Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA-2020)
IEEE Xplore Part Number: CFP20J88-ART; ISBN: 978-1-7281-6387-1

Design and Implementation of Voice Assisted Smart Glasses for Visually Impaired People Using Google Vision API

P. Selvi Rajendran
Prof, CSE
Hindustan Institute of Technology and Science
Chennai, India
selvir@hindustanuniv.ac.in

Padmaveni Krishnan
Asso. Prof, CSE
Hindustan Institute of Technology and Science
Chennai, India
kpadmaveni@hindustanuniv.ac.in

D. John Aravindhar
Prof, CSE
Hindustan Institute of Technology and Science
Chennai, India
jaravindhar@hindustanuniv.ac.in

Abstract - Generally, visually challenged people tends to have difficulties in traveling and managing many kinds of challenges in their routine life. Mostly, wooden Sticks are used to sense barriers and obstacles next to them. As a result, visually impaired people cannot know exactly what kind of challenges they face and must thus rely entirely on lead sticks and training to navigate safely and in the right direction. This research work focuses on the development of a guidance system that uses smart glass paired with a sensor to continually capture images from the environment by the user wearable smart glass. The smart glass is equipped with a processor to process the captured images and objecst will be detected to inform the user about the results of the image and the user would have a much more comprehensive view of the method. This system allows visually impaired people not only to inform about traveling route and distance to the obstacle, but it also can inform about what the obstacle is. This smart glass can sense the distance from the obstacle and produce a warning to alert the user in advance. This application is developed to provide such a speech-based interface for the user, i.e. the user sends a voice that interprets his destination location when and when he is about to reach the destination. Here, instead of an alarm signal, the blind man can hear the location recorded by the user.

Keywords—Assistive device, Image Recognition, Arduino, Visually Impaired People, Smart Glass

I. Introduction

India is a world where vision disability is a big concern due to pervasive extremist violence and because of water and food poisoning these types of defects arisen for newborn babies. However, technical developments are required to help the people those are living in hard surroundings. As a whole, visually challenged people are gradually capable of performing everyday activities in their own way, such they can drive on the highways and they can move inside their homes independently. It is known that people with disabilities need the help of others to recognize items. Even then, the latest research has presented a variety of approaches to overcome the problems of visually impaired persons (VIPs) and to give them more mobility, but these methods have not sufficiently discussed safety measures when VIPs travel around their own. Moreover, many alternative systems are available but no method assists disabled people to be in constant touch with their family and relatives and is also typically complicated and costly.

The main challenges faced by blind people come into the context of managing indoor and outdoor conditions, comprising of numerous barriers and the awareness of the individual in front of them. It is difficult to distinguish items or individuals with only perceptive and audio knowledge. Smart, lightweight, less expensive, self-contained guidance and facial recognition systems are widely sought-after by people who are blind. This system will be helped for blind people to navigate in indoors and outdoors environments with the support of a Smartphone, this system is equipped with ultrasonic sensors and a GPS facility to provide location information.[1,2]. This proposed device has many features that provide the visually challenged people with a means of secure and autonomous movement and it has a feature to provide constant communication with their family members and caregivers who can monitor their location where they struck out. The device depends on cameras that collect photos to identify items at a distance from the user, warn the family and caregivers when the user comes to a standstill, and notify the family, caregivers, and others when the user wants assistance. This research proposes a simple mobile guidance system to resolve direction-finding problems for visually challenged people and also to reduce hurdles for helping visually impaired people to travel effectively from the starting location to reach the destination location with a better understanding of their environments.

II.RELATED WORK

Hsieh-Chang[3] has carried out an obstacle recognition research that is based upon applying knowledge to enable visually impaired people to escape obstacles while traveling through an unknown environment. The machine consists of three parts: the identification of situations, the identification of barriers, and the voice announcement. They reduced the over-segmentation problem by adding the ground plane. This approach solves the issue of segmentation by eliminating the edge and the original seed location issue for the area growth process by applying the Connected Node Process (CCM). Van-Nam [2] increased the precision of the procedure by sensing obstacles and localizing them and then transmitting information on obstacles to visually disabled people using various styles such as speech, tactile, vibration. They introduced the Visually Driven Assistance Device based on the electrode matrix and the handheld Kinect. The method consists of two key components: the processing and analysis of the context and the representation of information. The very first component tries to analyze the world with the use of mobile Kinect for recognizing the predefined obstacles for visually challenged persons, while the second aspect aims to portray obstructive data in the form of an electrode matrix.

Muhammad Asad[1] has implemented a methodology to find the direction among several straight paths in various environments in the real world scenario. These straight paths will have the corresponding description about the path when image has been captured, it will be appeared for converging to a predetermined location from the vanishing point. The proper isolation and practical simulation of the trapped frame contribute to the analysis of these parallel edges. The vanishing point is then measured and a decision-making mechanism is created which shows the visually challenged person his / her divergence from the straight line.

