Third Eye: A Device for Visually Impaired Persons

[1]Ms. Sabari L Umamaheshwari, [2]Amritha A, [3]Jayasree E, [4]Umess Ravi

[1]Ast.Profesor, Department of Electrical and Electronics Engineering, R.M.K. Engineering College, Kavaraipettai-601206,
Tamil Nadu, India

[2,3] Student, Department of Electrical and Electronics Engineering, R.M.K. Engineering College, Kavaraipettai-601206,
Tamil Nadu, India

Email:[1]slu.eee@rmkec.ac.in,[2]amri18106.ee@rmkec.ac.in, [3]jaya18129.ee@rmkec.ac.in, [4]umes18325.ee@rmkec.ac.in

Article Info
Page Number: 5910 - 5928
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received*: 25 March 2022*
Revised*: 30 April 2022*
Accepted*: 15 June 2022*
Publication*: 19 August 2022*

Abstract

The virtual assistant for the visually impaired is a advancement that assists outwardly weakened individuals with exploring with pace and assurance by identifying the nearby deterrents with the help of ultrasonic waves and telling them with a noise. They just got to wear this gadget as a band. It is a straightforward wearable innovation for blind individuals. These days, many brilliant gadgets are accessible for outwardly weakened individuals, yet a large portion of them dislike conveying, and the significant downside is that they need a ton of preparation to utilize. One of the principal idiosyncrasies of this development is that reasonable for everybody. It very well may be wearable like fabric, with minimal expense and effortlessness. It will radically help the local area. As indicated by WHO, 39 million individuals are assessed as outwardly weakened around the world. In their day-to-day existence, they are enduring a ton of difficulty. Many impacted individuals have been involving a conventional white stick for a long time, yet it has a lot of burdens. So, the venture plans to create a modest and more productive because to help the outwardly impeded explore with more prominent solace, speed, and certainty.

Index Terms – Raspberry Pi, YOLO v3, Google Text to Speech, Object Detection, Sonar.

I. Introduction

All around the world, the primary driver of vision misfortune is constant eye illnesses while the best two reasons for visual impedance are in-adjusted refractive blunders and unworked waterfall. In this quick world, outwardly debilitated individuals are abandoned and not treated similarly. To help them and furnish them with some degree of solace, numerous arrangements and strategies have been attempted and created. One of these procedures is called direction and versatility. In this method, an expert aid the outwardly impeded and blind individuals and trains them to continue all alone. They are prepared to rely upon their other excess faculties to move freely and securely. Another strategy is through utilizing aide canines. In this strategy, the canines are prepared uniquely to help the development of visually impaired individuals. The canines explore the impediments as an alarm to the client to fundamentally impact his direction. Be that as it may, it is challenging for outwardly debilitated and blind individuals to comprehend the complicated course given by these canines.

Moreover, the expense of these prepared canines is extremely high. These days, various kinds of sticks are likewise being utilized like the white stick, the shrewd stick, and the laser stick. Be that as it may, these apparatuses additionally have a few requirements. The stick is unhandy since it has a long length which additionally makes it hard to keep out in the open spots. Moreover, the stick has restrictions on perceiving deterrents. Numerous procedures have been grown as of late to upgrade visually impaired individuals' versatility. These methods are created because of sign handling and sensor innovation. Notwithstanding, these devices will generally flop in a jam-packed region or in a space where there are high volumes of electronic waves. To help blind and outwardly disabled individuals' versatility indoor and open air, this work proposes a basic electronic direction inserted vision framework which is configurable and proficient. The framework uses three kinds of gadgets including an IR sensor, sonar sensor, and camera.

A raspberry pi 4 model B microcontroller processes the reflected signs from all gadgets to characterize an obstruction. The proposed direction framework can decide the deterrent distance, notwithstanding the substance and form attributes of the snag. Besides, the framework can name a portion of the identified items. Additionally, neither stick nor other stamped apparatus is required to have been conveyed by the client. Like different frameworks, the proposed framework can be secured to a cap or a pen-sized hand smaller than a normal stick. It has high insusceptibility to both surrounding glow and the article's tone.

