A Synopsis on

Smart Assistive Device for the Visually Challenged

Submitted in partial fulfillment of the requirements
of the degree of

Bachelor of Engineering

in

Information Technology

by

Sameer Dev (16104030)
Sudama Jaiswal (161040315)
Yogendra Kokamkar (16104016)

Prof. Kiran B. Deshpande
Prof. Kaushiki Upadhyaya

Department of Information Technology
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

2019-2020

CERTIFICATE

This is to certify that the project Synopsis entitled ***'Smart assistive device for visually challenged"*** submitted by ***"Sameer Dev (16104030)", "Sudama Jaiswal (161040315)", "Yogendra Kokamkar (16104016)"*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Information Technology*** to the University of Mumbai,is a bonafide work carried out during academic year 2019-2020

(Prof. Kaushiki Upadhyaya)
Co-Guide

(Prof. Kiran B. Deshpande)
Guide

Prof. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place:A.P.Shah Institute of Technology, Thane
Date:

Declaration

We declare that this written submission represents my ideas in my our words and where others' ideas or words have been included we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

(Sameer Dev - 16104030)
(Sudama Jaiswal - 16104015)
(Yogendra Kokamkar - 16104016)

Date:

Abstract

Blind, the dictionary defines it in one simple word, sightless. The life for a visually challenged person is extremely hard for obvious reasons. In this era of cutting-edge technology, it is still extremely difficult for visually challenged people to carry out day to day chores or enjoy the simple pleasures of life such as going for a walk, socializing, and so on. Hence, developing new solutions that allow those individuals to interact with sighted people and the sighted world, in a way that lessens any of the problems that can arise from being visually impaired is becoming increasingly important. This paper presents a Smart Device built using a Raspberry Pi that can be controlled via Voice Commands and carry out various tasks such as Object Detection, Navigation and notify the user through Haptic and Audio feedback. The device will also take help of Image Recognition and Image Processing in order to convey information about specific places to the user as soon as the user is in that particular vicinity, hence allowing the person using the device know their surrounding environment in a better way.

Introduction

According to the October 2018 article of World Health Organization (WHO) globally, it is estimated that approximately 1.3 billion people live with some form of distance or near vision impairment. With regards to distance vision, 188.5 million have a mild vision impairment, 217 million have moderate to severe vision impairment, and 36 million people are blind. With regards to near vision, 826 million people live with near vision impairment. Population growth and aging will increase the risk that more and more people acquire vision impairment in the near future.

In this high-tech era, technology has made it possible for everyone to live a comfortable life. From entertainment to sports, studies to arts, technology has had a huge impact on our lives which has certainly made everybody's life enjoyable as well as easy. And yet, even after so many technological advancements and breakthroughs physically challenged people still need to depend upon other people for day to day functioning which ultimately makes them less confident in unfamiliar settings.

There are many definitions of Assistive Technology in common use, they range from formal technical definitions maintained by organizations such as the WHO to informal definitions often popularized by users themselves. As outlined by Wikipedia, Assistive Technology is an umbrella term that includes assistive, adaptive, and rehabilitative devices for people with disabilities or the elderly population while also including the process used in choosing, identifying, and using them.

A growing variety of special devices are readily available for use by visually impaired people. They vary in cost from solely a couple of Hundred Rupees to Thousands of Rupees for a single device. Hardly ever does a man-made device, by itself, make the difference in whether or not a visually impaired person can do a job. Devices do, however, provide added independence and adaptability to visually impaired persons in numerous positions.

Visually impaired people have limited scope of reading and understanding text and images; hence a Voice interface will be a very important medium of communication for them. With the help of the voice interface, they will be able to get various kinds of information from the device. This interface will be enabled by the increasingly available Speech Recognition as well as Speech-to-Text APIs which allow a person to extract relevant features from audio signals.

With the advent of computing power and the humongous amount of data being generated every day, Deep Learning has come into much prominence in the past few years giving rise to powerful models. This, in turn, has made way for Transfer Learning which allows people to use the same models, with minor readjustments, for other purposes thus reducing their overhead of having to train a model from scratch. A pre-rained image classifier, coupled with new domain-specific data can do wonders while prediction new unseen images hence enabling Object Detection at amazing speeds.

This combination of Voice Recognition, Speech-to-Text and Object Detection when coupled into a single device and optimized in such a way so as to provide real-time analysis, as well as feedback, can benefit the needy in a plethora of ways. The device will act as a medium of communication and assistance, connecting visually impaired people to the outside world.

Objectives

The objectives behind developing this project are:

1. **Build a device for travelling safely around the campus.**

   Help a visually challenged individual walk around the campus, on his own and without any external human help.

