Aeye - A Smart Glasses

Abhash Kumar Jha[1], Pranay Jain[2], Advaith Shankar [3], Shreenivas B[4]
Dept. of Telecommunication Engineering
BMS College of Engineering
Bangalore, India

Abstract - The sight impaired face a slew of challenges when it comes to getting around. The only options they have are a draconian era walking staff and a guiding dog, both of which have severe drawbacks. The downsides of using a walking stick include the fact that it is unwieldy and does not allow the user to move around with much fluidity. It also becomes a hindrance when the roads the user is travelling on are unevenly laid out and have cracks and potholes, among other things. The guide dogs on the other hand are expensive and difficult to maintain. As a result, we want to provide a cost-effective and intelligent solution to this problem by offering 'A-eye,' a wearable smart glass. For the blind in contemporary culture, this is a convenient, cost-effective, and technologically advanced choice. In a nutshell, the 'A-eye' smart glasses will have a camera mounted on the device's frame that will continuously watch items in front of the user, as well as an IR sensor to determine the distance between the object and the user. These observations will be turned to voice to allow the user to move around with more ease.

Keywords- AEye, Blind, Ultrasonic sensors, ESP32 CAM, Object Detection, Single Shot Detector (SSD)

1. INTRODUCTION

Currently, in India alone, there are approximately 12 million blind people which constitutes around 33% of the world's entire population. Moreover, this is increasing by the day. Blind people go through many restrictions to their movement, and the only assistant device they have been using is a walking stick. The walking stick compared to the modern age of science is critically draconian. While there are guide dogs, but their problem is the fact that they have a lot of monetary requirements in the form of day-to-day maintenance, including the cost to acquire making them expensive. The walking stick also is presented with its own set of problems. They are not helping the blind section of society to keep their movements as fluid as we hope they would. Furthermore, if encountered with any uneven surface or any cracks on the surface the user is moving on, there might be physical risks involved with using a walking stick. Hence, we have designed a piece of equipment that is lightweight, affordable, and also relieves the user of using a walking stick. With the help of AI and object detection, we have produced the product- 'AEye- shades to aid the blind.' This equipment is user-friendly and light, making it extremely easy to wear and use. This wearable equipment is in the form of any other glasses available in the market. However, the uniqueness of this product is the fact that it is equipped with a camera that continuously monitors objects in front of the user. It is also equipped with two ultrasonic sensors on either side of the face to obtain an accurate reading of the distance of the objects. The equipment is connected to a processor which computes all the values and hence provides the observations in an audio format to the user or wearer. The foundation of our project is that of object detection- a technique that allows computers to identify defined objects within an image pulled from a live feed. Bounding boxes are made around the detected objects through the output from the deep learning model concocted on the processor. The project is affordable, hassle-free, and counters all the disadvantages of the walking stick.

This paper will first review the previous works along the parallel lines and present their findings in the literature survey section of paper. It will go on to give a brief overview about the methodology of our system in proposed system section. We will not go into much depth but enough to get a lucid background of each of the parameters covered in our work. The fourth section of the paper will give out the proposed work and consequently the paper will present the testing of the model and provide experiment results.

2. LITERATURE SURVEY

Object detection and combining them with wearable devices such as glasses have been an arduous task for a very long time. There have been several attempts to aid the blind section of society, but they have done so only for material purposes. Many devices and equipment were hailed which gave the visually impaired the ability to read text out to them but none of these devices can be used for traveling and moving about. Our project draws along the same line but we attributed the use of object detection technology to make a wearable device that the blind can comfortably use.

The authors of [7] have designed an architecture for object detection by getting inspired by the idea of the working of radars. In the case of this technical paper, the components used were the ESP-32 camera module, the Ultrasonic sensor, and the Arduino UNO was used as the microprocessor. The design included that of a servo motor that is connected to the ultrasonic sensor to act as a rotating radar. The ultrasonic sensor emits out ultrasonic waves in various directions and this wave is then again detected by the sensor. The readings and parameters of the reflected ultrasonic wave are computed by the Arduino UNO microcontroller which gives us the readable values of the distance, angle, and position of the objects around the device. The readings and parameters from the ultrasonic sensor are transferred to the Arduino UNO serially. The results observed by the author were extremely accurate hence making object detection simple and inexpensive.