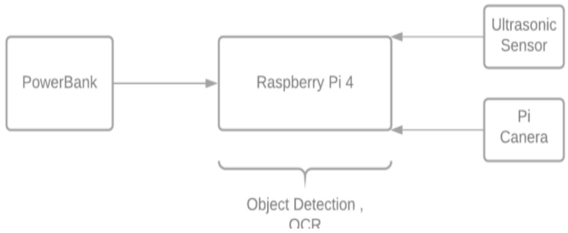


Figure 1.1 Block Diagram of Third Eye

II. METHODOLOGY

This venture proposes the plan of a compact computer-based intelligence-based direction framework for blind - Third Eye, which helps the outwardly impeded local area and helps in their everyday versatility. The third eye gives the outwardly debilitated local area a better approach to envision the world by making sense of them their environmental factors. The entire framework is constrained by the Raspberry Pi microcontroller. The third eye tackles the most extreme capacities of the Raspberry Pi microcontroller which can up hold the framework with one benefit being the inbuilt realistic card. The model purposes different sensors, for example, IR Sensors, Sonar Sensors, and a Camera module which assist the framework with the social occasion of the necessary information. Moreover, text to discourse module is utilized to converse with the client.

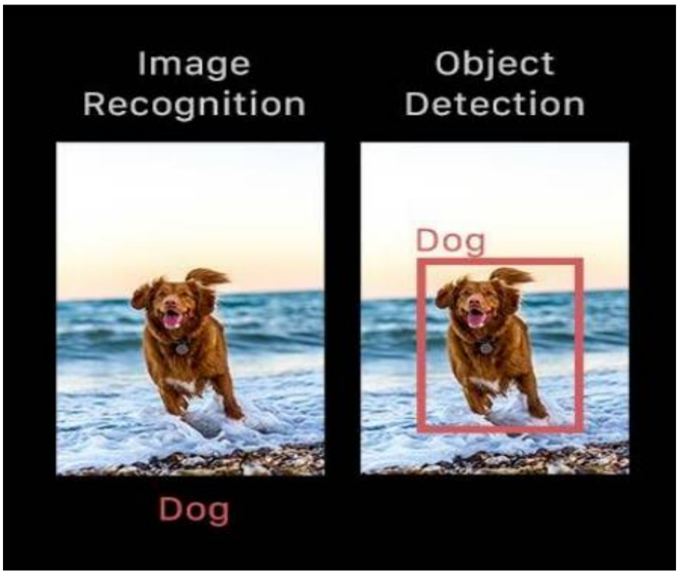


Figure 2.1. Figure of object detection

Python frames the core of the framework. It then processes the gathered information and converts it into data which is at last conveyed to the end client.

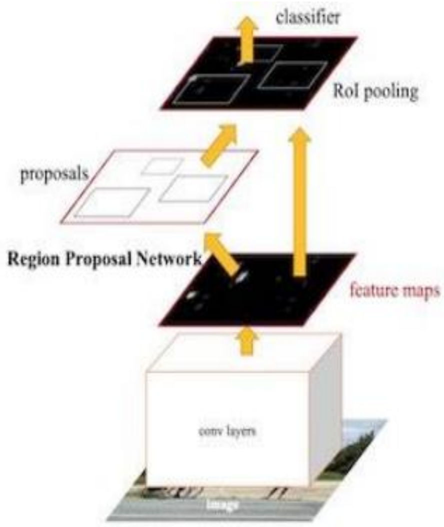


Figure 2.2. Figure of Single shot detectors

The camera module assumes a significant part as it takes the photos which are then handled utilizing a picture handling strategy to imagine the article appropriately. All the data is handled and changed over completely to a message which is then taken care of into a message to discourse module. The message to discourse module conveys this data to the end client in their ear utilizing earphones.

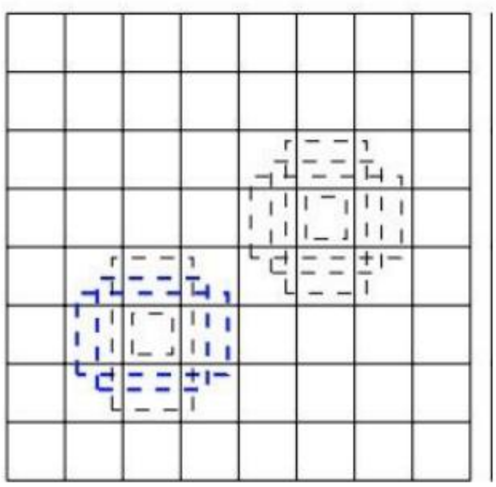


Figure 2.3. Figure of output of Single shot detectors

A. System Design

This framework is created utilizing different advances, which cooperate to flawlessly run Third Eye.