2. **Overcome Environmental Challenges.**

   Physical movement is one of the biggest challenges for visually impaired people. Our device can help them get an overview of the surroundings so they can be more confident about the path and surroundings they are travelling in

3. **Overcome Technological Challenges.**

   The small touch screens of many tablets and smartphones may be particularly difficulty for the visually impaired. Our device can be fully controlled by voice or by an app with a subtle and simple user interface, overcoming the technology usage barrier

4. **Overcome Social Challenges.**

   Blindness can cause significant social challenges, typically because there are activities in which blind people can't easily participate. Our device can assist the visually impaired in regards to other people, things, animals and surroundings by which the he/she can decide a way of social interaction

Literature Review

The papers referred to while developing the system are mentioned below:

1. **Voice-controlled smart assistive device for visually impaired individuals. D., Munteanu, Ionel, R., 2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC)**

   -This paper presents the modelling, implementation and testing of an experimental microcontroller (MCU) based smart assistive system which can be used by the visually impaired or blind people. This device includes haptic and audio feedback options from which the user can select. A Smart Phone can be used to control the device using predefined voice commands and Bluetooth connectivity. The device is portable and the purpose of its usage is to warn the user when objects are present on the walking path so collision can be avoided. Distance measurements, between the user and possible obstacles, are performed using ultrasonic echolocation and the data provided by the ultrasonic sensor is processed by a microcontroller, which also handles the feedback part. The hardware design, software architecture and mechanical design of the enclosure as well as the breadboard prototyping are covered in this material. Experimental results performed in different functionality scenarios demonstrate that the proposed system can be successfully used to fulfil its purpose.

   The device is powered by a lithium-ion battery that can provide an output voltage of 3.7 V. This voltage is raised by the high efficiency POLOLU-2115 step-up converter at nominal 5 V - which is the supply voltage for the MCU. A charging solution for the battery is also provided using the TP4056 module. The charging module uses the LTC5056 chip with thermal feedback from Linear Technologies to provide a constant 4.2 V charging voltage and can be supplied with 5 V at 1 A from any charger with mini USB type B connector.

   The enclosure of the device was designed in Creo and 3D printed and has the following dimensions: a length of 60 mm, a width of 49 mm and 29 mm in height. It has a leather band attached and can be wrist mounted. In the back side of the case two hinges are designed such that the PCB, the battery, the ultrasonic rangefinder, the battery charger module and the Bluetooth module are easily accessible.

   Preferably, the whole component should have a uniform thickness - the nominal wall thickness. This was not possible for this design but variations were kept to a minimum. Most of the walls of the enclosure are 2 mm thick, but there are areas with 4 mm made to be used as mechanical support for the PCB. The thinnest part of the enclosure has 1 mm and it is located on the bottom side.

2. **Blind assistive device - Smart Lazy Susan, Lee, C.-N., Chu, Y.-T., Cheng, L., Lin, Y.-T., Lan, K.-F., 2017 International Conference on Machine Learning and Cybernetics (ICMLC).**

   -The Lazy Susan is a common device used in Chinese dining occasions. Currently, commercially available Lazy Susan mostly consists of two overlapping tables, where the upper one is used to carry dishes and turn upon the lower one. Therefore, diners can manually turn the upper table to position a dish in front of them, allowing an easier access to the food in the dish.

   To minimize the inconvenience of manual turning, current technology has entailed electric versions of the Lazy Susan that enables diners to control the rotation of a Lazy Susan without manually turning it. However, both manual and electric Lazy Susan are designed for diners with proper vision. For visually impaired people, dining with family and friends using a Lazy Susan always entails relying on their friend or family member sitting next them to select dishes for them or turn the Lazy Susan to change the dish in front of them.

   Statistics have shown that currently Taiwan has a visually impaired population of approximately 56,529. Among this population, approximately 42% suffered total blindness before the age of 5 mainly due to genetic inheritance or accidental injury. Considering that Taiwan has a large group of visually impaired people in need of care and assistance, the motivation of this study is to care for them in daily lives and save the dining inconveniences faced by them. By designing and realizing a set of table aids to assist visually impaired people with dining, this study can provide dining information for visually impaired diners, and enhance their dining convenience and safety. Visually impaired people can learn the dishes independently and choose the dish they prefer comfortably without relying on others. Thus, they would feel more secure and fulfilled rather than socially isolated when dining with family and friends.

3. **An intelligent walking stick for the visually challenged people, H., Sahoo, N., Lin, H.-W., IEEE International Conference on Applied System Innovation 2018, IEEE ICASI 2018- Meen, Prior Lam (Eds)**

   -In this high-tech era, technology has made it possible that everyone can live a comfortable life. But somehow the physically challenged people need to depend upon others in their daily life which ultimately makes them less confident in an unfamiliar environment. But nowadays the explosion of innovative technology provides many opportunities for them to live confidently without feeling as a burden. So, in this paper, an intelligent device is represented for visually challenged people to guide them to reach their destination place safely without facing any difficulties. It consists of Raspberry Pi and PIC as the controller, Global Positioning System (GPS) along with sensors like Ultrasonic and other supportive sensors and an Android-based Application (APP).

