The purpose of this stage is for the machine's representation to be precise while also removing any undesired aberrations. The concept is subsequently transformed to a black and white rendition, evaluated for bright vs. dark regions (characters). The image is then segmented into individual pieces, such as spreadsheets, text, or inset graphics, using an OCR system.
- **AI Character recognition**- AI analyzes the image's dark portions to recognize characters and numerals. Typically, AI uses pattern recognition and feature recognition.
- **Post Processing** - AI corrects flaws in the final file during Post-Processing. One approach is to teach the AI a glossary of terms that will appear in the paper. Then, limit the AI's output to those words/formats to verify that no interpretations are beyond the vocabulary.

Tesseract has unicode (UTF-8) support and can recognize many languages along with various output formats: plain-text, hocr (HTML), pdf, tsv, and invisible-text-only pdf. The underlying engine uses a long short-term memory (LSTM) network. LSTM is part of a recurrent neural network, which is a combination of some unfolded layers that use cell states in each time steps to predict letters from an image. The captured image is divided into horizontal boxes, and in each time step, the horizontal boxes are being analyzed with the ground truth value to predict the output letter. LSTM uses gate layers to update the cell state, at each time step, by using several activation functions. Therefore, the time required to recognize texts can be optimized.

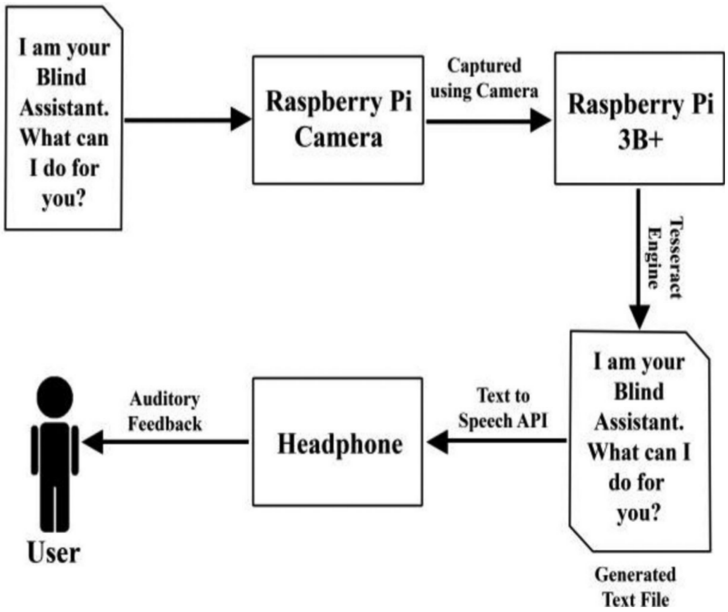

Figure 3.7: Workflow of reading asssitant

Figure 3.7 shows the working principle of the reading assistant. An image is captured from the live video feed without interrupting the object detection process. In the background, Tesseract API will extract the texts from the image and save them in a temporary text file. Then it reads out the text from the text file using the textto-speech engine eSpeak. The accuracy of the Tesseract OCR engine depends on ambient lighting and background and usually works well in the white background and brightly illuminated places. Therefore, the raspberry Pi gets a single frame from the camera module and runs through the tesseract OCR engine. The test output is then converted to the audio.

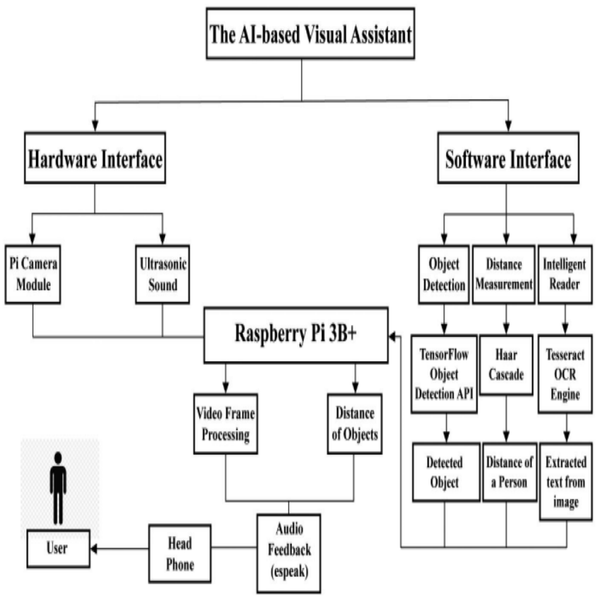

Figure 3.8: Complete workflow of the proposed system

After understanding the key steps and the workflow of the essential features implemented , the overall workflow of the proposed design can be depicted as shown in figure 3.8 where in the AI based visual assistant proposed is comprising of hardwre and software interface having their respective functions and features which integrate with the raspberry pi