The framework configuration thinks about the accompanying equipment parts:

• Raspberry Pi 3

• Sonar sensors

• Camera module

• Headset

The framework configuration thinks about the accompanying programming parts:

• Python programming language Raspberry Pi working framework and programming stage

• OpenCV

• Text to Discourse Module

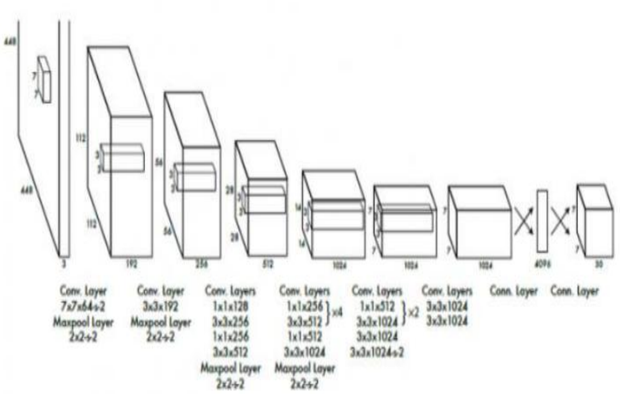


Figure 2.4. Figure of model of single shot detectors

B. The Raspberry Pi:

The Raspberry Pi is a minimal expense, versatile PC that can utilize a typical television or a PC screen as its presentation and standard console and standard mouse as info gadgets. It upholds different programming dialects, for example, python, java, scratch, and so on. It can play out any errand that one would anticipate from a personal computer. It tends to be utilized to peruse the web and to play top-quality video.

It also can create spreadsheets, perform word processors, and play games. Additionally, the Microcontroller has a feature that standard desktop PCs do not. It can communicate with people and things outside of it. It has been applied to several projects of various kinds, including weather stations and music machines.

The Raspberry Pi architecture has undergone numerous iterations, each of which has modifications in the type of Processor, memory capacity, network connectivity, and compatibility for peripherals.

Figure 2.5. Raspberry Pi 4 Model B

C. Sonar Sensor:

A sonar detector is a tool that uses sound waves to estimate distances to objects. The detector waits for the audio wave to return after already being emitted at a certain frequency from the item. It calculates the distance here between the sonar sensor as well as the item by measuring the delay during emission and reception. However, there are some things that the sonar device might detect. This is due to the possibility that a waveform may occasionally fail to return adequately to the sensor and travel elsewhere. Sonar sensors will not be able to precisely forecast the distance in this situation.

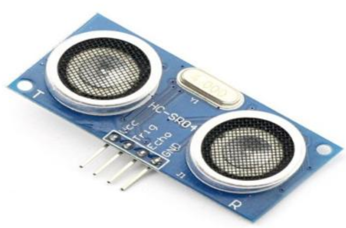


Figure 2.6 Sonar Sensor

D. Camera

The microcontroller pictures were taken by this module. Depending on the application and use, the frequency of image capture could be changed. This module has an 8mp Sony IMX219 sensor. Additionally, it could be used to record high-quality videos. It is also capable of 1080p30, 720p60, and VGA90 video modes and supports low-light clicking.

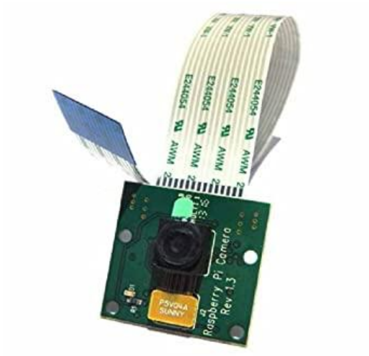


Figure2.7 Raspberry Pi Camera

III. USE CASE

A utilization case graph is a chart that portrays the connections among use cases and entertainers inside a framework. Using a case graph gives an outline of all or part of the framework or association using prerequisites as a fundamental model or a plan of action. It likewise permits one to impart the extent of an improvement project.

In this framework, there is just an entertainer named Client. The client will want to utilize every one of the capabilities or contributions of this framework.