[8] Describes another method of implementation of smart glasses without the use of AI-powered object detection but

rather using about 10 VL53L1X Time-of-Flight (ToF) sensors which are mounted around the wearable glasses. The paper's project and method of implementation involve the user to be wearing a pair of smart glasses that are equipped with Time-of-flight (ToF) sensors and a belt that consists of an ESP- WROOM-32 microcontroller. The wearable glasses consist of another WROOM-32 microcontroller. The sensors detect an object and send the observation to the microcontroller equipped within the glasses. Using Bluetooth, the observations are then transmitted to the second microcontroller in the wearable belt which computes the data and helps the user move by signalling the vibration motors arranged in a square grid.

[4] Utilizes object detection and identification in their proposed architecture. It presents a system that finds the important characteristics of a certain object to reduce the complexity of the overall system. Although the authors of this paper have used the SIFTS key point extraction, the drawbacks were evident in the paper. It did not consider the relationship of many objects for understanding scenes or for detecting everything that belongs to a certain place or location. The test results of this technical paper concluded that the authors received high accuracy in the feature extraction and the object detection but they also happen to present few cases where the model failed. The failure in their detection was due to the low image quality and the target frame being smaller than the original size of the object. Moreover, if the object detected contained some texture and if it did not have enough feature points for detection, the system failed to recognize the object.

The authors of [1] suggested a system using OpenCV in python to detect the objects in real-time along with a real-time text reader. The object detection was done using the single-shot multi-box detector which is what we are implementing in our project as well. The voice was outputted using a speech synthesizer and the synthesized speech can be formed by joining certain fragments of recorded voice and speech which will already be stored in a database on a server. The results obtained were very satisfactory and the speech was audible and understandable as well. While the idea of our project is almost the same, we have proposed a device that produces a stimulus that can be felt by the user. This aspect has been proved highly inefficient with low precision for [1]. Another facet that isn't covered by it is the identification of their surroundings. Our model can calculate the distance to the object providing a real feel to the object even in case of an inaccurately detected object.

[6] Designed a system to assist the visually impaired to travel independently without any external aid and monitors the real-time location of the individual wearing the device. It also detects the sudden fall and notifies the user's relative in such a fall incident. The model consisted of a set of ultrasonic sensors for distance, a PIR motion sensor for obstacle detection, an accelerometer for fall detection, and an endpoint smartphone application for location-based services. The system had proved to obtain an accuracy of 98.34% for an obstacle 50cm away.

[2] implicates the use of a new and upcoming technology of Bluetooth low energy (BLE) and the ultra-wideband (UWB)

beacons to help a visually challenged person move around with higher fluidity within a closed infrastructure (ideally the person's home). It suggests installing several beacons that work by using the UWB technology around the home. The architecture works by sending signals from the beacons called anchors to a wearable (worn by the person) which is UWB compatible to locate the person that is accurate to a distance of approximately 10cm. The anchors then determine the location of the person and translate the data to a server set up which further communicates the observations to an application on the user's phone. The observations are converted to speech with the help of the Google Cloud Text-To-Speech and hence the speech is read out to help the user move about. The only drawback of this is the constant need for internet connectivity on the smartphone where the application is installed for it to work seamlessly.

[5] Introduces a technique where GPS navigation can be used for helping visually challenged individuals to move about in an outdoor environment but as all the GPS techniques are currently computer-centric the authors propose a hybrid methodology. The authors explore the idea of integrating the current wayfinding techniques with social networking websites such as SoNavNet where users share their individual experiences for the navigation of a particular location to make the methodology more effective and efficient.

The Envision AI glasses [3], is an extremely technologically advanced device that uses the technology same as this project. It is AI-powered spectacle that assist the visually impaired in providing text detection as well as facial recognition to detect the people in front of the visually impaired. The device also has an application that is supportive and can be used by blind and as well as the sighted people.