   Vision is a precious gift from God that one can able to see and enjoy this beautiful world. But many people throughout the world are deprived of this. According to October 2017 report of World Health Organization (WHO) an estimated 253 million people live with vision impairment: 36 million are blind and 217 million have moderate to severe vision impairment. Un-operated cataract is the main reason for blindness in low income and developing countries. So, in this case most of the visually challenged people cannot afford an expensive device to use as their supporter. So, in this paper we have proposed a cost-effective intelligent device. This device mainly consists of a walking stick and an APP.

   The walking stick is based on the integration of sensors which helps to detect obstacles in their path. A GPS module is also used to know the location of the blind person. For this, an APP is created, which makes them feel safe and secure because it helps in navigation as well as the relatives of blind person can also able to track their real-time location. Also, in an emergency condition the blind person can contact his guardians immediately. In early days the blind persons used trained dog to guide their path which is costly as well as not so efficient.

   Again, the traditional white cane is only able to detect objects by touch, so it has also limitations to get less time to react the situation after detecting an obstacle. There are many other mobility aids known as electronic travel aids (ETAs) in the market. But most of the commercial ETAs are expensive and lack of accuracy.

4. **Design and development of smart assistive device for visually impaired people., Yadav, A. B., Bindal, L., Namhakumar, V. U., Namitha, K., Harsha, H. (2016)., IEEE International Conference on Recent Trends in Electronics Information Communication Technology, May 20-21, 2016, India**

   -Blindness is a state of lacking the visual perception due to physiological or neurological factors. In this proposed work, a simple, cheap, friendly user, smart stick will be designed and implemented to improve the mobility of both blind and visually impaired people in a specific area. This multipurpose model is designed to help the blind person to navigate alone safely and to avoid any obstacles that may be encountered, whether fixed or mobile, to prevent any possible accident.

   The device provides voice output giving direction to the blind Using RFID technology, the destination of the bus is detected and voice announcement is given regarding the destination of the bus. The location of stick is added advantage to the current multipurpose device. Using RFID technology, the location of the stick is achieved.

   The blind is provided with a push button to locate the stick. People with low vision or complete blindness face difficulty in navigating surroundings they are not familiar with and usually require someone to help them navigate. They often bump into the obstacles present in their way thus hindering their free movement.

   The conventional white sticks that are used by the blind do not help them to avoid the obstacles efficiently. Only those obstacles that are hit by the stick are identified and can be avoided. However, the obstacles in the surroundings are at different heights and distances, and sometimes cannot be identified by the white canes used.

   In order to navigate independently and confidently in unfamiliar environments it is required that the blind people are well aware about the obstacles in their path from a distance. This can be achieved by implanting sensors in the traditional white cane, which can then be used to detect the obstacles. There are many technologies that can be used to detect the obstacles in the path from a distance. Smart assistive device indicates an intelligent device that will help the blind in his easy mobility and to carry out his work like any other person.

5. **Smart device for visually impaired people, Kasthuri, R., Nivetha, B., Shabana, S., Veluchamy, M., Sivakumar, S., 2017 Third International Conference on Science Technology Engineering Management (ICONSTEM).**

-In this system enables blind people to handle the Android phone effectively. The blind people wanted to make use of the services like calling, getting notification of battery level, hearing music and to get latest updates on the Android phone. The proposed system in fig. 1. enables to obtain all the services through their voice command. The Selendroid enables the communication between smart phone and the various web servers. It has the ability to identify the spoken languages and convert into machine understandable format. This is done by the SRE.

The individual speaker input is read and isolated into vocabulary. This system performs an action that is usually performed by a normal person. The VIPs who need to use Android phone have to give their input through voice to the SRE through microphone or headset. The SRE converts the speech into text. The text is given as input to the command recognition module. The command is recognized and identified by using morphological analyzer. SRE output controls the dialler, music player, Selendroid architecture (SA) and Google maps.

The dialer manager gives the option of dial, hold, disconnect. The music manager include play, stop, pause, forward and rewind the music tracks. The battery notification is given when it indicates battery level is low to certain range i.e., 20%. Google maps are used by the system to retrieve the longitude and latitude of the VIPs and given to the concern resources, events and to remember certain incident that has person.

The concern person is the known person of VIP. By making use of the coordinates, the concern person can easily identify the accurate position where the user will be present. Thus, it overcomes the scenario of missing situation to a minimal extent. After knowing the location, the concern person makes a call to the user to acknowledge the situation of missing scenario.

6. **Low cost ultrasonic smart glasses for blind, Agarwal, R., Ladha, N., Agarwal, M., Majee, K. K., Das, A., Kumar, S., Saha, H. N. (2017). 2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).**

   -A smart ultrasonic glasses for blind people comprises of a pair of wearable glasses, ultrasonic sensors for detection of obstacles in the way of blind man, a buzzer to give the sound as per the direction of the obstacle from the man, a central processing unit comprising of Arduino NANO which takes the information from the sensor about the obstacle distance and processes the information according to the coding done and sends the output through the buzzer, power supply is given to the central unit which distributes the power to different components. The sensor is mounted in between of the top bar and bridge present in optical glasses as shown in the figure.