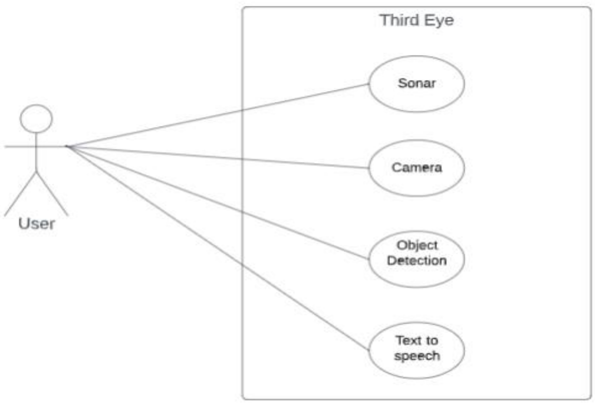


Figure 3.1 Use Diagram

IV. SYSTEM ARCHITECTURE

A framework design is the reasonable model of the task that characterizes the construction, conduct, and functionalities. This part centres around framework engineering and makes sense of the usefulness of each section of the framework. It additionally communicates how sensors speak with one another and cooperate to give the ideal result

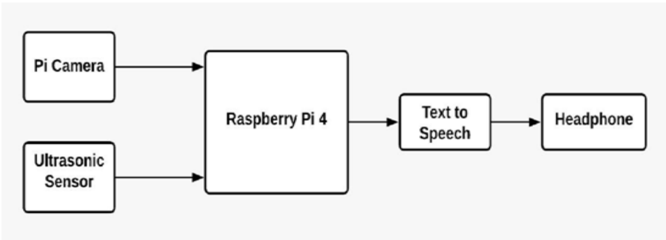


Figure 4.1. System Architecture

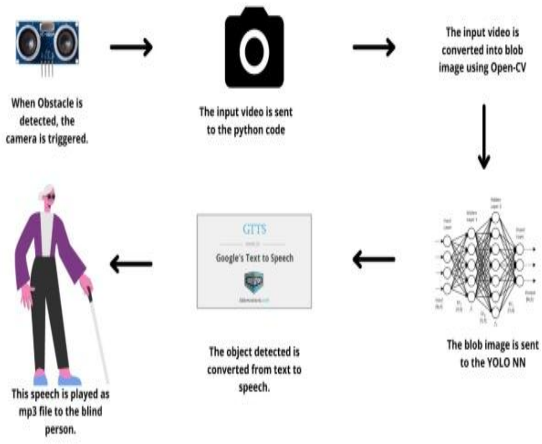


Figure 4.2 Flow Chart

The camera is referred to as this system's eye. It collects still photos continually and sends them to the Raspberry Pi microcontroller. These photograph are taken by the raspberry pi microcontroller, which uses AI image processing algorithms to process them. A corresponding answer is produced and sent to the module. The AI model that creates the Cafe model for the Raspberry Pi is created using the OpenCV framework. To switch the communication connecting the model and the put in, a Python application is utilized to collect data from the camera. The output is then shown as text on the screen. The text-to-speech module takes the user's input as text and changes it into an audio output that is played back to them. Infrared sensors are positioned to mirror a human's angle of view. They are employed to determine the separation between a user and a far-off item. The downward-facing sonar sensors are positioned at a 30-degree angle from the vertical plane. This 30-degree inclination aids sonar sensors in searching for things put on the ground close to the user. They can therefore be used to determine how far an object is from the user's legs if it is on the ground. The Raspberry Pi receives the data from both sensors, analyses it, and then delivers the preferred output to the text-to-speech module. Once more, the text is used as the input for this conversion. Once more, the text is used as the input and is transformed into an auditory response that is delivered to the user through a headset. The entire process continues until the user decides to turn the gadget off simultaneously and constantly.

V. SPEECH SYNTHESIS ON RASPBERRY PI

GTTS (Google Text-to-Discourse), a Python library and CLI device to communicate with Google to interpret the text-to-discourse Programming interface. Compose spoken mp3 information to a document, a record-like item (byte string) for additional sound control, or stdout. Or on the other hand pre-create, Google Interprets TTS demand URLs to take care of to an outside program.

VI. IMPLEMENTATION

The framework is made from three unique parts specifically: sonar sensor, infrared sensor, and camera. Every one of the information from each of the three parts is shipped off raspberry pi which process it and gives the result to the client. The design, execution, and carrying out are comparative for sonar and infrared sensors. Camera deals with an alternate design and calculation. The highlights of the framework include:

- recognize the distance of the article (sonar sensors)
- Recognize the sort of the article (utilizing object recognition calculation)
- Pass the result on to the client through earphones.