3. PROPOSED SYSTEM

After analysing the problem set, we have taken up, we aimed to build an efficient, inexpensive, and simple to use device or equipment to assist the visually impaired section of our society to help them move with more fluidity and with no hassle by combining modern techniques of real-time object detection technology with modern techniques of compressing and compiling equipment. The device we intended to build is lightweight and easily wearable as it is in the form of glasses hence making it compact.
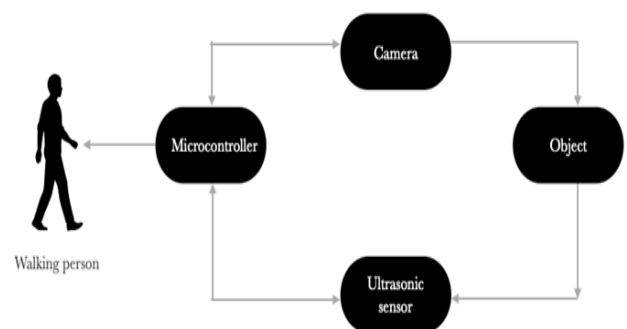

Fig. 3.1. Proposed architecture of the system

As shown in figure 3.1, there are 3 main components, i.e., Microcontroller, Camera, and Ultrasonic sensors. On

moving from right to left of the flow diagram, we can see that the distance of the object is being detected by the ultrasonic sensors, and the object detection is done by the camera module. On combining the outputs of both, we obtain a result that is commuted to the end-user, by a master controller, in an audio form. The controller is responsible for linking all the peripherals and text-to-speech conversion.

4. IMPLEMENTATION
4.1. Hardware Implementation

Circuit and Schematic Designing

The circuit of the proposed model is designed to be as simple as possible. The schematic of the work is designed using an electronic design automation software, KiCad 5.1.10. This software can be used in designing anything from a basic electronic circuit to a complicated PCB layout of an electronic design. The schematic of our work is shown in Figure 4.1.
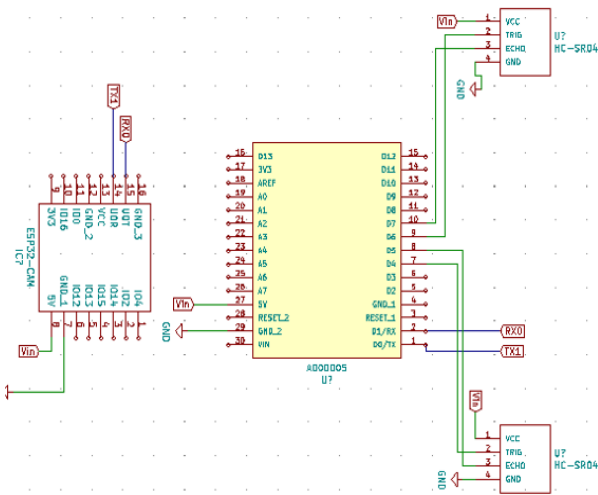

.

Fig. 4.1. The electronic circuit system

As shown in figure 4.1, Arduino Nano is connected to two HC-SR04 ultrasonic sensors and one ESP32 CAM module. Two ultrasonic sensors are placed in the right and left side of the device corresponding to two human eyes. The TRIG and ECHO pins of the right ultrasonic sensor are connected to the Digital I/O pins D4 and D5 respectively. On the other hand, TRIG and ECHO pins of the left ultrasonic sensor are connected to the Digital I/O pins D6 and D7 respectively. The distance of an object is detected using SONAR technique and is obtained at pins D5 and D7 for right and left sensors respectively.

The second major section of the schematic is the ESP32 CAM module, which is serially connected to the master controller, Arduino Nano. The Rx/Tx of the master controller is connected to Tx/Rx UART ports of the camera module which enables serial connection between the two controllers. Arduino Nano is powered by an external battery and all other peripherals (ESP32 CAM, HC-SR04) are being powered by the 5V power supply port available in the master controller.

Printed Circuit Board (PCB) Designing

The 2-layer PCB (Printed Circuit Board) of the proposed model is designed to be as simple and compact as possible. The PCB of the work is designed using an electronic design automation software, KiCad 5.1.10. This software can be used in designing anything from a basic electronic circuit to a complicated PCB layout of an electronic design. The PCBs of our work is shown in Figure 4.2.
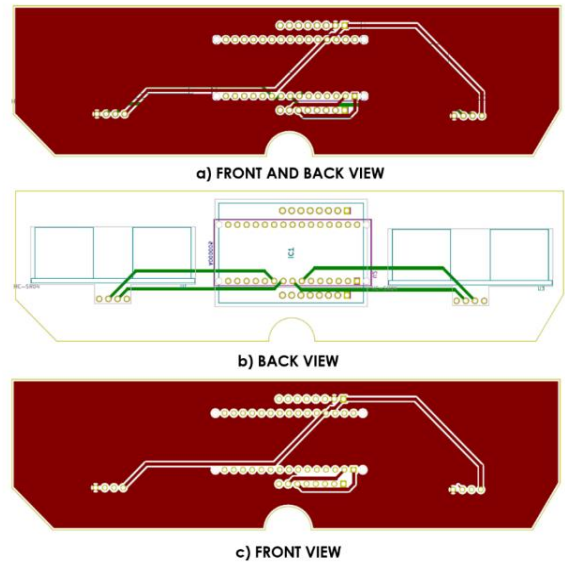