   All the components are connected to the central unit using single strand copper wires and the power is given to the central unit using a USB cable. The best sensors that can be used will be ultrasonic sensors because ultrasound is a strong point, the energy consumption of slow wave propagating in the medium relatively far distance. Therefore, often it is used to measure the distance over big length. At the same time, ultrasound for the object in the dark, dust, smoke, electromagnetic interference, toxic and other harsh environments have a certain ability to adapt, with a wide range of applications.The ultrasonic sensor is fixed at a perpendicular from the glasses. According to claim 1, as the blind man goes closer to the obstacle the distance sent by the sensors to the central unit will decrease. Hence the beeping of the buzzer will take shorter intervals and hence the beeping will be faster.

   At present, many of the navigation device using seeing-eye guide dogs, guide dogs by seeing some extent, although the trip to ensure the safety of the blind. But there are still some problems, training a guide dog larger difficulty, generally have to spend 3-6 months, training a skilled guide dogs will need to spend about two years, with dog the daily life of consumer spending, the cost it takes to reach the million, while the limited life cycle of guide dogs. These smart glasses are very easy to use and very simple to understand. If a blind uses it for 2-3 times then he/she will understand the working and can handle it easily

Problem Definition

Almost all of the works of literature mentioned have a fundamental requirement or dependency, i.e. every smart device that has been proposed needs to be connected to an android application for proper functioning. This gives rises to the question as to what should a visually impaired person do if he/she has no access to a smartphone to run the required application? Should they suffer for not having a smartphone?

We propose a novel device that is standalone in nature and can function without having to connect to an Android application. Although there will be an Android app for the device to connect with, it will not be a required dependency as is the case with most smart devices. The main purpose of the app will be to change certain system settings, debug the device and view system generated logs and stats.

The proposed assistive device for visually impaired people will let them get a better sense of the surroundings and environments of our college and similar premises. It will also give them haptic and audio feedback regarding the obstacles that come in their path. The device can also classify nearby objects, thanks to the camera module and the power of Deep Learning, which will help give the user a basic idea about the objects surrounding him/her. The device will also be fully operable with the help of voice commands that will be enabled thanks to the various Text-to-Speech and Speech-to-Text APIs available for development purposes.

Proposed System Architecture/Working

The proposed device will help visually impaired people get a better sense of their surroundings and overcome the various challenges they face daily for doing the simplest of tasks. The device is portable and the purpose of its usage is to warn the user when objects are present in the walking path so a collision can be avoided and also guide the user in the right directions while walking.

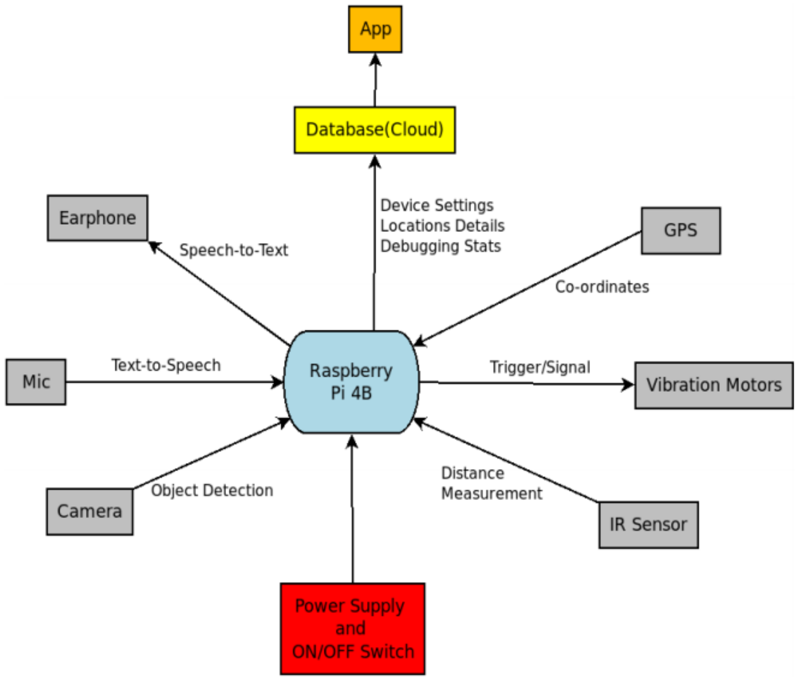


Figure 1: Block Diagram of Proposed System

Distance measurements, between the user and possible obstacles, are done using an IR Sensor. This module is chosen instead of an ultrasonic sensor because its maximum range is considerably higher than that of an ultrasonic sensor. Object Detection and Image Classification are performed using Deep Learning models based on Convolution Neural Networks (CNN) implemented via OpenCV and the data provided by the modules is processed by the Raspberry Pi, which also handles the audio and haptic feedback part.

In practice, very few people train an entire Convolutional Network from scratch (with random initialization), because it is relatively rare to have a dataset of sufficient size. Instead, it is common to pretrain a ConvNet on a very large dataset (e.g. ImageNet, which contains 1.2 million images with 1000 categories), and then use the ConvNet either as an initialization or a fixed feature extractor for the task of interest.

