For(E)Sight : A Perceptive Device to Assist Blind People

Jajini M[1], Udith Swetha S[2], Gayathri k[3], Harishma A[4]

[1] Assistant Professor, Velammal College of Engineering and Technology
Viraganoor, Madurai, Tamil Nadu.
[2,3,4]Student, Electrical and Electronics Engineering, Velammal College of Engineering and Technology
Viraganoor, Madurai, Tamil Nadu.

Abstract- *Blind mobility is one of the major challenges faced by the visually impaired people in their daily lives. Many electronic gadgets have been developed to help them but they are reluctant to use those gadgets as they are expensive, large in size and very complex to operate. This paper presents a novel device called FOR(E)SIGHT to facilitate normal life to the visually impaired people. FOR(E)SIGHT is specifically designed to assist visually challenged people in identifying the objects, text ,action with the help of ultrasonic sensor and Camera modules as it converts them to voice with the help of RASPBERRY PI 3 and OPENCV. FOR(E)SIGHT helps the blind person to communicate with the mute people effectively, as it identifies the gesture language of mute using RASPBERRY PI 3 and its camera module. It also locates the position of visually challenged people with the help of GPS of mobiles interfaced with RASPBERRY PI 3. While performing the above mentioned processes, the data is collected and stored in cloud. The components used in this device are of low cost, small size and easy incorporation, making it possible for the device to be widely used as consumer and user friendly device.*

Keywords- **Object detection, OPENCV, Deep learning, Text detection, Gesture.**

1. INTRODUCTION

Visual impairment is a prevalent disorder among people of different age groups. The spread of visual impairment throughout the world is an issue of greater significance. According to the World Health Organization (WHO), about 8 percent of the population in the eastern Mediterranean region is suffering from vision problems, including blindness, poor vision and some sort of visual impairment. They face many challenges and difficulties. It includes the complexity of moving in complete autonomy and the ability to identify objects. According to Foulke "Mobility is the ability to travel safely, comfortably, gracefully and independently through the world."Mobility is an important key to personal independence. As visually challenged people face a problem for safe and independent mobility, a great deal of work has been devoted to the development of technical aids in this field to enhance their safe travel. Outdoor Travelling can be very demanding for blind people. Nonetheless, if a blind person is able to move outdoors, they gain confidence and will be more likely to travel outdoors often leading to improved quality of their life. Travelling can be made easier if the visually challenged people are able to do the following tasks. They are:

- Identify objects that lie as an obstacle in the path.
- Identify any form of texts that convey traffic rules.

Most blind people and people with vision difficulties were not in a position to complete their studies. Special schools for people with special needs are not available everywhere and most of them are private and expensive thus it also becomes a major obstacle for them to live a scholastic life. Communication between people with different defects has been impossible decades ago. Particularly the communication between visually challenged people and deaf-mute people is a challenging task. If complete progress is made to enable such communication, the people with impairment forget their inability and can live an ambitious life in this competitive world.

The rest of the paper is organized as follows. Section II will provide an overview of the various related works in the area of using deep learning based computer vision techniques for implementing smart devices for the visually impaired. The detail of the methodology used is discussed in section III. Proposed system design and implementation will be presented in Section IV. Experimental results are presented in Section V. Future scope is discussed in Section VI. Conclusion is provided at section VII.

2. RELATED WORK

The application of technology to aid the blind has been a post-World War II development. Navigation and positioning techniques in mobile robotics have been explored for blind assistance to improve mobility. Many systems rely on heavy, complex, and expensive handheld devices or ambient infrastructure, which limit their portability and feasibility for daily uses. Various electronic devices emerged to replace traditional tools with augmented functions of distance measurement and obstacle avoidance, however, the functionality or flexibility of the devices are still very limited. They act as aids for vision substitution. Some of such devices are:

2.1 Long cane:

The most popular mobility aid is, of course, the long cane. It is designed primarily as a mobility tool to detect objects in the path of user. Its height depends on the height of the user.

2.2 eSight 3:

It was developed by CNET. It has High resolution camera for capturing image and video to help people with poor vision. The demerit is that it does not improve vision as it just an aid. Further improvements are made to produce a Water proof version of eSight 3.

2.3 Aira:

It was developed by SumanKanuganti. Aira uses smart glasses to scan the environment Aira agents help users to interpret their surroundings by smart glasses. But it takes large waiting time for connecting it to the Aira agents in order to be able to sense to include language translation features.

2.4 Google Glasses:

It was developed by Google Inc. Google Glasses show information without using hands gestures. Users can communicate with the Internet via normal voice commands. It can capture images and videos, get directions, send messages, audio calling and real- time translation using word lens app. But It is expensive (1,349.99$).Efforts are made to reduce costs to make it more affordable for the consumers.