Fig. 4.2. PCB layouts

As shown in Figure 4.2, figure 4.2.a corresponds to the front and back view, figure 4.2.b corresponds to the back view and figure 4.2.c corresponds to the front view of AEye shades PCB layout. It can be observed that the sensor connections are done in the bottom layer of PCB, shown as green trace connecting two ports. On the other hand, all the power connections and serial connection between the two controllers are done in the top layer, shown by red traces.

4.2. Software Implementation

Controller application software

There are 2 controllers used in our design, i.e., Arduino Nano and ESP32 CAM. Arduino IDE is the platform used to code both the controllers. In order to operate on Arduino Nano, the processor is set to ATmega328P and the board is Arduino Nano for hardware configuration. Arduino Nano works at 9600 baud rate and uses the same rate for communicating to other devices such as ESP32 in our work. The ultrasonic sensors are connected to the digital pins as stated before and works on the given formula:

$$Distance = (Speed * Time)/2$$

In the above formula:
- Speed = 320 m/sec (Speed of sound)
- Time = Amount of time taken by the wave to come back at pin ECHO (microseconds).

- Distance = The distance between the object and the sensor.

On the other hand, ESP32 CAM module is also coded using the same platform, Arduino IDE. In order to operate on ESP32 CAM module, the partition scheme is set to Huge APP with a flash frequency of 40 MHz and the board is ESP32 Wrover Module for hardware configuration. The camera module works at Baud rate of 115200, but the communication between the controller is done at 9600 Baud rate. This hardware is powered by the 5V supply from the master controller and is serially connected to the Arduino Nano.

ESP32 essentially creates a http server on the Wi-Fi network it is connected to. A set of addresses are created on this local server that are used as endpoint from the main processing server to access the stream from the camera and getting the ultrasonic readings from Arduino via ESP32. 5 addresses are used to handle different type and resolution of the stream-cam-lo.jpg for getting 320 x 240 image, cam-hi.jpg for getting 800 x 600, cam.bmp for getting low quality bmp image, cam.jpg for getting a high-quality jpeg image and cam.mpeg to get the live stream of the video captured by the camera. The address getDistance is used to access the ultrasonic sensor's readings.

Main Processing Server

This sever is responsible for all the processing that is the object detection and the distance calculation. The server continuously hits the URLs (local IP/cam-hi.jpg and local IP/getDistance) and extracts the image and the ultrasonic sensor's readings. At each hit, it processes the image using the SSD MobileNet to get the bounding boxes, classes and accuracy scores for the objects within the image. Further, it extracts the top 2 classes having accuracy above 70% to present it to the user. Moreover, the distance from the lens for each of the two objects are calculated based on the formulation as depicted in the figure 4.3.

$Height\ of\ dog(in\ px) = abs(y_{max} - y_{min})$

$Average\ Real\ Height\ of\ dog = 1000\ mm$

$Focal\ length\ for\ ESP32\ cam = 3.6\ mm$

$Sesnsor\ height\ for\ ESP32\ cam = 2.7\ mm$

$$Distance\ to\ the\ dog(mm) = \frac{focal\ length(mm) \times Real\ Height\ of\ Dog(mm) \times Image\ Height\ (in\ px)}{Height\ of\ Dog\ (in\ px) \times Sensor\ Height(mm)}$$