Every frame of the video is being processed through a standard convolutional network to build a feature representation of the original image or the frame. This backbone network is then pretrained on Image-Net in the SSD model, as an image classifier, to learn how to extract features from an image using SSD. Then, the model manually defines a collection of aspect ratios for bounding boxes, at each grid cell location. For each bounding box, it predicts the offsets for the bounding box coordinates and dimensions. Along with this, the distance measurement is processed using both the depth information and the ultrasonic sensor. In addition, the reading assistant works without interrupting any of the prior processes. All the three features run in the software interface with the help of the modules from the hardware interface.

Chapter 4

System Evaluation and Experiments

This chapter illustrates the conditions and cases of evaluation of the proposed setup, depicting how it works in testing environments

4.1 Evaluation of implemented features

4.1.1 Evaluation of object detection

The proposed model (SSDLite) is pretrained on the Image-Net dataset for the image classification. It draws a bounding box on an object and tries to predict the object type based on the trained data from the network. It directly predicts the probability that each class is present in each bounding box using the softmax activation function and cross entropy loss function. The model also has a background object class when it is classifying different objects. However, there can be a large number of bounding boxes detected in one frame with only background classes. To avoid this problem, the model uses hard negative mining to sample negative predictions or downsampling the convolutional feature maps to filter out the extra bounding boxes.

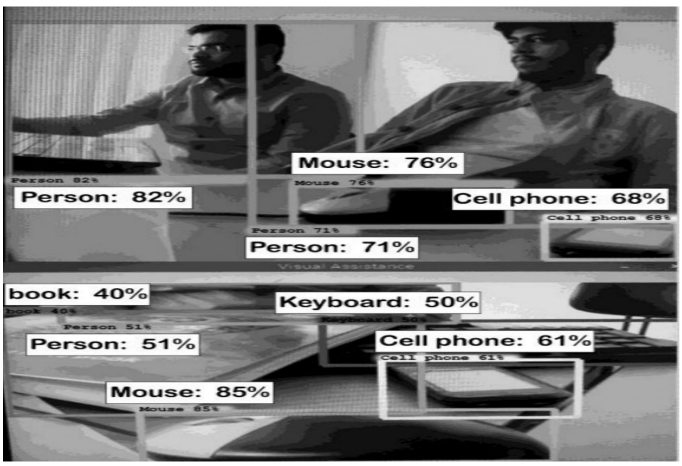

Figure 4.1: Detecting multiple objects with various confidence levels form a single frame

As shown in Figure 4.1, The model can easily identify up to four or five objects, simultaneously, from a single video frame. The confidence level indicates the percentage of times the system can detect an object without any failure.

Table 4.1: Detecting multiple objects with various confidence levels from a single frame

Actual Objects	predicted objects	Failure Case(s)
Person	Person	None
Mouse	Mouse	None
Person	Person	None
Notebook	Notebook	None
Cellphone	Cellphone	None
Person,chair,mouse	Person,chair,mouse	None
Cellphone,notebook	Cellphone,notebook,laptop	Laptop
Notebook,Person	Notebook,Person	None
Pen,Mouse,Keyboard	Pen,Mouse,Keyboard	None
Bottle, Cellphone	Bottle,Cellphone	None

Table 4.1 summarizes the results from single and multiple object detection, for 10 unique cases, consisting of either a single item or a combination of items, commonly found in indoor and outdoor setups.

The system can identify single items with near 100 percent accuracy with zero failure cases. Where multiple objects are in the frame, the proposed system can recognize each known object within the view. For any object situated in the range of 15–20 m from the user, the object can be recognized with at least 80 percent accuracy.Whenever there are multiple objects, in front of the user, the system will generate feedback for the object, which is closest to the user. An object with a higher ground truth value has a higher priority. As indicated in the paper work, the pretrained model, however, is subject to failure due to variation in the shape and color of the object as well as changes in ambient lighting conditions.

4.1.2 Evaluation of distance measurement

Figure 4.3 shows the device measuring the distance between a computer mouse and the blind person using the ultrasonic sensor. If the distance measured from the sensor is less than 40 cm, the user will get a voice alert saying that the object is within 40 cm. The sensor can measure distances within a range of 2–120 cm by sonar waves. Fig. 10 demonstrates the case where the combination of camera and ultrasonic sensor is used to identify a person's face and determine how far the person is from the blind user. The integration of the camera with the ultrasonic sensor, therefore, allows simultaneous object detection and distance measurement, which adds novelty to our proposed design. We have used the Haar cascade algorithm to detect face from a single video frame. It can also be modified and used for other objects. The bounding boxes, which appear while recognizing an object, consist of a rectangle. The width w, height h, and the coordinates of the rectangular box (x0, y0) can be adjusted as required.