2.5 Oton Glass:

It was developed by Keisuke Shimakage, Japan. It has Support for dyslexic People as it Converts images to words and then to audio. It looks like a normal glasses and it also supports English and Japanese languages. But it supports only people with reading difficulty and no support for blind people. It can be improved to support blind people also by including proximity sensors.

3. METHODOLOGY

FOR(E)SIGHT uses OPENCV(OPENSOURCE COMPUTER VISION). Computer Vision Technology has played a vital role in helping visually challenged people to carry out their routine activities without much dependency on other people. Smart devices are the solution which enables blind or visually challenged people to live normal life. FOR(E)SIGHT is an attempt in this direction to build a novel smart device which has the ability to extract and recognize object, text, action captured from an image and convert it to speech. Detection is achieved using the OpenCV software and open source Optical Character Recognition (OCR) tools Tesseract based on Deep Learning techniques. The image is converted into text and the recognized text is further processed to convert to an audible signal for the user. The novelty of the implemented solution lies in providing the basic facilities which is cost effective, small-sized and accurate. This solution can be potentially used for both educational and commercial applications. It consists of a Raspberry Pi 3 B+ microcontroller which processes the image captured from a camera super-imposed on the glasses of the blind person.

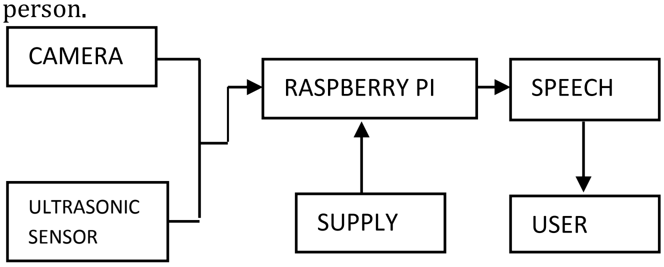


Fig.1

The block diagram indicates the basic process taking place in FOR(E)SIGHT.(Fig.1)

The components used in FOR(E)SIGHT can be classified into hardware and Software. They are listed and described below,

3.1 Hardware Requirements:

It includes

1) Raspberry Pi 3B+
2) camera module
3) ultrasonic sensor
4) Headphones.

1) Raspberry Pi 3B+:

Raspberry Pi 3B+ is a credit card-sized computer that contains several important functions on a single chip. It is a system on a chip (SoC).It needs to be connected with a keyboard, mouse, display, power supply, SD card and installed operating system. It is an embedded system that can run as a no-frills PC, a pocket coding computer, a hub for homemade hardware and more. It includes 40 GPIO (General Purpose Input/output) pins to control various sensors and actuators. Raspberry Pi 3 used for many purposes such as education, coding, and building hardware projects. It uses [1]Python through which it can accomplish many important tasks. The Raspberry Pi 3 uses Broadcom BCM2837 SoC Multimedia processor. The Raspberry Pi's CPU is the 4x ARM Cortex-A53, 1.2GHz processor. It has internal memory of 1GB LPDDR RAM.

2) Camera Module:

The Raspberry Pi camera is 5MP and has a resolution of 2592x1944 pixels. It is used to capture stationary objects. On the other hand IPcam can be used to capture actions. IPcam of mobilephones can also be used for this purpose. The IPcam has a view angle of 60�with a fixed focus. It can capture images with maximum resolutions of 1289x720 pixels. It is compatible with most operating systems platforms such as Linux, Windows, and MacOS.

3) Ultrasonic Sensor:

The ultrasonic sensor is used to measure the distance using ultrasonic waves. It emits the ultrasonic waves and receives back the echo. So, by measuring the time, the ultrasonic sensor will measure the distance to the object. It can sense distances in the range from 2-400 cm. In FOR(E)SIGHT, the ultrasonic sensor measures the distance between the camera and an object or text or action to decode it into an image. It was observed based on experimentation that the distance to the object should be from 10 cm to 150 cm to capture a clear image.

4) Headphones:

Wired headphones was used since Raspberry Pi 3 B+ come with Audio jack of 3.5mm.The headphones will be used to help the user listen to the text that is been converted to audio after it has been captured by the camera or to listen to the translation of the text. The headphone used is small, lightweight so the user will not feel uncomfortable wearing them.