The accompanying segments elaborate on the total interaction. Recognize the distance of the item (Sonar Sensor)

The result of the sonar sensor should be visible progressively on the screen at genuine distance esteem from the article.

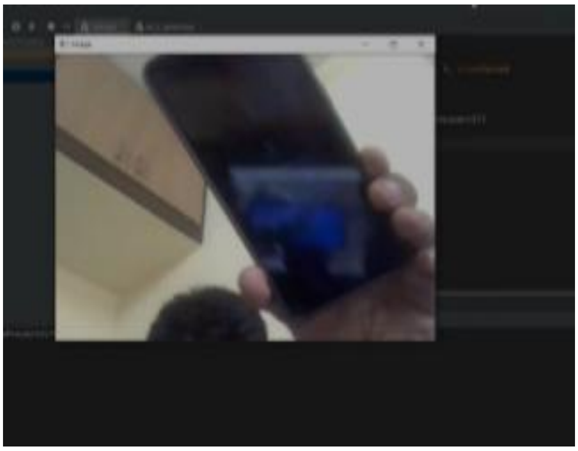


Figure 6.1 Output of Sonar

The above yield display shows the space of the client from the item like clockwork. Simultaneously, this data is being handled by the raspberry pi and sent to the client through earphones. This guides the client about the place of the article.

Variation in the distance

The sonar sensors are generally not exact as there is no optimal condition in the real world. There may be a variation between the distinguished distance and the real distance. These varieties can be diminished by utilizing better and more precise sensors.

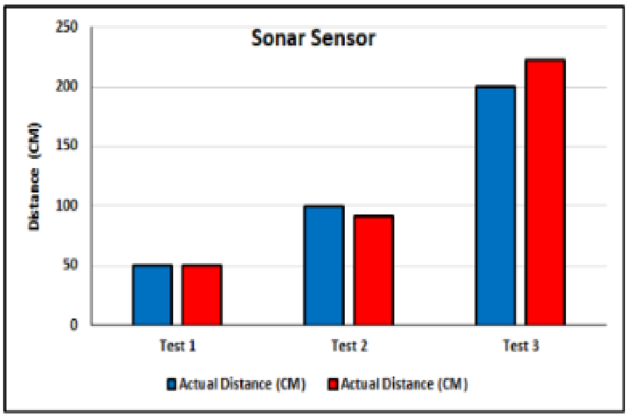


Figure 6.2 Actual versus Detected Distance through Sonar Sensors

VII. OBJECT DETECTION

One computer vision technique that enables us to spot and place things in an picture or tape is object identification. With this type of classification and restriction, object discovery will be able to

ISSN: 2094-0343
2326-9865

tally the items, determine where they are located, and follow those locations while correctly describing them. Picture acknowledgment awards an image with a mark.

Models for item recognition based on profound learning typically comprise two elements.

A photo is taken by an encoder as information, which is then put through a series of blocks and layers to determine how to take out the information needed to locate and identify items. yields are passed from the encoder to the decoder, which forecasts the bounding boxes and names for every item.

VIII. YOLO ALGORITHM

YOLOv3 (You Just Look Once, Variant 3) is an constant thing recognition estimate that distinguishes precise articles in recordings. Consequences are damned purposes highlights advanced by a profound convolutional brain organization to identify an item.

The main adaptation of Just go for it was made in 2016, and variant 3, which is examined broadly in this article, was made two years after the fact in 2018. YOLOv3 is a superior adaptation of Consequences be damned and YOLOv2. Just go for it is executed utilizing the Keras or OpenCV profound learning libraries.