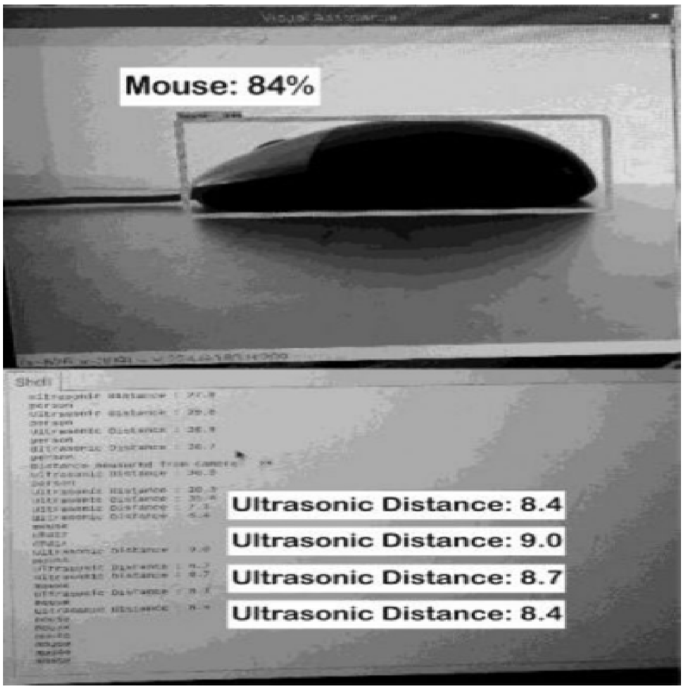

Figure 4.2: Detecting multiple objects with various confidence levels form a single frame

Figure 4.2 demonstrates how the distance between the object and the blind user can be simultaneously measured by both the camera and the ultrasonic sensor. The dotted line (6 m) represents the distance measured by the camera and the solid line (5.6 m) represents the distance calculated from the ultrasonic sensor. Width w and height h of the bounding box are defined in the .xml file with feature vectors, and they vary depending on the distance between the camera and the object. In addition to the camera, the use of the ultrasonic sensor makes object detection more reliable.

The following equation, which can be derived by considering the formation of image, as light passes through the camera lens, is used to calculate the distance between the object and user: distance(inches) = $(2*3.14*180) \div (w + h*360)*1000+3$

4.1.2.1 Evaluation of reading assistant

The integrated reading assistant in our prototype is tested under different ambient lighting conditions for various combinations of text size, font, color, and background.

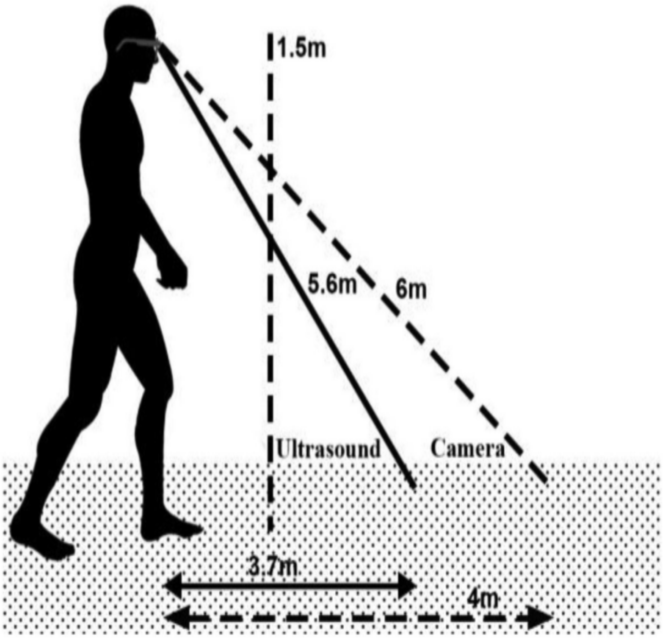

Figure 4.3: Demonstration of distance measurement mechanism using camera and ultrasonic sensor

The OCR engine performs better in an environment with more light as it can easily extract the text from the captured image. While comparing text with different colored background, it has been shown that a well-illuminated background yields better performance for the reading assistant.

