A VISION DISABLED INDIVIDUAL'S ASSISTIVE MECHANISM

Kirana J P, Keerthi Reddy B V, Hariprita K S, Gagana V, Priyanka Bharti

School of Computer Science and Engineering, REVA University

Bengaluru, India

Abstract—**Visually challenged people make up a sizable population segment, with estimates ranging from tens of millions to hundreds of millions worldwide. Integration into society is a major and ongoing goal for them. A considerable deal of effort has gone into ensuring a model of primary care. To facilitate visually impaired people in living a full life, many navigating system approaches have been introduced. Often, these systems are built with a particular goal in mind. Nonetheless, these solutions can significantly improve the mobility and safety of such individuals. The plan establishes a platform based on vision recognizing real-world items both indoors and outdoors in order to help people who are visually handicapped. The software is created with OpenCV library and Python functionalities and then ported to a laptop. YOLO is a novel way of object detection. The image is converted to a scanned image for further processing using the application for interpreting the contents. The output of the image detected is a scanned image, which then uses Tesseract OCR (Optical Character Recognition) software as an input, which converts the image to text, along with that we add road sign recognition for outdoor guidance while walking on the road. We use the TTS (Text to Speech) engine to translate text to speech after detecting texts and objects.**

Keywords—Deep learning, Object Detection, OpenCV, YOLO, Tesseract OCR, Road sign detection, TTS.

I. INTRODUCTION

Individuals who are visually impaired (VI) or blind have a lower quality of life since they are unable to sense the terrain and their surroundings. In their daily lives, they would require constant assistance and walking support systems. For decades, solutions have been introduced, and they are rapidly improving because of technological evolution and integration. An outsized variety of help aids are deployed in the real world whereas other ideas remained as analysis concepts.

The event of progressive steering the sophisticated image acquisition and computer vision techniques, as well as device and unit processor speed performance, are all linked to systems that assist visually impaired people. Regardless of the terminology used, the software must work in real-time, with swift judgments, because speed is crucial when taking action.

Obtaining the simplest answer is essentially a trade-off between the software component's performance and the complexity of the solution possibilities of hardware. It is necessary to standardize the parameters to the greatest extent possible. One of the goals of the aided system during a visually impaired person's indoor movement is to identify and distinguish objects or impediments, which is followed by an audio alert.

This system's planned vision module for image processing is an associate integrated with a portion of the system dedicated to serving blind and visually impaired folks. Furthermore, regardless of the integrated platform, the suggested system could be used outside of the shell. Throughout experimentation and iterative optimization, the image decision to develop has been designed, developed, and proven to be effective. The module refers to the notion of creating a high-performance gadget that is also cost-effective when used properly. The module employs disruptive technologies to allow for changes and the addition of new features.

II. OBJECTIVE AND SCOPE

1. Identify, recognize and produce valuable captions for a certain picture by means of Deep Learning Techniques.
2. Identify the object's position in the frame with the object label/object name.
3. Programming both the objects detected and the position of the objects to a speech output using text-to-speech convert.
4. Analyzing the written text and converting it into speech format.
5. This paper's one of the objectives is to describe a sign detection approach that can convey regions of interest to a classifier.
6. Sign recognition is been said in voice format so that a blind person will have no difficulties in understanding the sign.

III. LITERATURE REVIEW

More than a quarter of the world's 36 million blind individuals live in India. One of the most biggest obstacles that blind institutions face these days is educating the blind to avoid states within their total number. Despite the fact that many schools employ Braille to combat, it is quite unapproachable. According to India's Braille Literacy Statistics, only 10 percent of the country's twelve million blind people, know Braille. The challenges faced by the blind and visually impaired to browse and learn without the utilization of Braille is one of the most significant issues they face. As a result, The approach presented by

Megha P Arakeri, Keerthana N S, Madhura M, Anusha Sankar, and Tazeen Munnavar is to create a low-cost gadget that utilizes ai algorithms to browse any words in various orientations and illumination circumstances surrounding the user. The technology uses a Raspberry Pi and an appropriate webcam, to gather stuff all across people who are blind and read it aloud in their native language to them. The device has a sensor that alerts the user to the closeness of the closest item at a neutral position and simultaneously calculates the number of objects in its view. Methods for image segmentation, algorithms, and voice synthesizer are all used in the system. Eighty-four percent efficiency was found when optical character recognition and visual perception algorithms were integrated. [1]

According to Effy Maria, Ani R, Sakkaravarthy V, and J Jameema Joyce of the World Health Organization, 285 million individuals worldwide are vision-impaired out of a population of 7.4 billion, and they've been experiencing difficulty managing their everyday routines. They have proposed smart spectacles for blind people that can identify text and generate an audio output as a means of assisting them. It allows visually challenged people to scan any textual content using their voices. Using the device's built-in camera, the text image from the written text is collected, and the obtained picture is examined using Tesseract-OCR (Optical Character Recognition). eSpeak, a simple open-source speech synthesizer, is then used to convert the detected text to voice. Ultimately, by applying the TTS technique, the headphones generate synthesized speech. The primary objective for implementation in this project is Raspberry Pi which serves as a gateway between the sensors, camera, and image processing results, as well as performing operations to control peripheral units (USB, Keyboard, etc.,). [2]