This technique is known as Transfer Learning wherein we take a pre-trained, state-of-the-art image classifier such as VGG16, InceptionV3, ResNet50, DenseNet, NasNet, etc. and instead of training the whole model we only train the fully connected layers with our data while keeping the weights of the previous layers unchanged. This technique, while being faster, also proves to be more accurate than creating a CNN from scratch.

Delving a little bit more into the topic of transfer learning, the three major transfer learning scenarios look as follows:

1. ConvNet as fixed feature extractor: Take a ConvNet pretrained on ImageNet, remove the last fully-connected layer and then treat the rest of the ConvNet as a fixed feature extractor for the new dataset
2. Fine-tuning the ConvNet: The second strategy is to not only replace and retrain the classifier on top of the ConvNet on the new dataset, but to also fine-tune the weights of the pre-trained network by continuing the back-propagation
3. Pre-trained models: Since modern ConvNets take 2-3 weeks to train across multiple GPUs on ImageNet, it is common to see people release their final ConvNet checkpoints for the benefit of others who can use these networks for fine-tuning.

Using OpenCV we can perform a frame by frame, real-time video analysis with the help of which we will be able to detect objects in front of us. For illustration purposes, in Fig 4.2 we have shown bounding boxes along with the class/category to which an object belongs. Also included is a numerical value. This value is the confidence with which our model is able to identify an object in an image.

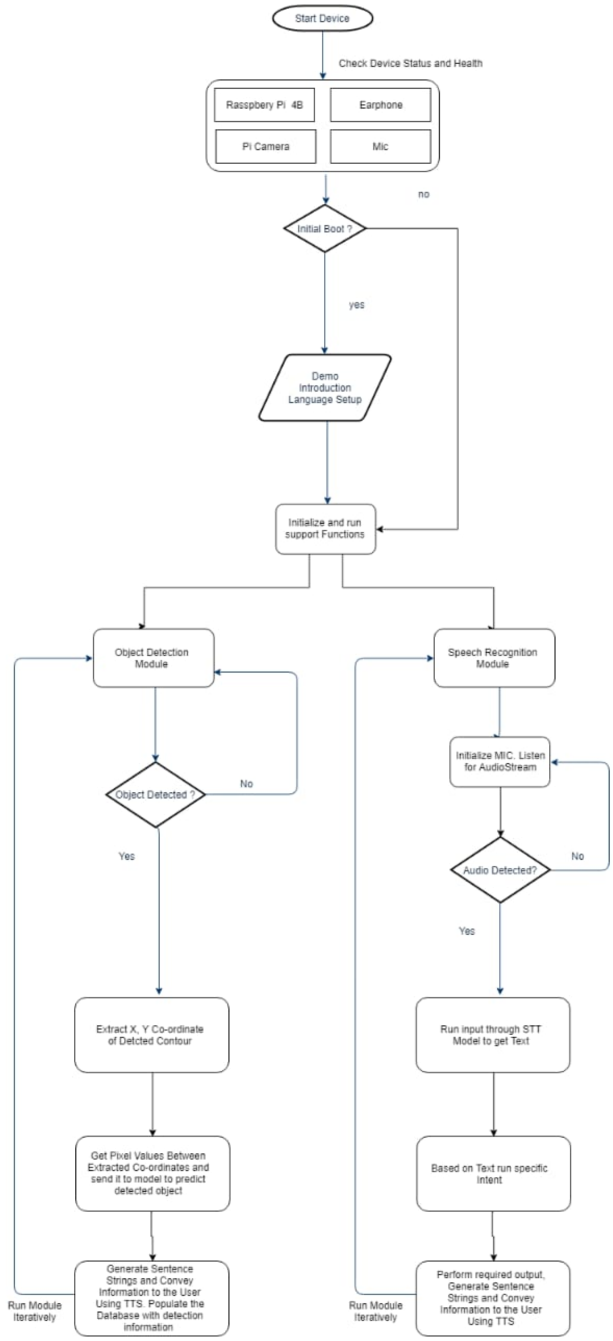
Figure 2: Output of the Object Detection Module

Once an object has been detected the device will then calculate the distance between the object and the user using the IR Sensor and the information will be conveyed to the user via the audio feedback system. In case multiple objects are detected, the information about the nearest object will be conveyed first followed by information about the rest of the objects. The program will also analyze the direction in which the user must go in order to avoid the obstacle. The information about the directions to be taken will be conveyed via audio as well as haptic feedback. The direction in which the user has to go is based on the obstacles ahead of him which will also determine which motor vibrates (left motor in-case of left direction and right motor in-case of right direction).

Since we are using OpenCV, we can also perform minor Image Processing activities in order to parse sign-boards, hoardings, etc. and extract useful information from them. This parsed information can then be conveyed to the user pretty conveniently.