Table 4.2: Performance of the reading assistants

Text	Text color	Paper color	Ambient Lighting	Performance
what can i do for you ?	Black	White	Bright,Slight	Reads accurately
what can i do for you ?	Black	White	Dark	Does not read accurately
I am doing well	Green	white	Bright,slight	reads accurately
I am doing well	Green	white	Dark	Does not read accurately

As given in Table 4.2 , the performance of the reading assistant is tested under three different illuminations: bright, slightly dark, and dark, using the green and black-colored texts, written on white pages. When the text color is black, the device performed accurately in bright and even in a slightly dark environment but under the dark condition, it failed to read the full sentence. For the green-colored text, the reading assistant had no issues in the brightly lit environment but failed to perform accurately in slightly dark and dark conditions.

4.2 Experimental Setup

The usability and performance of the prototype device is primarily tested in controlled indoor settings that mimic real-life scenarios. Although the proposed device functioned well in a typical outdoor setting, as shown in Figure 4.4, the systematic study and conclusions, discussed in the following sections, are based on the indoor setup only.

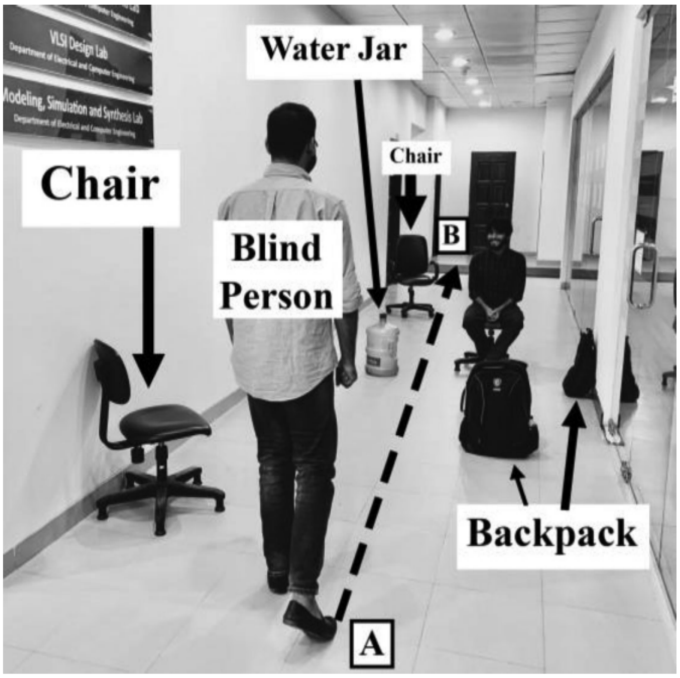

Figure 4.4: Testing the prototype in an indoor setting

A total of 60 completely blind individuals (male: 30 and female: 30) volunteered to participate in the controlled experiments. The influence of gender or age on the proposed system is beyond the scope of our current work and has, therefore, not been investigated here. However, since the gender-based blindness studies have shown blindness to be more prevalent among women than in men, it is important to have female blind users represented in significant numbers, in the testing and evaluation of any visual aid. Dividing 60 human samples into 30 males and 30 females to study separately, could, therefore, prove useful to conduct the gender-based evaluation study of the proposed system in future endeavors.

A short training session, over a period of 2 hours, is conducted to familiarize the blind participants with the prototype device. During the training, the evaluation and scoring criterion were discussed in detail. The indoor environment, as shown in Figure 4.4, consisted of six stationary obstacles of different heights and a moving person The position of the stationary objects was shuffled to create ten different indoor test setups, which were assigned, in random, to each user. A blind individual walks from point A to point B, along the path AB (15 m in length), first with our proposed blind assistant, mounted on a pair of eyeglasses, and then with a traditional white cane. For both the device and the white cane, the time taken to complete the walk was recorded for each participant. Based on the time, the corresponding velocity for each participant is calculated.