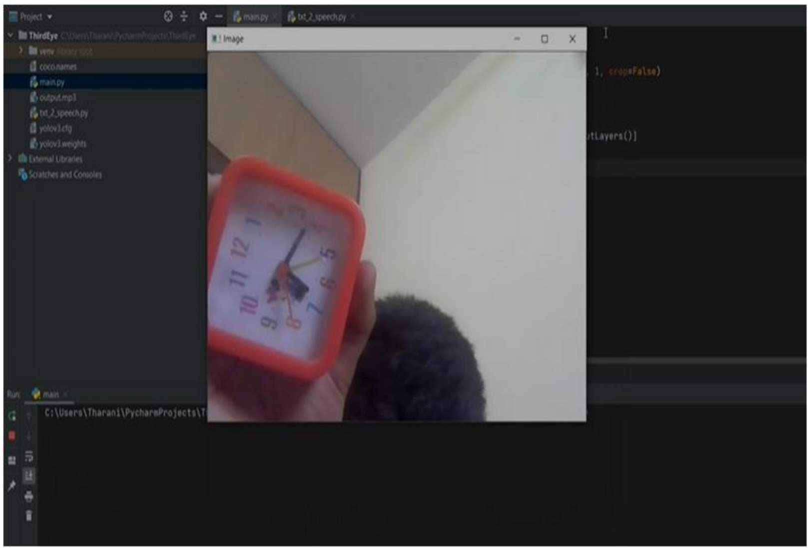


Figure 8.1 Output Detection

IX. IMPLEMENTATION

Circuit of the framework is planned and carried out so that every one of the parts works all the while without slowing down one another. Current to sonar sensors is controlled utilizing resistors and every one of the parts are painstakingly positioned to make it as little as could really be expected and to give handiness.

Figure9.1 Third Eye Module

According to the design outlined in this part, the Third Eye is created and integrated. Raspberry Pi, sonar, ultrasonic, and video sensors are used as the system's hardware foundation. The sonar sensor and camera are two independent but functional parts of the system. The sonar sensor is managed via Python. Python programmers may operate the camera and carry out object detection using the OpenCV framework for object detection.

X. RESULT AND DISCUSSION

This paper has discussed the outcome of wearables used for the visually impaired for navigation. The demand for this project will be huge due to its advantages like affordable price, easy to use, and its efficiency.

XI. PROGRAM

```
import cv2
import numpy as np
from gtts import Gtts
import os

mytext = "There is a"
language = 'en'


cap = cv2.VideoCapture(0)
whT = 320
confThresold = 0.5
nmsThreshold = 0.3

classesFile = 'coco.names'

with open(classesFile, 'rt') as f:
    classNames = f.read().rstrip('\n').split('\n')

modelConfiguration = 'yolov3.cfg'
modelWeights = 'yolov3.weights'
```

```
net = cv2.dnn.readNetFromDarknet(modelConfiguration, modelWeights)
net.setPreferableBackend(cv2.dnn.DNN_BACKEND_OPENCV)
net.setPreferableTarget(cv2.dnn.DNN_TARGET_CPU)


def findObjects(outputs, img):
    Ht, Wt, Ct = img.shape
    bbox = []
    classIds = []
    confs = []

    for output in outputs:
        for det in output:
            scores = det[5:]
            classId = np.argmax(scores)
            confidence = scores[classId]
            if confidence > confThresold:
                w,h =int(det[2]*Wt), int(det[3]*Ht)
                x, y = int((det[0]*Wt) – (w/2)), int((det[1]*Ht) – (h/2))
                bbox.append([x,y,w,h])
                classIds.append(classId)
                confs.append(float(confidence))

    indices = cv2.dnn.NMSBoxes(bbox, confs, confThresold, nmsThreshold)
    for i in indices:
        i = i[0]
```

```
    box = bbox[i]
    x, y, w, h= box[0],box[1], box[2], box[3]
    cv2.rectangle(img, (x,y), (x+w, y+h), (255,0,255), 2)
    cv2.putText(img, f'{classNames[classIds[i]].upper()}  {int(confs[i]*100)}%',
(x,y-10), cv2.FONT_HERSHEY_SIMPLEX, 0.6, (255,0,255), 2)


    mytext = 'There is a' + classNames[classIds[i]] + 'near to you'
    myobj = Gtts(text=mytext, lang=language, slow=False)

    myobj.save("output.mp3")

    # Play the converted file
    os.system("start output.mp3")


while True:
  success, img = cap.read()

  blob = cv2.dnn.blobFromImage(img, 1 / 255, (whT, whT), [0, 0, 0], 1, crop=False)
  net.setInput(blob)

  layerNames = net.getLayerNames()
  outputNames = [layerNames[i[0] – 1] for i in net.getUnconnectedOutLayers()]

  outputs = net.forward(outputNames)
```

```
findObjects(outputs, img)

cv2.imshow('Image', img)
cv2.waitKey(1)
```