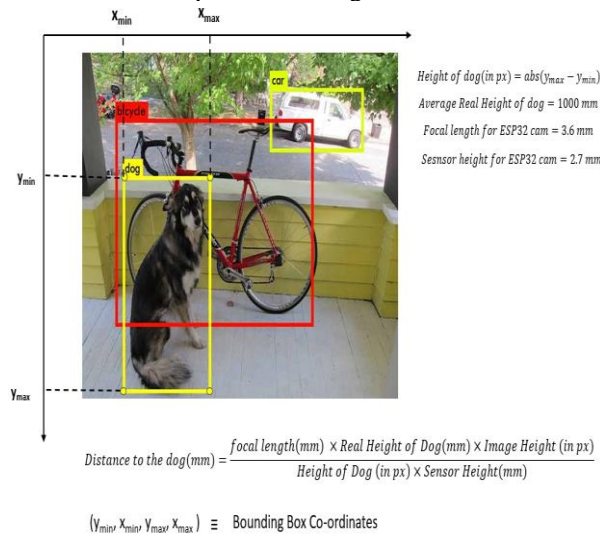$(y_{min}, x_{min}, y_{max}, x_{max}) \equiv$ Bounding Box Co-ordinates

Fig. 4.3 Distance calculation from the bounding boxes.

We have curated a list of average heights for different types of objects so that we can calculate the distance of the object from the camera lens. It is to be noted that these calculations yield only approximate distance as the orientation and shape of objects can differ vastly. Thus, it is to be noted that the readings from the ultrasonic sensor are used as a ground truth values which can be contrasted to the calculations done by the server. These observations are converted to the audio format using google text to speech API and saved onto the server. After every 50 successful detection (3 sec), the speech is played to the user.

5. RESULT

The outcomes of our work are as follows:

- The schematic of the system was successfully converted into a working prototype.

Fig. 5.1. Aeye prototype

- The PCB was designed in 2 ways- one that can be accommodated inside a compact box and the other was in shape of the glasses that will be feasible to wear.
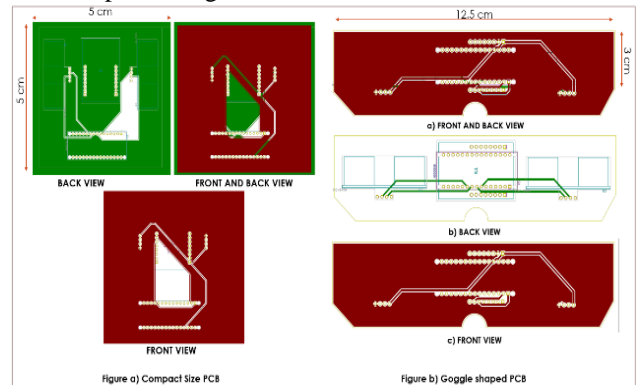

Fig. 5.2. PCB layouts (two models)

- The objects were detected at the server. It is to be noted that the upper limit for the accuracy was set to be 60%.

fig. 5.3. Object detection window

- Serial Communication between ESP32 and Arduino Nano was successful and the distance was logged in the command window at the main processing server.
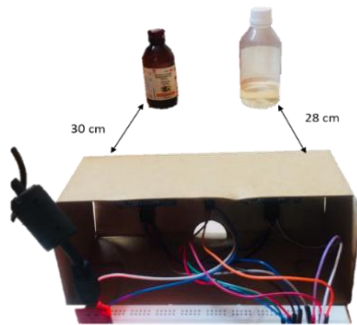

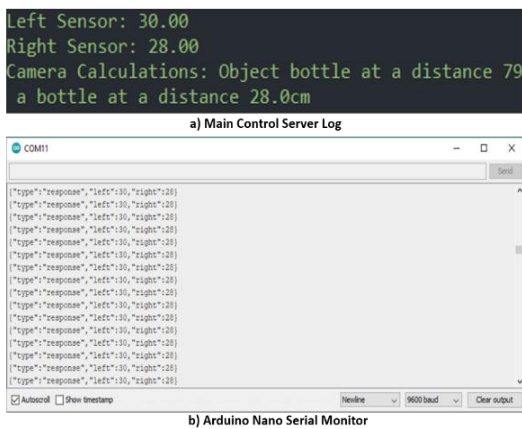Fig. 5.4. Object's distance from the Aeye prototype