Chapter 5

Results and Conclusion

The performed experiments to evaluate the setup as illustrated in the previous chapter needs a metric and organized analysis as it is important to study the behaviour of the developed device in cohesion to the users resonse. The methodology of assesment criteria to determine the performance of the device is as follows -
Blind participants were instructed to rate the device based on its comfort level or ease-of-use, mobility, and preference compared with the more commonly used traditional white cane. Ratings were done on a scale of 0–5 and the user experiences for comfortability, mobility, and preference over the white cane are divided into the following three categories based on the scores:

- Worst (Score: 0-2)
- Moderate (score: 3)
- Good (Score: 4 and 5)

The preferability scores also refer to the likelihood that the user would recommend the device to someone else. For example, a score of 3 for preferability means that the user is only slightly impressed with the overall performance of the device, while a score of 1 means that the blind person highly discourages the use of the device. The accuracy of the reading assistant was also scored on a scale of 0–5, with 0 being the least accurate and 5 being the most. The total score, from each user, is calculated by summing the individual scores for comfort, mobility, preferability, and accuracy of the reading assistant. In the best-case scenario, each category gets a score of 5 with a total score of 20. Depending on the total score, the proposed blind assistant is labeled as "not helpful (total score: 0–8)," "helpful (total score: 9–15)," and "very helpful (total score: 16–20)." These labels were set after an extensive discussion with the blind participants prior to conducting the experiments. Also , it was surveyed that almost all the blind users were participating in such a study for the first time with no prior experience of using any form of ETA. Therefore, it was necessary to set a scoring and evaluation criterion that could be easily adopted without the need for advanced training and extensive guidelines.

The results obtained can essentially be depicted based on two chosen governing factors of performance assesment those being - **Velocity** and **User rating**

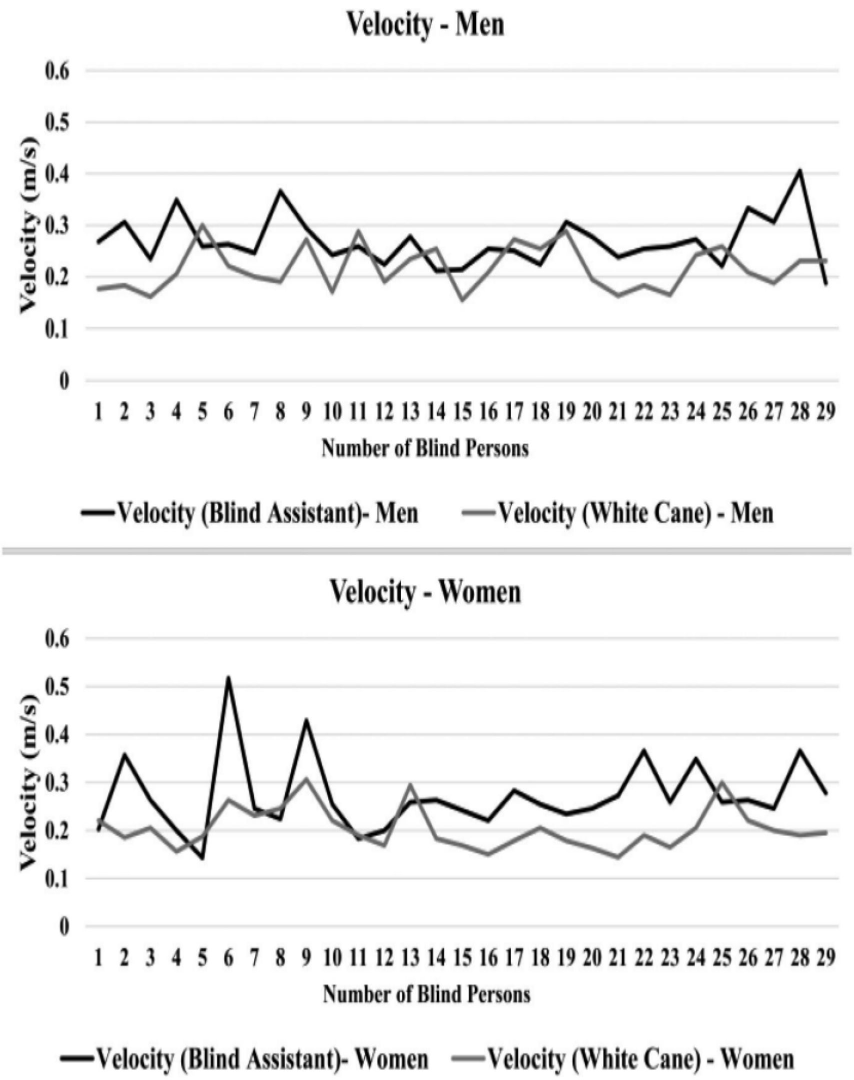

Figure 5.1: Velocity of blind participants walking from point A to B

The results therefore obtained are seen in figure 5.1

For each user, the speed achieved using the blind assistant and the white cane is plotted. The plots for male and female users are shown separately.