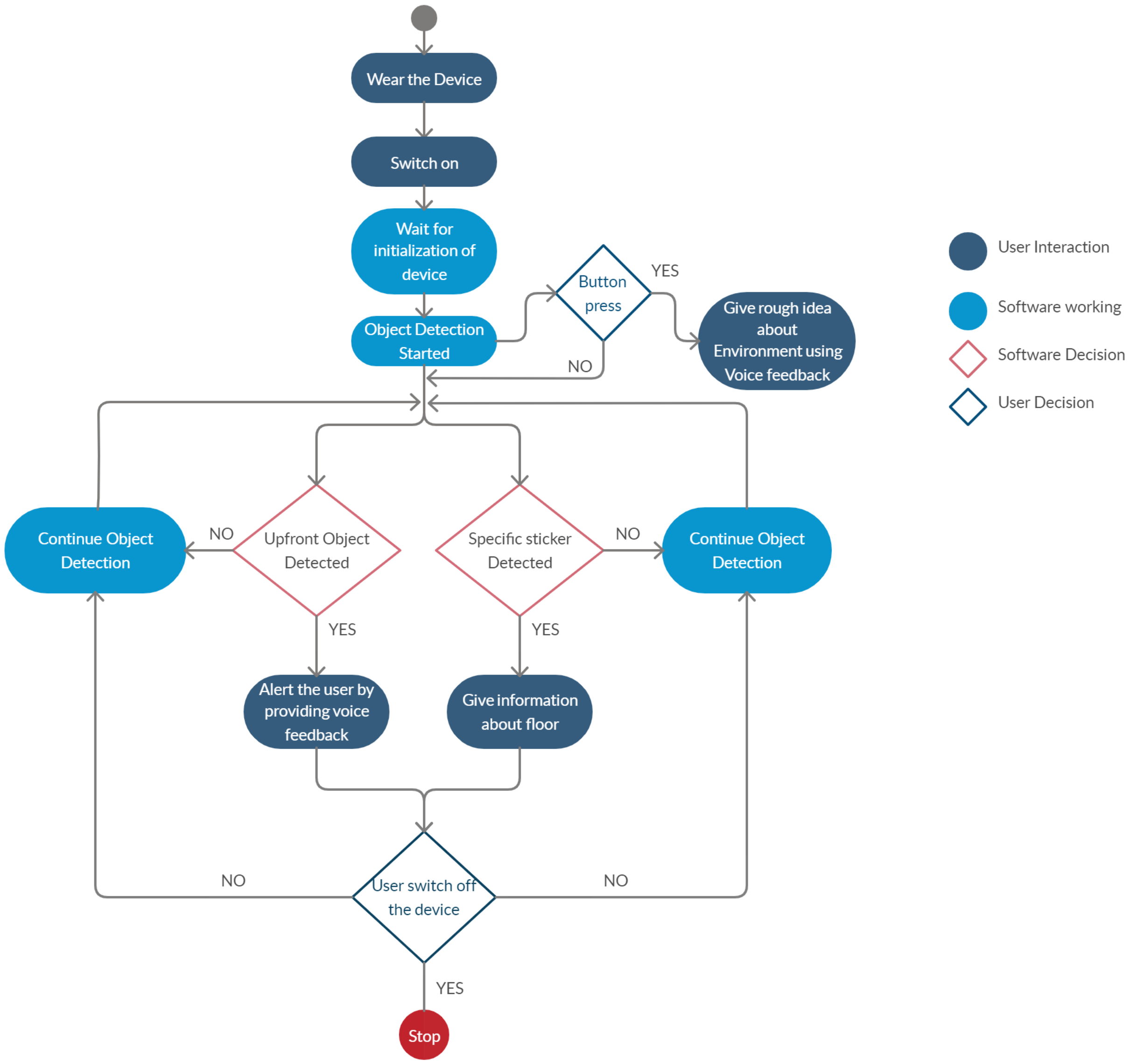
The advantage of adding Speech Recognition to the device also opens up the possibility of creating a Conversational AI or chat-bot which can carry out certain tasks based on the needs of the user, the input for which will be available via the Speech-To-Text module.

Flowchart



Start Device
Check Device Status and Health
Rasspbery Pi 4B
Earphone
Pi Camera
Mic
Initial Boot ?
no
yes
Demo
Introduction
Language Setup
Initialize and run
support Functions
Object Detection
Module
Speech Recognition
Module
Object Detected ?
No
Yes
Initialize MIC. Listen
for AudioStream
Audio Detected?
No
Yes
Extract X, Y Co-ordinate
of Detcted Contour
Run input through STT
Model to get Text
Get Pixel Values Between
Extracted Co-ordinates and
send it to model to predict
detected object
Based on Text run specific
Intent
Generate Sentence
Strings and Convey
Information to the User
Using TTS. Populate the
Database with detection
information
Run Module
Iteratively
Perform required output,
Generate Sentence
Strings and Convey
Information to the User
Using TTS
Run Module
Iteratively

Activity Diagram



Design and Implementation

The final project/program is a culmination of various modules, namely, Text-To-Speech (TTS), Object Detection, Voice Recognition, Environment Description. These 4 core modules work in tandem in order to provide the user with the required information in a swift and timely manner and help the user make his/her decisions.