This proposed device has many features that provide the visually challenged people with a means of secure and autonomous movement and constant communication with their families and caregivers who can monitor their whereabouts. This proposed smart glass completely depends on a camera which is embedded in the front and center portion of the smart glasses. Here, cameras are used to collect photos of the environment which is in front of the blind people to identify the obstacles on the path. The control unit is responsible for sending the notification to the family members and caregivers when the user fails to detect the obstacle and face the problem. The information will be communicated to the family, caregivers, and others when the user wants assistance by sending the prerecorded voice message.

A. Maintaining the Integrity of the Specifications

1) SmartCane

SmartCane[4,16,19] is an interactive navigation assistance device that works like a white cane handle. Since white cane can only detect obstacles up to the height of some level of human, Smartcane complements its versatility by detecting obstacles from the knee to the height of the head. Electrical signals have been used to detect the obstacles which may create a harmful effect, and the position of obstacles is informed to the users as a vibrational pattern. All the components are connected with a battery such as a cell phone that can be used in both indoor and outdoor navigation modes. It was developed to accommodate various types of business grips that are widely used by visually disabled individuals.

2) Tap tapsee

TapTapSee [4, 16, 20] is a handheld camera technology developed exclusively for blind and visually disabled people, operated by the CloudSight Image Recognition API. Tap TapSee uses the camera and VoiceOver features of the smartphone to record a picture or video of something and to recognize it clearly for a blind person. In the screen, double-tap the top portion of your screen to take the photo, or double-tap the left side of your screen to take a recording. Tap TapSee will reliably visualize and describe any two or three-dimensional objects in any direction within seconds. The VoiceOver unit then talks about the ID loud.

3) Smart shoes

Smart shoes [4,16,21] are coordinated with the person's device, and the software that piggybacks on Google Maps enables the shoes to keep a record of where they're headed. When you arrive at your destination and pick a route, you can turn off your mobile and run or walk also with your left or right foot buzzing.

III. TALKING SMART GLASS EMBEDDED WITH COMPUTER VISION

This proposed wearable system includes a pair of glasses and an obstacle detection module to be mounted in the processing unit and an output screen. The output panel element is a beeping part that warns the consumer of the obstruction intrusion and the power supply. The obstacle detection module and the output panel assembly are attached to the processor unit. The energy supply shall be used to produce electricity for the central processing unit.

The ultrasonic sensor and is used to implement the control module. The control unit controls the ultrasonic sensors, which gather information on the obstacle in front of the person, evaluates the information, and sends feedback through the buzzer. The Ultrasonic Smart Glasses for Blind People are portable, easy-to-use, lightweight, user-friendly, and low-cost devices. These glasses can easily guide the blind and assist the blind.

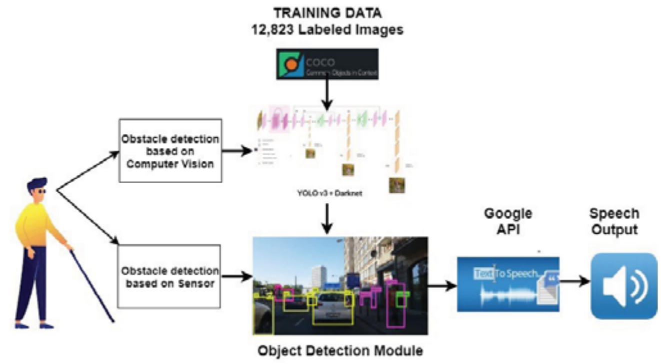

Figure 1. Architecture diagram of Designing of voice Assisted Smart Glasses

A. Training Data

The training data used for the model is taken from the Context Common Objects (COCO) dataset. The You Just Look Once (YOLO) v3 is used for programming and the value of the weights are obtained from the file of 200+mb.[7,8,9,10]

B. Model

The concept implemented used the You Just Look Once (YOLO) algorithm that operates in an extremely complicated Convolutional Neural Network architecture called the Darknet. Besides, the python cv2 module is used for setting up Darknet in the yolov3.cfg format.

C. Input Data

The sensor that is mounted in the smart glass is being used to transmit images to this equipped model. The transfer speed is thirty frames per second, and can only be programmed to handle any other frame to improve performance.

D. API

The class description of objects observed for each frame will be a set of sequence images, e.g. "car." The various position of the image such as top, left, right, center and bottom also received with the image and it will be added to the class predictor "auto." The text concept can then be sent to the Google Text-to-Speech API using the gTTS package. [7,8,9,10] for the transformation of text form into speech form to alert the blind person.