Table 5.1: Average velocity of blind participants

Gender	Average velocity of blind assistant (m/s)	Average velocity of white cane(m/s)
male	0.2627	0.2172
female	0.2690	0.2144

Table 5.1 lists the average velocity for 30 male and It is evident from the table that, on an average, the blind assistant provides slightly faster navigation than the white cane, for both the genders. To compare the performances between our proposed blind assistant and the white cane, a t-test is performed with a sample size of 60.

With a t-value equal to 4.9411, the two-tailed P value is less than 0.0001. Therefore, by the conventional criteria and at 95 percent confidence interval, the difference in velocity between the blind assistant and white cane can be considered statistically significant.

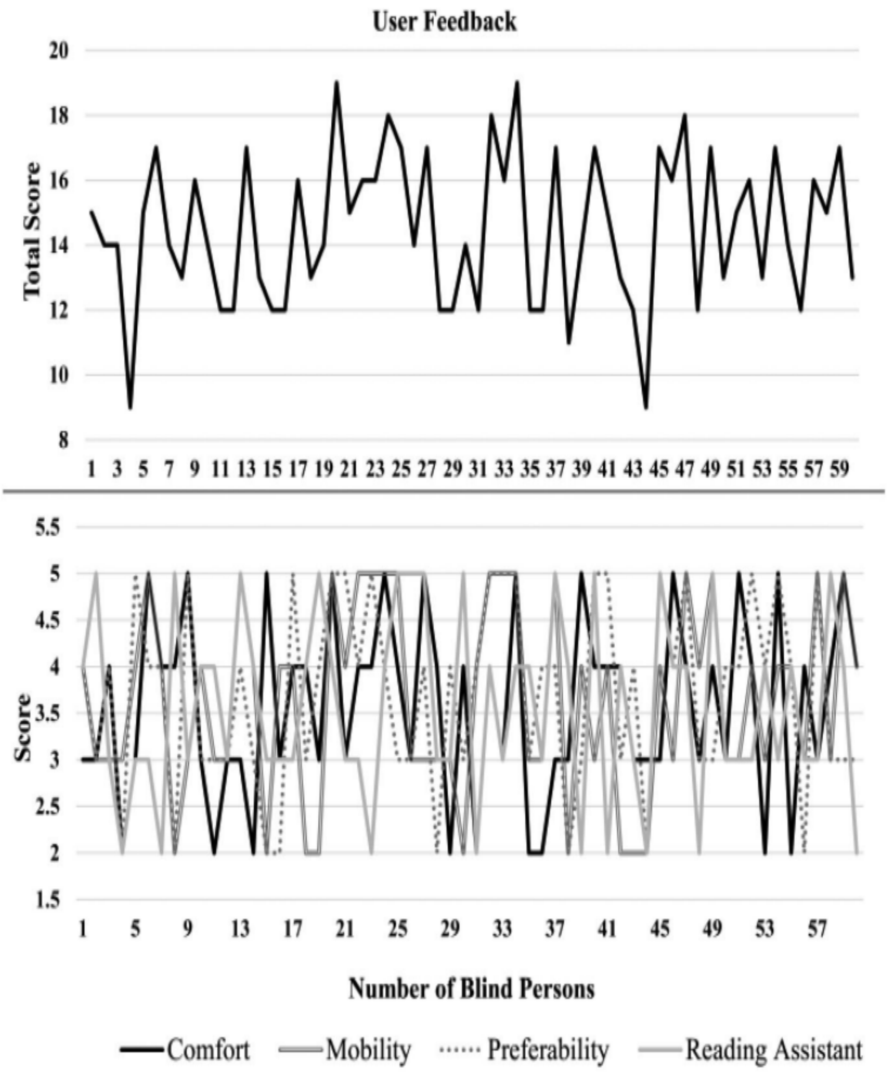

Figure 5.2: Velocity of blind participants walking from point A to B

The user ratings are plotted in Figure 5.2, which shows the individual scores for comfort, mobility, preference, and accuracy of the reading assistant, on a scale of 0–5, for each of the 60 users. In addition, the total score, rated on a scale of 0–20, is also shown. The average of all scores is 14.5, which deems our proposed device as “helpful” based on the criterion defined. The pretrained model, which was used, could be retrained with more objects for better performance.

5.1 Limitations

Considering the above implemented prototype in the proposed project work, the major limitations of the same include -

- Comfort level was compromised since it is a prototype and recieved comparatively less user rating
- The reading assistant performed well under brightly illuminated settings but one major limitation of the reading assistant, as pointed out by the users, is that it was unable to read texts containing tables and pictures.

5.2 Comparison with other alternatives

In order to compare the proposed prototype with its other visual aid alternatives, a cost analysis was done with similar state-of-the-art assistive navigation devices. Table compares the cost of the blind assistant with some of the existing platforms.