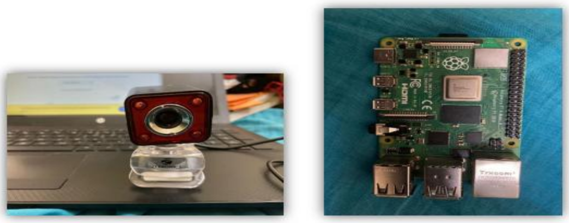
XII. CONCLUSION AND FUTURE ENHANCEMENT

Currently, a handicap of any sort for any individual can be hard and it is a similar case with visual impairment. Blind individuals are for the most part left oppressed. Giving a dream to a visually impaired person is extremely challenging. In this document, another man-made intelligence-based framework called Third Eye to manage the route of a visually impaired individual has been projected and created.

This man-made intelligence-based framework offers a straightforward electronic direction inserted vision framework which is configurable and effective. The framework assists blind and outwardly hindered individuals with being exceptionally self-subordinate by helping their versatility paying little mind to where they are; open air. consequences demonstrate that every one of the sensors works appropriately and furnish precise readings, however, the scope of the model sensors isn't elevated. Object recognition calculation uses a 100 percent computer processor which makes the raspberry pi hot and hence, in prospect frameworks, two raspberry pi are prescribed to be utilized; one for object discovery and one for every one of the sensors.

XIII. HARDWARE AND SOFTWARE

HARDWARE:



SOFTWARE:

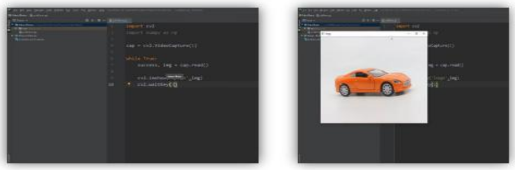


Figure 13.1 Hardware and software

XIV. REFERENCES

[1] N. Ambady and R. Rosenthal, "Slight Cuts of Expressive Way of behaving as Indicators of Relational Results," A Meta-Investigation: Mental Bull., vol. 111, no. 2, pp. 256-274, 1992.

[2] J. Ang et al., "Prosody-Based Programmed Discovery of Irritation and Dissatisfaction in Human-PC Exchange," Proc. Eighth Int'l Conf. Communicated in Language Handling (ICSLP), 2002.

[3] A.B. Ashraf, S. Lucey, J.F. Cohn, T. Chen, Z. Ambadar, K. Prkachin, P. Solomon, and B.J. Theobald, "The Agonizing Face: Agony Demeanor Acknowledgment Utilizing Dynamic Appearance Models," Proc. 10th ACM Int'l Conf. Multimodal Connection points (ICMI '07), pp. 9-14, 2007.

[4] T. Athanaselis, S. Bakamidis, I. Dologlou, R. Cowie, E. DouglasCowie, and C. Cox, "ASR for Close to home Discourse: Explaining the Issues and Upgrading Execution," Brain Organizations, vol. 18, pp. 437-444, 2005.

[5] A. Austermann, N. Esau, L. Kleinjohann, and B. Kleinjohann, "Prosody-Based Feeling Acknowledgment for MEXI," Proc. IEEE/RSJ Int'l Conf. Wise Robots and Frameworks (IROS '05), pp. 1138-1144, 2005.

[6] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias, "Feeling Examination in Man-Machine Communication Frameworks," LNCS 3361, pp. 318-328, 2005.

[7] R. Banse and K.R. Scherer, "Acoustic Profiles in Vocal Feeling Articulation," J. Character Social Brain Research, vol. 70, no. 3, pp. 614-636, 1996.

[8] M.S. Bartlett, G. Littleworth, P. Braathen, T.J. Sejnowski, and J.R. Movellan, "A Model for Programmed Acknowledgment of Unconstrained Facial Activities," Advances in Brain Data Handling Frameworks, vol. 15, pp. 1271-1278, 2003.

[9] M.S. Bartlett, G. Littleworth, M. Forthright, C. Lainscsek, I. Fasel, and J. Movellan, "Perceiving Look: AI and Application to Unconstrained Way of behaving," Proc. IEEE Int'l Conf. PC Vision and Example Acknowledgment (CVPR '05), pp. 568-573, 2005.