1. Text-To-Speech (TTS)

Text-To-Speech (TTS) is the conversion of written text into spoken voice. The TTS module is necessary to communicate information to the user. It first generates the text that is to be spoken, hits an API that returns the audio to be played, in Base64 format, and then plays that audio for the user which he/she can listen through the earphone. The information relayed could range from detected objects to describing the environment.

Code Snippets:

```
if __name__ == '__main__':
    q = Queue()
    t1 = Thread(target = object_detection,args =(q, ))
    t2 = Thread(target = txtospeech,args =(q, ))
    t1.start()
    t2.start()
```

Figure 3: The TTS function getting called in the main Thread

```
def txtospeech(in_q):
    CLASSES = ["background", "aeroplane", "bicycle", "bird", "boat",
        "bottle", "bus", "car", "cat", "chair", "cow", "diningtable",
        "dog", "horse", "motorbike", "person", "pottedplant", "sheep",
        "sofa", "train", "tvmonitor"]
    idx = in_q.get()
    engine = pyttsx3.init()
    engine.say(CLASSES[idx])
    engine.runAndWait()
```

Figure 4: The actual TTS function

2. Object Detection

It is the main module that is responsible for detecting objects in the users path and calling the TTS function in order to convey the information to the user in a timely manner. One of the important fields of Artificial Intelligence is Computer Vision. Computer Vision is the science of computers and software systems that can recognize and understand images and scenes. Computer Vision is also composed of various aspects such as image recognition, object detection, image generation, image super-resolution and more. Object detection is probably the most profound aspect of computer vision due the number practical use cases.

Code Snippets:

```python
if __name__ == '__main__':
    q = Queue()
    t1 = Thread(target = object_detection,args =(q, ))
    t2 = Thread(target = txtospeech,args =(q, ))
    t1.start()
    t2.start()
```

Figure 5: The Object Detection function running continuously in the main Thread

```python
# initialize the list of class labels MobileNet SSD was trained to
# detect, then generate a set of bounding box colors for each class
def object_detection(out_q):
    CLASSES = ["background", "aeroplane", "bicycle", "bird", "boat",
        "bottle", "bus", "car", "cat", "chair", "cow", "diningtable",
        "dog", "horse", "motorbike", "person", "pottedplant", "sheep",
        "sofa", "train", "tvmonitor"]
    COLORS = np.random.uniform(0, 255, size=(len(CLASSES), 3))

    # load our serialized model from disk
    print("[INFO] loading model...")
    net = cv2.dnn.readNetFromCaffe(args["prototxt"], args["model"])

    # initialize the video stream, allow the cammera sensor to warmup,
    # and initialize the FPS counter
    print("[INFO] starting video stream...")
    vs = VideoStream(src=0).start()
    time.sleep(2.0)
    fps = FPS().start()

    # loop over the frames from the video stream
    while True:
        # grab the frame from the threaded video stream and resize it
        # to have a maximum width of 400 pixels
        frame = vs.read()
        frame = imutils.resize(frame, width=400)

        # grab the frame dimensions and convert it to a blob
        (h, w) = frame.shape[:2]
        blob = cv2.dnn.blobFromImage(cv2.resize(frame, (300, 300)),
            0.007843, (300, 300), 127.5)
```

Figure 6: The actual Object Detection function

3. Voice Recognition

Voice or speaker recognition is the ability of a machine or program to receive and interpret dictation or to understand and carry out spoken commands. Voice recognition has gained prominence and use with the rise of AI and intelligent assistants, such as Amazon's Alexa, Apple's Siri and Microsoft's Cortana. This module is responsible for understanding the requests of the user and calling the appropriate function/API based on the result. The module being used the Google Cloud Speech API. It is pretty easy to setup and the responses time as well as latency is much better as compared to other similar service providers such as Amazon Web Services or Microsoft Azure.

Code Snippets:

```python
#Sample rate is how often values are recorded
sample_rate = 48000
#Chunk is like a buffer. It stores 2048 samples (bytes of data)
#here.
#it is advisable to use powers of 2 such as 1024 or 2048
chunk_size = 2048
#Initialize the recognizer
r = sr.Recognizer()

#generate a list of all audio cards/microphones
mic_list = sr.Microphone.list_microphone_names()

#the following loop aims to set the device ID of the mic that
#we specifically want to use to avoid ambiguity.
for i, microphone_name in enumerate(mic_list):
    if microphone_name == mic_name:
        device_id = i

#use the microphone as source for input. Here, we also specify
#which device ID to specifically look for incase the microphone
#is not working, an error will pop up saying "device_id undefined"
with sr.Microphone(device_index = device_id, sample_rate = sample_rate,
                        chunk_size = chunk_size) as source:
    #wait for a second to let the recognizer adjust the
    #energy threshold based on the surrounding noise level
    r.adjust_for_ambient_noise(source)
    print "Say Something"
    #listens for the user's input
    audio = r.listen(source)
```