Table 5.2: Cost of proposed device versus existing visual aids

Device	Estimated cost
Proposed device	68
Lan et al. [26]	240
Jiang et al. [17]	97
Rajesh et al. [31]	70
White Cane	25

The total cost of making the proposed device is roughly US $68, whereas some existing devices, with a similar performance, appear more expensive. The service dogs, another viable alternative, can cost up to US $4000 and require high maintenance. Although the white canes are cheaper, they are unable to detect moving objects and do not include a reading assistant.

5.3 Conclusion

It can be concluded that this research article introduces a novel visual aid system, in the form of a pair of eyeglasses, for the completely blind. The key features of the proposed device include the following.

- The hands free, wearable, low power, low cost, and compact design for indoor and outdoor navigation.
- The complex algorithm processing using the low-end processing power of Raspberry Pi 3 Model B+.
- Dual capabilities for object detection and distance measurement using a combination of camera and ultrasound sensors.
- Integrated reading assistant, offering image-to-text conversion capabilities, enabling the blind to read texts from any document.

5.4 Future Scope

Assistive technology for the visually impaired and blind people is a research field that is gaining increasing prominence owing to an explosion of new interest in it from disparate disciplines. The field has a very relevant social impact on our ever-increasing aging and blind populations Assistive technology for the visually impaired (VI) and Blind people is concerned with “technologies, equipment, devices, apparatus, services, systems, processes and environmental modifications” [43] that enable them to overcome various physical, social, infrastructural and accessibility barriers to independence and live active, productive and independent lives as equal members of the society. Vision being an extremely vital sensory modality in humans, the loss of it affects the performance of almost all activities of daily living (ADL) and instrumental activities of daily living (IADLs); thereby hampering an individuals’ quality of life (QoL), general lifestyle, personal relationships and career.

Therefore, technology that facilitates accessibility, safety, and an improved quality of life has a very relevant social impact. Moreover, with our ever-increasing ageing and blind populations, it has the potential to broadly impact our quality of life in the future. This has driven novel research across many disparate disciplines, from cognitive psychology and neuroprosthetics to computer vision and sensor processing to rehabilitation engineering. More recently, advances in computer vision, wearable technology, multisensory research, and medical interventions have facilitated the development of numerous assistive technology solutions of both kind, invasive and non-invasive.

The proposed device has achieved a significant feat in assisting the blind and visually impaired by utilizing some key aspects of modern embedded computing yeilding satisfactory performance as illustrated in the previous sections.

However ,there can be further enhancements that would improve the performance of the proposed device such as introducing an architecture specific image processing algorithm for object detection that would improve the efficiency of classification as well as the time taken . This research article also motivates further work in the domain of electronic aids of the similar kind involving many other techniques.

The limitations of brightness classification and ease of use can be countered with a much better comfortable setup by surveying the needs of the visually impaired and introducing lightweight comfortable materials for eyewear as well as better approaches for OCR and object detection

References

[1] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," IEEE Transactions on . Systems, Man and Cybernetics. Part C Applications, vol. 40, Jan. 2010

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

[3] E. E. Pissaloux, R. Velazquez, and F. Maingreaud, "A new framework for cognitive mobility of visually impaired users in using tactile device," IEEE Transactions on Human-Machine Systems, vol. 47, Dec. 2017.

[4] K. Patil, Q. Jawadwala, and F. C. Shu, "Design and construction of electronic aid for visually impaired people," IEEE Transactions on Human-Machine Systems, vol. 48, Apr. 2018.

[5] Blindness and vision impairment, World Health Organization, Geneva, Switzerland, Oct. 2019.

[6] R. K. Katzschmann, B. Araki, and D. Rus, "Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device," IEEE Transactions on Neural Systems and Rehabilitation Engineering , vol. 26, Mar. 2018.

[7] J. Villanueva and R. Farcy, "Optical device indicating a safe free path to blind people," IEEE Transactions on Instrumentation Measures, vol. 61, Jan. 2012.

[8] X. Yang, S. Yuan, and Y. Tian, "Assistive clothing pattern recognition for visually impaired people," IEEE Transactions om Human-Machine Systems, vol. 44, Apr. 2014.

[9] S. L. Joseph et al., "Being aware of the world: Toward using social media to support the blind with navigation," IEEE Transactions on Human-Machine Systems, vol. 45,Jun. 2015.