3.2 Software Requirements:

1) OpenCV Libraries:

OpenCV is a library of programming functions for real-time computer vision. The library is cross-platform that is, it is available on different platforms including Windows, Linux, OS X, Android, iOSetc and it supports a wide variety of programming languages like C++, Python ,Java etc. OpenCV is free for use under the open-source BSD license It performs real-time applications like image processing, Human–computer interaction (HCI), Facial recognition system, Gesture recognition, Mobile robotics, Motion tracking, Motion understanding, Object identification, Segmentation and recognition, Augmented reality and many more.

2) OCR:

OCR (Optical Character Recognition) is used to convert typed, printed or handwritten text into machine-encoded text. There are some OCR software engines which try to recognize any text in images such as Tesseract and EAST. FOR(E)SIGHT uses Tesseract version 4, because it is the best open source OCR engines. It is used directly, or using an API to extract data from images. It supports a wide variety of languages. It is open source software that has a framework for building efficient speech synthesis systems.

3) TTS API:

Last function is text to voice conversion. In order to implement this task, TTS (Text-to-Speech) is installed. It is a python library that interfaced with certain API .TTS has many features such as conversion of text to voice, provide error pronunciation using customizable text pre-processors.

4. PROPOSED SYSTEM

FOR(E)SIGHT is a device that is designed to use the computer vision technology to capture an image and extract English text and convert it into audio signal with the help of speech synthesis. The overall system can be viewed in three stages. They are,

4.1 Primary Unit:

1) Sensing:

The first unit is sensing unit. It comprises of ultrasonic sensor and camera. Whenever the object or text or action is noticed within a limited region of space, the camera captures the image.

[1]*The code is available at :*https://github.com/Udithswetha/foresight/tree/master/.gitignore

2) Image acquisition:

In this step, the camera captures the images of the object or text or action. The quality of the image captured depends on the camera used.

4.2 Processing unit:

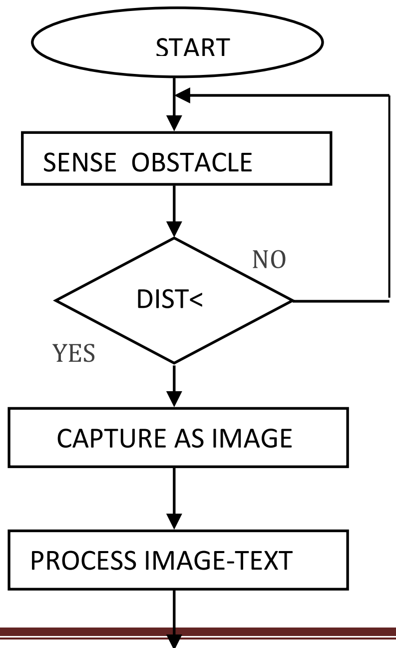
It is a technique through which data of Image is obtained in the form of numbers. It is done using OpenCV and OCR tools. The libraries which are commonly used for this are numpy, matplotlib and pillow.
This step consists of

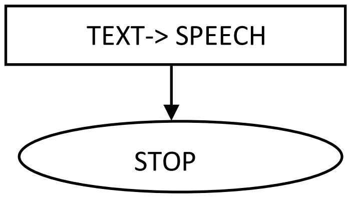
1) Gray scale conversion:

The image is converted to gray scale because OpenCV functions tend to work on the gray scale image. NumPy includes array objects for representing vectors, matrices, images and linear algebra functions. The array object allows the important operations such as matrix multiplication, solving equation systems, vector multiplication, transposition and normalization, that are needed to align and warp the images, model the variations and classify the images. Arrays in NumPy are multi-dimensional and can represent vectors, matrices, and images. After reading images to NumPy arrays, mathematical operations are performed.

2) Supplementary Process:

Noise is removed using bilateral filter. Canny edge detection is performed on the gray scale image for better detection of the contours. The warping and cropping of the image are performed to remove the unwanted background. In the end, Thresholding is done to make the image look like a scanned document to allow the OCR to convert the image to text efficiently.





FOR(E)SIGHT implements a sequential flow of algorithms. The flow of the process is depicted.

4.3 Final unit:

The last step is to convert the obtained Text to speech (TTS). TTS program is created in python by installing the TTS engine (Espeak).The quality of the spoken voice depends on the speech engine used.

5. EXPERIMENTAL RESULTS

The figure (Fig.2) shows the final prototype of FOR(E)SIGHT.

Fig.2

It is noted that the hardware circuit (Raspberry Pi) is placed on the glass head piece and the glasses can be worn on the face which consists of the cameraand the ultrasonic sensor. The sensor senses the object or text or gesture and the pi camera captures it as an image. Though it is not highly comfortable, it is comparatively cost effective and beneficial.