Fig. 5.5. Serial communication verification

The distance calculated from the image is not as accurate as the size of the object can vary vastly. Take for example the bottle shown in the image. It is a small medicine bottle which is approximately 110 mm. Compared to the average length of bottle that is 304.8 mm, it is obvious that there would be a difference in the distance to object. If, however, the object was close enough the average height of the respective object, the calculations would fetch us nearly the same values. Coming to the frame rate and processing speed, there was a problem that occurred during the serial communication. The Arduino Nano sent the sensor readings every 1.5 seconds (1.5 seconds was the threshold set by us to reduce load on the http server) and the processing of the image from getting the image from http server to mapping the bounding box on the OpenCV application nearly took only 0.3 seconds. To compensate for the readings sent by Arduino Nano, we had to further put a 1.5 second sleep in the main processing server code to get a clean output. Moreover, after every 4-frame processed, the detection was converted into audio format to help the blind person hear his surroundings. Coming to the frame rate and processing speed, there was a problem that occurred during the serial communication. The Arduino Nano sent the sensor readings every 1.5 seconds (1.5 seconds was the threshold set by us to reduce load on the http server) and the processing of the image from getting the image from http server to mapping the bounding box on the OpenCV application nearly took only 0.3 seconds. To compensate for the readings sent by Arduino Nano, we had

to further put a 1.5 second sleep in the main processing server code to get a clean output. Moreover, after every 4-frame processed, the detection was converted into audio format to help the blind person hear his surroundings.

6. CONCLUSION

In conclusion, this project produced a light, affordable and effective device to help the visually impaired section of our society. The results showed that the 'A-Eye shades' assist user personnel to help move with utmost fluidity. Besides, this could be exploited as an optimum replacement for the arcane walking stick or the expensive guide dog option, thus, fulfilling the aim of our project. The model produced standard results as expected from the SSD mobile-net 320 model, processing a frame in 22 milliseconds and having mean average precision of 22.2. However, the bottleneck of the model was the serial communication over HTTP that delayed the response. Despite that, the model calculated the average distance from the camera to the object to a near precision given the object used was of standard size (near average height). Further, the readings from the ultrasonic sensors were highly accurate, with camera distance calculation to be used as a backup for the bad response from the server. For the future scope, the device can be used for various other purposes by adding on multiple features. A faster processor can be integrated with the prototype such as the Nvidia Jetson nano to increase the feed frame rate greatly and to make the device more efficient in terms of self-computing efficacy. The device can also be branched as a reading tool to help the visually impaired by integrating text detection with the model as seen in [9]. Another upgrade would be the installation of gyroscopes that are commonly used in a mobile phone as an orientation sensor. In an unfortunate case of fall, it can alert any medical, close acquaintances, and family. A structural upgrade can be also done to replace the current equipment on the chassis by replacing it with lightweight carbon fibre to increase the toughness and as well as making it extremely light.

REFERENCES

[1] Rajeshwari Kumar Dewangan, Siddharth Chaubey, "Object Detection System with Voice Output using Python," *2021 International Journal of Science & Engineering Development Research,* Vol. 6, Issue 3, pp. 15-20, 2021.

[2] E. Barri, A. Gkamas, E. Michos , C. Bouras, C. Koulouri, S.A.K. Salgado, "Text to Speech through Bluetooth for People with Special Needs Navigation," *International Conference on Ne8tworking and Services,* 2010.

[3] Envision, *Envision Glasses,* 2017.

[4] H. Jabnoun, F. Benzarti, H. Amiri, "Object detection and identification for blind people in video scene," *2015 15th International Conference on Intelligent Systems De- sign and Applications (ISDA),* pp. 363-367, 2015.

[5] H. A. Karimi, M.B. Dias, J. Pearlman, G. J Zimmerman, "Wayfinding and Navigation for People with Disabilities Using Social Navigation Networks," *EAI Endorsed Transactions on Collaborative Computing,* Vol. 1, Issue 2, e5, October, 2014.

[6] Rahman Mohammad Marufur et al. "Obstacle and fall detection to guide the visually impaired people with real time monitoring," *SN Computer Science,* Vol. 1, pp 1-10 (219), 2020.

[7] Divya P, Bhavana N and George M, "Arduino Based Obstacle Detecting System," *International Conference of Advance Research & Innovation (ICARI) 2020*.

[8] Košale U, Žnidaršic P, Stopar K, "Detection of different shapes and materials by glasses for blind and visually impaired," *2019 6th Student Computer Science Research Conference*, Vol. 57, 2019.