[10] L. Tepelea, I. Buciu, C. Grava, I. Gavrilut, and A. Gacsadi, "A vision module for visually impaired people by using raspberry PI platform," 15th International Conference on Engineering , Modern Electronic Systems (EMES), Oradea, Romania, 2019,

[11] L. Dunai, G. Peris-Fajarnés, E. Lluna, and B. Defez, “Sensory navigation device for blind people,” J. Navig., vol. 66, no. 3, pp. 349–362, May 2013.

[12] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, “Virtual-blind-road followingbased wearable navigation device for blind people,” IEEE Trans. Consum. Electron., vol. 64, no. 1, pp. 136–143, Feb. 2018

[13] V.-N. Hoang, T.-H. Nguyen, T.-L. Le, T.-H. Tran, T.-P. Vuong, and N. Vuillerme, “Obstacle detection and warning system for visually impaired people based on electrode matrix and mobile kinect,” Vietnam Journal of Computer science Sci., vol. 4, no. 2, pp. 71–83, Jul. 2016.

[14] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, “Virtual-blind-road following based wearable navigation device for blind people,” IEEE Trans. Consum. Electron., vol. 64, no. 1, pp. 136–143, Feb. 2018

[15] B. Jiang, J. Yang, Z. Lv, and H. Song, “Wearable vision assistance system based on binocular sensors for visually impaired users,” IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1375–1383, Apr. 2019.

[16] A. Chalamandaris, S. Karabetsos, P. Tsiakoulis, and S. Raptis, “A unit selection text-to-speech synthesis system optimized for use with screen readers,” IEEE Transactions on Consumer Electronics, vol. 56 Aug. 2010.

[17] B. Andò, S. Baglio, V. Marletta, and A. Valastro, “A haptic solution to assist visually impaired in mobility tasks,” IEEE Transactions on Human-Machine Systems ,Oct. 2015.

[18] V. V. Meshram, K. Patil, V. A. Meshram, and F. C. Shu, “An astute assistive device for mobility and object recognition for visually impaired people,” IEEE Trans. Human-Machine Systems, vol. 49, no. 5, pp. 449–460, Oct. 2019.

[19] F. Lan, G. Zhai, and W. Lin, “Lightweight smart glass system with audio aid for visually impaired people,” in Proceedings of IEEE Region 10 Conference, Macao, China, 2015, pp. 1–4.

[20] M. M. Islam, M. S. Sadi, K. Z. Zamli, and M. M. Ahmed, “Developing walking assistants for visually impaired people: A review,” IEEE Sensors Journal, vol. 19,Apr. 2019.

[21] R. Keefer, Y. Liu, and N. Bourbakis, “The development and evaluation of an eyes-free interaction model for mobile reading devices,” IEEE Transactions on Human-Machine Systems, vol. 43,Jan. 2013.

[22] J. Xiao, S. L. Joseph, X. Zhang, B. Li, X. Li, and J. Zhang, “An assistive navigation framework for the visually impaired,” IEEE Transactions on Human Machine Systems, vol. 45, Oct. 2015.

[23] A. Karmel, A. Sharma, M. Pandya, and D. Garg, “IoT based assistive device for deaf, dumb and blind people,” Procedia Computer Science, vol. 165, Nov. 2019.

[24] C. Ye and X. Qian, “3-D object recognition of a robotic navigation aid for the visually impaired,” IEEE Transactions on Neural Systems Rehabilitations Eng, vol. 26, Feb. 2018

[25] Y. Liu, N. R. B. Stiles, and M. Meister, “Augmented reality powers a cognitive assistant for the blind” eLife, vol. 7, Nov. 2018.

[26] A. Adebiyi et al., “Assessment of feedback modalities for wearable visual aids in blind mobility,” PLoS One, vol. 12, no. 2, Feb. 2017

[27] C. I. Patel, A. Patel, and D. Patel, “Optical character recognition by open source OCR tool Tesseract: A case study” International Journal of Computer Applications, vol. 55, Oct. 2012.

[28] T-Y Lin et al. “Microsoft COCO: Common objects in context,” Feb. 2015.

[29] J. Han et al., “Representing and retrieving video shots in human-centric brain imaging space,” IEEE Transactions on Image Processing, vol. 22, Jul. 2013.

[30] J. Han, K. N. Ngan, M. Li, and H. J. Zhang, “Unsupervised extraction of visual attention objects in color images,” IEEE Transactions on Circuits Systems- Video Technology, vol. 16, Jan. 2006

[31] M. Rajesh et al., “Text recognition and face detection aid for visually impaired person using raspberry PI” in proceedings - International Conference of Circuit, Power and Computing Technologyies.,Kollam,India,2017.