E. Output

As the output, the bounding box coordinates of every object will be captured within the frames, and it connects the boxes to the objects detected, and returns the stream of the frames as a video replay. After a few seconds, voice response on the first frame of the second (instead of 30 fps) will be generated e.g. "lower left cat"—meaning a cat being located at the bottom left of my camera view.

F. YOLO algorithm

You Just Look Once (YOLO) seems to be a real-time object detection algorithm [12,17], amongst the most effective object detection algorithms that also combine some of its most innovative ideas in the computer vision and machine learning environment. The detection of obstacles is a core feature of autonomous vehicle technologies. YOLO is a smart neural convolutionary network (CNN) for real-time object identification. This algorithm works by implementing the several layers of a neural network for processing the whole image, then divides the image into several small regions and determines bounding boxes and probabilities between each region is calculated. These bounding boxes are weighted according to the projection.

G. Darknet

There seem to be several different versions of the YOLO algorithm on the internet. Darknet is also an open-access neural network system. Darknet was developed in C Language and CUDA technology, which makes it very fast and allows for GPU computing, which is important for real-time predictions.

Installation is easy and takes only 3 lines of code to run (it is important to change the settings in the Makefile script after cloning the repository to use the GPU). After installation, use a pre-trained model or create a new one from the scratch.

H. Working Principle

Computer Vision Glass consists of two major hardware components, Raspberry Pi and a compact HD sensor. The camera serves as an input sensor, which collects the live video stream from the neighborhood and afterward feeds the brain, i.e. Raspberry Pi, which monitors the entire network, detects, and recognizes objects which are presented in the neighboring area.

The most important component, sensing, and identifying objects were developed. The project uses the YOLO algorithm to identify artifacts and devices that are constraining tools, such as Raspberry Pi or Smartphone.

IV. IMPLEMENTATION

The proposed system is implemented by combing the numerous technology stacks discussed below. Android Studio is mostly used for the development of applications because it is an Approved Integrated Development Environment (IDE) primarily developed for the various android applications [7].

OpenCV library is used for image recognition since it offers support for real-time applications [8].Python scripting language has been used to build a machine learning model [9]. TensorFlow software is used for machine learning applications.

It has a scalable architecture that makes it easy to spread computing efficiently across a range of platforms [10].

YOLO (You Only Look Once) algorithm has been used to analyze the video streams in real-time. It uses a particular convolution neural network that is stronger than the R-CNN (Regional Convolutional Neural Network) and enables the target to be projected in the form of bounding boxes [11].

Google Cloud is being used to train machine learning models as it provides high computing power at a low cost. It also allows for a huge quantity of data to be processed in the cloud and can be used later to construct a more reliable model [12].

The computer uses two modules to display objects and a text reader in real-time. The text reader reads data in real-time using that same method, which lets users, easily interpret menu cards in hotels, hotel room numbers, or even find their properties.

A. Dataset

For the testing purpose, the Generic Object in Context dataset is being used to train the You Just Look Once model and text reader. It is capable of recognizing eighty distinct groups.

V. RESULTS AND DISCUSSION

The application can recognize and consider various classes of hazards that may be detected while driving, ordinary objects and machinery, various types of vehicles, food, etc. It can be seen in the Table. 1.

Using this method, an obstacle can be found in a specific area. After the barrier has been found, the user will be navigated to their destination. That's why the rationale is to establish two borders on the left and right border of the ROI. Therefore, the ROI will be divided into three parts. If the barrier is in the right side then the user move to the left, and if the barrier is in the left side then the user to move to the right.

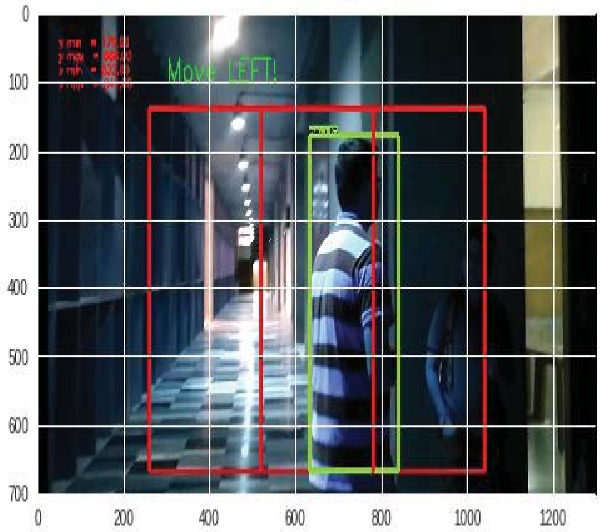

Figure 2. Obstacle Detection

The relation between the distance and the frequency of vibration is given in the below-mentioned table. It also includes the type of vibration which is provided as feedback to the user. Certainly, it has been known from the table values that as the distance from the obstacle to the user reduces, the vibration increases, users are alerted with the continuous vibration of the device.