[10] M.S. Bartlett, G. Littleworth, M.G. Plain, C. Lainscsek, I. Fasel, and J. Movellan, "Completely Programmed Facial Activity Acknowledgment in Unconstrained Way of behaving," Proc. IEEE Int'l Conf. Programmed Face and Signal Acknowledgment (AFGR '06), pp. 223-230, 2006.

[11] A. Batliner, K. Fischer, R. Hubert, J. Spilker, and E. Noth, "How to Track down Inconvenience in Correspondence," Discourse Comm., vol. 40, pp. 117-143, 2003.

[12] A. Batliner et al., "You Moronic Tin Box — Kids Communicating with the AIBO Robot: A Cross-Semantic Profound Discourse," Proc. Fourth Int'l Conf. Language Assets and Assessment, 2004.

[13] C. Blouin and V. Maffiolo, "A Concentrate on the Programmed Location and Portrayal of Feeling in a Voice Administration Setting," Proc. 10th European Conf. Discourse Comm. also, Innovation (INTERSPEECH '05), pp. 469-472, 2005.

[14] S. Burger, V. MacLaren, and H. Yu, "The ISL Meeting Corpus: The Effect of Meeting Type on Discourse Style," Proc. Eighth Int'l Conf. Communicated in Language Handling (ICSLP), 2002.

[15] C. Busso et al., "Investigation of Feeling Acknowledgment Utilizing Looks, Discourse and Multimodal Data," Proc. 6th ACM Int'l Conf. Multimodal Points of interaction (ICMI '04), pp. 205-211, 2004.

[16] G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Paouzaiou, and K. Karpouzis, "Displaying Naturalistic Full of feeling States through Facial and Vocal Demeanor Acknowledgment," Proc. Eighth ACM Int'l Conf. Multimodal Points of interaction (ICMI '06), pp. 146-154, 2006.

[17] Y. Chang, C. Hu, and M. Turk, "Probabilistic Articulation Investigation on Manifolds," Proc. IEEE Int'l Conf. PC Vision and Example Acknowledgment (CVPR '04), vol. 2, pp. 520-527, 2004.

[18] Y. Chang, C. Hu, R. Feris, and M. Turk, "Complex Based Investigation of Look," J. Picture and Vision Processing, vol. 24, no. 6, pp. 605-614, 2006.

[19] Y. Chang, M. Vieira, M. Turk, and L. Velho, "Programmed 3D Look Examination in Recordings," Proc. IEEE Int'l Studio Investigation and Demonstrating of Appearances and Motions (AMFG '05), vol. 3723, pp. 293-307, 2005.

[20] L.S. Chen, "Cooperative Handling of General media Data for the Acknowledgment of Profound Articulations in Human-PC Collaboration," Ph.D. paper, Univ. of Illinois, Urbana-Champaign, 2000.

[21] L. Chen, T.S. Huang, T. Miyasato, and R. Nakatsu, "Multimodal Human Inclination/Articulation Acknowledgment," Proc. IEEE Int'l Conf. Programmed Face and Signal Acknowledgment (AFGR '98), pp. 396-401, 1998.

[22] J.A. Coan and J.J.B. Allen, Handbook of Feeling Elicitation and Evaluation. Oxford Univ. Press, 2007.

[23] I. Cohen, F. Cosman, N. Sebe, M. Cirelo, and T.S. Huang, "SemiSupervised Learning of Classifiers: Hypothesis, Calculations, and Their Applications to Human-PC Cooperation," IEEE Trans. Design Investigation and Machine Knowledge, vol. 26, no. 12, pp. 1553-1567, Dec. 2004.

[24] L. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang, "Look Acknowledgment from Video Arrangements: Transient and Static Displaying," PC Vision and Picture Getting it, vol. 91, nos. 1-2, pp. 160-187, 2003.

[25] J.F. Cohn, "Groundworks of Human Figuring: Look and Feeling," Proc. Eighth ACM Int'l Conf. Multimodal Points of interaction (ICMI '06), pp. 233-238, 2006.

[26] J.F. Cohn, L.I. Reed, Z. Ambadar, J. Xiao, and T. Moriyama, "Programmed Examination and Acknowledgment of Temple Activities and Head Movement in Unconstrained Facial Way of behaving," Proc. IEEE Int'l Conf. Frameworks, Man, and Robotics (SMC '04), vol. 1, pp. 610-616, 2004.