5.1 Object Detection:

The main goal was to detect the presence of any obstacle in front of the user by using ultrasonic sensing and image processing techniques. FOR(E)SIGHT has successfully completed the task.(Fig.3)

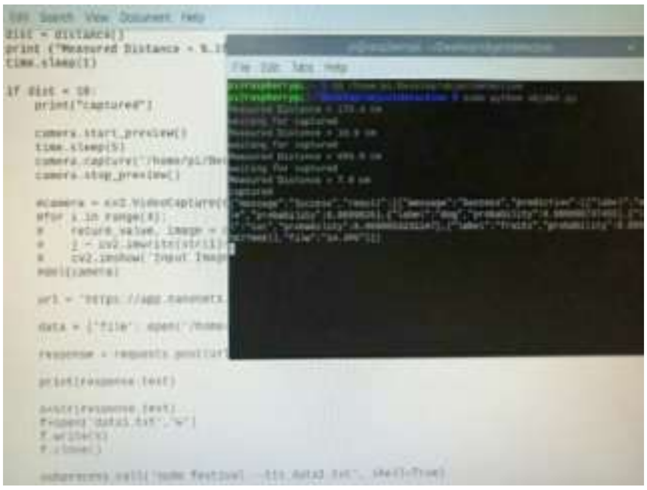


Fig.4

5.2 Text Detection:

In this result, the text detectorTessaract that was used in this solutionwas checked whether it was working well. The test showed mostly good results on big clear texts. It is found that the recognition depends on the clarity of the text in the picture, its font theme, size, and spacing between the words. (Fig.5)

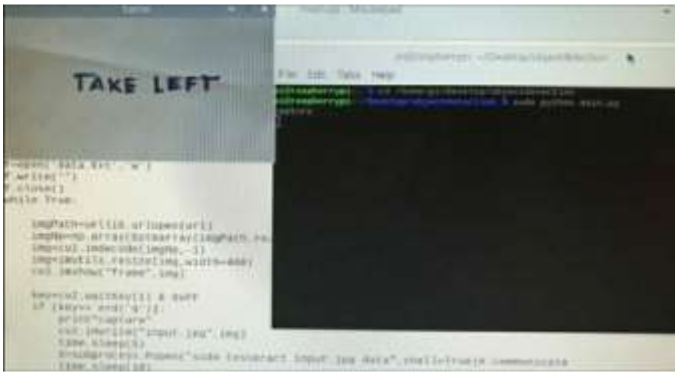


Fig.5

5.3 Gesture Detection:

In this test, the aim is to check the gesture is identified with the help OCR tools.FOR(E)SIGHT came out with positive results by detecting the gestures. But it is found that FOR(E)SIGHT identifies only the still images of gestures.(Fig.6)

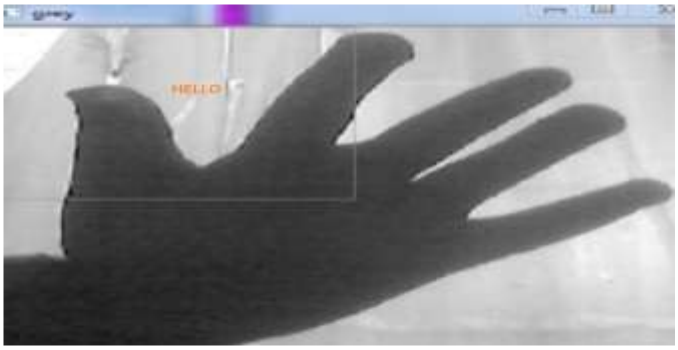


Fig.6

6. FUTURE SCOPE

However, there are some limitations in the proposed solution which can be addressed in the future implementations. Hence, the following features are recommended to be incorporated in the future versions.

- Time taken for each process can be reduced for a better usage.
- Translation module can be developed, in order to cater to a wide variety of users, as it would be highly beneficial if it were a multi-lingual featured device.
- To improve the direction and warning messages to the user, GPS-based navigation and alert system can be included.
- Recognizing more amounts of gestures would be helpful for performing more tasks.
- High space visibility can be provided by including a wide angle camera 270^0 degrees as compared to 60^0 currently used.
- To provide for more real-time experience, video processing instead of still images can be used instead of still images.

7. CONCLUSION

A novel device FOR(E)SIGHT for helping the visually challenged has been proposed in this paper. It is able to capture an image, extract and recognize the embedded text and convert it into speech. Despite certain limitations, the proposed design and implementation served as a practical demonstration of how open source hardware and software tools can be integrated to provide a solution which is cost effective, portable and user-friendly. Thus the device discussed in this paper has helped the visually challenged by providing maximum benefits at a minimal costs.