Table 1. Experimentation of Obstacle Detection

Distance	Time	Vibration (count)	Vibration (sec)	Vibration (Type)
1m	1sec	1 time	2sec	continuous
2m	1sec	1 time	2sec	continuous
3m	1sec	1time	2sec	continuous
4m	1sec	1time	2sec	continuous
5m	1sec	2 times	1sec	short
10m	2sec	3 times	1sec	short
15m	2sec	4 times	1sec	short

The above table 1 shows the Obstacle detection with vibration indication. Here, the experiment is performed for varying distances from the range of one meter to fifteen meters. Also, as an optional feature, the user can alert the system to stop vibration by supplying through the voice-oriented command like "Stop Vibration" till the user has connection with the object or human. There is also a variation in the form of acceleration where the target is in the furthest distance to the nearest point. When the distance between the obstacle and the user decreases, and then the vibration type changes from short to continuous mode to indicate the continuous alert about the unsafe situations to the user. There is also the option to provide continuous feedback to the consumer on the variance of the system.

VI. . CONCLUSION

The device presented here is a smart glass that incorporates the functionality of a machine vision and obstacle detection and recognition sensor. It can be conveniently advertised and made accessible to the visually disabled population. It would also help to deter future injuries. Smart devices can be transported comfortably and the system camera can be used to track objects from the surrounding environment and display in audio format. So, allowing visually disabled individuals to 'see without the eyePre-process Image

REFERENCES

[1] Muhammad Asad, Waseem Ikram," Smartphone based guidance system for visually impaired person", 2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA), 15-18 Oct. 2012.

[2] Hsieh-Chang Huang, Ching-Tang Hsieh and Cheng-Hsiang Yeh," An Indoor Obstacle Detection System Using Depth Information and Region Growth", Sensors (Basel). 2015 Oct; 15(10): 27116–27141

[3] Ales Berger, Andrea Vokalova, Petra Poulová,"Google Glass Used as Assistive Technology Its Utilization for Blind and Visually Impaired People",Book Chapter,Springer International Publishing.

[4] Aditya Raj, Manish Kannaujiya, Ajeet Bharti, Rahul Prasad, Namrata Singh, Ishan Bhardwaj " Model for Object Detection using Computer Vision and Machine Learning for Decision Making " International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 43,March 2019.

[5] Selman TOSUN, Enis KARAARSLAN "Real-Time Object Detection Application for Visually Impaired People: Third Eye". IEEE Conferences 2018.

[6] Android Developer, https://developer.android.com/

[7] OpenCV, https://opencv.org/

[8] Python programming language, https://www.python.org/

[9] Tensorflow, https://www.tensorflow.org/

[10] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi "You Only Look Once: Unified, Real-Time Object Detection". Cornell University,Jun 2015

[11] Google "Google Cloud Services", https://cloud.google.com/

[12] P.Selvi Rajendran, "Virtual Bulletin Board using Man-Machine Interface (MMI) for Authorized Users", Indian Journal of Science and Technology, Vol.12, No.34, September 2019.

[13] P. Selvi Rajendran, "AREDAI: Augmented Reality based Educational Artificial Intelligence System", International Journal of Recent Technology and Engineering (IJRTE), Vol.8, No 1, May 2019.

[14] P.Selvi Rajendran, "Virtual Information Kiosk Using Augmented Reality for Easy Shopping", International Journal of Pure and Applied Mathematics (IJPAM),Vol. 118, No. 20, pp. 985-994, 2018.

[15] Suresh A., Arora C., Laha D., Gaba D., Bhambri S.,"Intelligent Smart Glass for Visually Impaired Using Deep Learning Machine Vision Techniques and Robot Operating System (ROS)", In: Kim JH. et al. (eds) Robot Intelligence Technology and Applications 5. RiTA 2017. Advances in Intelligent Systems and Computing, vol 751.Springer.2019.

[16] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788. 2016.

[17] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In European conference on computer vision, pp. 21-37. Springer, Cham, 2016.

[18] Mohamed Dhiaeddine Messaoudi , Bob-Antoine J. Menelas and Hamid Mcheick," Autonomous Smart White Cane Navigation System for Indoor Usage", Technologies 2020, 8, 37

[19] Deepti Patole, Sunayana Jadhava, Saket Gupta, Khyatee Thakkar," Touch Tap See: A Navigation Assistant for the Visually Impaired Person", January 2019 ,SSRN Electronic Journal

[20] Moaiad Khder ,"Smart Shoes for Visually Impaired/Blind People", November 2017Conference: International Conference on Sustainable Futures ICSF 2017At: bahrain