Figure 7: The Voice Recognition module that runs continuously

4. Environment Description

Describing the environment (generating caption) is a challenging artificial intelligence problem where a textual description must be generated for a given photograph. It requires both methods from computer vision to understand the content of the image and a language model from the field of natural language processing to turn the understanding of the image into words in the right order. Recently, deep learning methods have achieved state-of-the-art results on examples of this problem. Deep learning methods have demonstrated state-of-the-art results on caption generation problems. What is most impressive about these methods is a single end-to-end model can be defined to predict a caption, given a photo, instead of requiring sophisticated data preparation or a pipeline of specifically designed models. The model has been trained on 8000 images available in the Flickr8K dataset. The image is first passed to a Convolution Neural Network (CNN) which is a pre-trained InceptionV3 model that helps us extract the features of the image. The extracted features are then passed on to a Recurrent Neural Network (RNN) that predicts the captions/text for that particular image.

Code Snippets:

```
def generate_caption_beam_search(model, tokenizer, image, max_length, beam_index=3):
    # in_text --> [[idx,prob]] ;prob=0 initially
    in_text = [[tokenizer.texts_to_sequences(['startseq'])[0], 0.0]]
    while len(in_text[0][0]) < max_length:
        tempList = []
        for seq in in_text:
            padded_seq = pad_sequences([seq[0]], maxlen=max_length)
            preds = model.predict([image,padded_seq], verbose=0)
            # Take top (i.e. which have highest probailities) `beam_index` predictions
            top_preds = np.argsort(preds[0])[-beam_index:]
            # Getting the top `beam_index` predictions and
            for word in top_preds:
                next_seq, prob = seq[0][:], seq[1]
                next_seq.append(word)
                # Update probability
                prob += preds[0][word]
                # Append as input for generating the next word
                tempList.append([next_seq, prob])
        in_text = tempList
        # Sorting according to the probabilities
        in_text = sorted(in_text, reverse=False, key=lambda l: l[1])
        # Take the top words
        in_text = in_text[-beam_index:]
    in_text = in_text[-1][0]
    final_caption_raw = [int_to_word(i,tokenizer) for i in in_text]
    final_caption = []
    for word in final_caption_raw:
        if word=='endseq':
            break
        else:
            final_caption.append(word)
    final_caption.append('endseq')
    return ' '.join(final_caption)
```

Figure 8: Function that calls the trained model to predict the caption based on the image passed to it

Summary

The main motive of our work is to create a trustworthy, efficient and real-time device for the visually impaired so that they can roam around their surroundings freely, without the need of being helped or relying on others even for the simplest pleasures of life. We hope to remove the drawbacks of all the previously mentioned illustrations wherein an Android Application is a required dependency for even setting up the device. The powerful model that we built using a pre-trained model gives us higher accuracy due to the concept of Transfer Learning. Since we are relying on the model to predict the objects in every frame of the captured videos, the task of predication is computationally very expensive which means that most of the resources of the Raspberry Pi are consumed by the Object Detection module itself. A newer model of the Raspberry Pi can be used to fulfill the hardware requirements.

References

[1] World Health Organization (WHO), https://www.who.int/newsroom/fact-sheets/detail/blindness-and-visual-impairment, last accessed on 29.07.19.

[2] Wikipedia, https://en.wikipedia.org/wiki/Assistive$_{t}echnology, lastaccessedon$29.07.19.

[3] Chang, Y.-H., Sahoo, N., Lin, H.-W. (2018). An intelligent walking stick for the visually challenged people. 2018 IEEE International Conference on Applied System Invention (ICASI)

[4] Yadav, A. B., Bindal, L., Namhakumar, V. U., Namitha, K., Harsha, H. (2016). Design and development of smart assistive device for visually impaired people. 2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)

[5] Munteanu, D., Ionel, R. (2016). Voice-controlled smart assistive device for visually impaired individuals. 2016 12th IEEE Internationa Symposium on Electronics and Telecommunications (ISETC)

[6] Agarwal, R., Ladha, N., Agarwal, M., Majee, K. K., Das, A., Kumar, S., Saha, H. N. (2017). Low cost ultrasonic smart glasses for blind. 2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)

[7] Lee, C.-N., Chu, Y.-T., Cheng, L., Lin, Y.-T., Lan, K.-F. (2017). Blind assistive device - Smart Lazy Susan. 2017 International Conference on Machine Learning and Cybernetics (ICMLC)

[8] Kasthuri, R., Nivetha, B., Shabana, S., Veluchamy, M., Sivakumar, S. (2017). Smart device for visually impaired people. 2017 Third International Conference on Science Technology Engineering Management (ICONSTEM)

[9] CS231n, Stanford University, http://cs231n.github.io/transfer-learning/, last accessed 16.08.19

1 Publication

Paper entitled **"Voice Based Smart Assistive Device for the Visually Challenged"** is submitted at **"IEEE Conference ID 45521. International Conference on Convergence to Digital World - Quo Vadis (ICCDW- 2020)"** by **"Sameer Dev, Yogendra Kokamkar and Sudama Jaiswal".**