REFERENCES

[1] Saransh Sharma, Samyak Jain, Khushboo," A Static Hand Gesture and Face Recognition System For Blind People", 6th International Conference on Signal Processing and Integrated Networks (SPIN), pp.534-539, 2019.

[2] Jinqiang Bai, Shiguo Lian, Zhaoxiang Liu, Kai Wang, Dijun Liu," Virtual-blind Road following based wearable navigation device for blind people ", IEEE Transactions on consumer electronics, vol. 31, pp. 132-140, February 8, 2018.

[3] Arun Kumar,Anurag Pandey," Hand Gestures and Speech Mapping of a Digital Camera For Differently Abled", IEEE 6th International Conference on Advanced Computing, pp.392-397, 2016.

[4] H. Zhang and C. Ye, "An indoor navigation aid for the visually impaired,"in 2016 IEEE Int. Conf. Robot. Biomimetics (ROBIO), Qingdao, pp.467-472, 2016.

[5] Hanene Elleuch, Ali Wali, Anis Samet, Adel M. Alimi, "A static hand gesture recognition system for real time mobile device monitoring", 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 195 – 200, 2015.

[6] Hongsheng he, Van li, Vong Guan, and Jindong Tan, " Wearable Ego-motion Tracking for blind navigation in indoor environments", IEEE transactions on Automation science and engineering, vol. 12, no.4, pp.1181- 1190, October 2015.

[7] X. C. Yin, X. Yin, K. Huang, H. W. Hao, "Robust text detection in natural scene images " IEEE transactions on Pattern Analysis and Machine Intelligence, 36(5), pp. 970-980, 2014.

[8] Samleo L. Joseph, Jizhong Xiao, Xiaochen Zhang, Bhupesh Chawda, Kanika Narang, Nitendra Rajput, Sameep Mehta, and L. Venkata Subramaniam," Being Aware of the World: Toward Using Social Media to Support the blind With Navigation", IEEE Transactions on Human-machine systems, vol. 33, no. 4, pp.1-7, November 30, 2014.

[9] H. He and J. Tan, "Adaptive ego-motion tracking using visual-inertial sensors for wearable blind navigation," in ACM Proceedings of Wireless Health Conerence, pp.1-7, 2014.

[10] L. Neumann, J. Matas, "Scene text localization and recognition with oriented stroke detection", Proceedings of IEEE International Conference on Computer Vision, pp. 97-104, 2013.

[11] Bissacco, M. Cummins, Y. Netzer, H. Neven, "Reading text in uncontrolled conditions", Proceedings of the IEEE International Conference on Computer Vision, pp. 785-792, 2013.

[12] Y. T. Chucai Yi, "Text extraction from scene images by character appearance and structure modeling," Computer Visual Image Understanding, vol. 117, no. 2, pp. 182–194, 2013.

[13] World Health Organization—Visual Impairment and Blindness. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs282/en/ (2013).

[14] L. Neumann, J. Matas, "Real-time scene text localization and recognition ", Proceedings of IEEE International Conference onComputer Vision and Pattern Recognition (CVPR), pp. 3538-3545, 2012.

[15] Dimitrios Dakopoulos and Nikolaos G. Bourbakis," Wearable obstacle avoidance electronic travel aids for blind: A survey," IEEE Transactions on systems, man, and cybernetics-part c: applications and reviews", vol. 40, no. 1, pp. 25-35, January 2010.

[16] David J. Calder," Assistive technology interfaces for the Blind", 3rd IEEE International conference on Digital Ecosystems and Technologies,pp.318-323,2009.

[17] B. And⍰," A Smart Multisensor Approach to Assist Blind People in Specific Urban Navigation Tasks", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 16, no. 6, pp 592-594, December 2008.

[18] Hui Tang and David J. Beebe," An Oral tactile interface for blind navigation", IEEE Transactions on neural systems and rehabilitation engineering, vol. 14, no.1, pp.116-123, March 2006.

[19] Mohammad Shorif Uddin and Tadayoshi Shioyama," Detection of Pedestrian Crossing Using Bipolarity Feature-An Image-Based Technique", IEEE Transactions on intelligent transportation systems, vol. 6, no. 4, pp.439-445, December 2005

[20] Y. J. Szeto, "A navigational aid for persons with severe visual impairments: A project in progress," presented at the Proceedings of the 25th Annual International Conference on IEEE Engineering Medical Biology Society., Cancun, Mexico, Sep. 17–21, 2003.

[21] Andò, "Electronic sensory systems for the visually impaired," IEEE Magazine on Instrumentation and Measurement, vol. 6, no. 2, pp. 62–67, Jun. 2003.