L. Tepelea et al.[4] explained how to help visually impaired people using a CNN-based correlation method. Given the large number of data that will be collected from images acquired. A visual processing unit must be included in the structure of systems that aid people with visual impairments, regardless of the version presented. It proposes a correlation method based on the usage of CNNs(cellular neural networks), which can enhance the options of supporting systems and give visually impaired persons additional information from their surroundings. The majority of the operations (calculations) in the suggested algorithm is been performed by parallel processing. As a result, the computing time will be reduced, and the computing time will not rise correspondingly as the template picture size increases.[3]

P Szolgay et al.[5], designed the portable system which was based on a smartphone but also incorporates the usage of external sensing modules. It covers visually challenged people's interior and outdoor mobility. The system's effectiveness has been demonstrated in tests, and it can be further enhanced with the advancement of Android-based portable devices. The research described a portable solution for assisting visually impaired people in both indoor and outdoor settings. It used a variety of sensors to detect obstructions and, with the help of GPS and a compass, lead them through their movements. A multicore Android smartphone serves as the central component of the system. Alternative sensory modules detect barriers and relay pertinent information to the main portion. The technology might also interact remotely to allow for distant monitoring. [4]

According to Ahmed Abdelgawad, Michael Trent, and Kumar Yelamarthi, smart gadgets have grown significantly more prevalent in our everyday lives in recent years. They're being implemented into houses, cars, buildings, and public areas. Furthermore, the IoT, (Internet of Things), is a technological revolution, that provides us all with new opportunities. To assist blind people, a variety of navigating systems have been designed. Despite this, none of these systems are linked to the Internet of Things. Their goal was to create a low-cost, low-power IoT navigation system for those who are blind. Their solution included a network of ultrasonic sensors installed on a waist belt that surveyed the scene, iBeacons that specifies the position, as well as a Raspberry Pi that processed the data. Ultrasonic sensors on the Raspberry Pi detected the obstructions and provided audible cues to the user via a Bluetooth headset. iBeacons were placed in a variety of locations, each with its own unique ID. There was a database in the cloud for all the iBeacons that were linked to the corresponding information, such as the address and data about the location. The ID of the iBeacon is detected by the Raspberry Pi and delivered it to the cloud, which then resends the information associated with the existing ID to the Raspberry Pi, which converts the textual content to speech and plays it to the user through a Bluetooth headset. Tests showed that the technology worked as a navigation assistant and was accurate within the minimum radius. [5]

A variety of machine vision has been created to assist the blind and visually handicapped. The viewers' movement causes all scene objects, whether static or non-static, to move. As a result, detecting moving objects with a moving observer is required. Yogesh H Dandawate and Sanket Khade have proposed a camera-based prototype system for assisting blind people in detecting impediments utilizing motion vectors in this context. To accomplish object detection, they gathered a collection of their indoor and outdoor environments and evaluated the light flow. Furthermore, they were able to recognize the items inside the zone of interest without the need for expensive depth cameras or sensors. The Raspberry Pi 2-B is the hardware used in the proposed project, and the object identification algorithms are implemented in MATLAB (for simulation reasons) and Python. [6].

IV. PROPOSED SYSTEM

For people with visual impairments, this system features a voice-assisted word-based device. The suggested framework includes a camera module, an image processing module, an optical character recognition module, the Yolo framework, and a text-to-speech module.

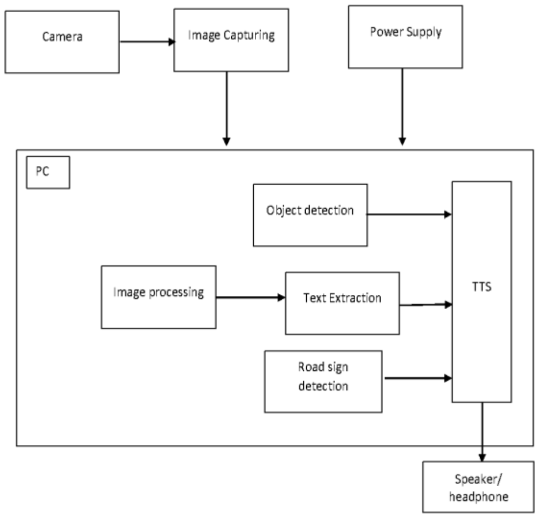

Figure 1. System Architecture

A. Image capturing

A Raspberry Pi camera with a 5mp resolution is used to capture the text picture at this step. In terms of shape and scale, the acquired image isn't perfect, nor in an acceptable state for text extraction data. As a result, the image-processing module is used to process the acquired picture first. The photograph was captured in jpg format.

B. Detection of object

Many artifacts are present in the captured shot. The JSON.parse library is used to identify all objects, although not all of them are interpreted, however, those with higher accuracy can be read.

C. Image processing

The image magic implementation is used to eliminate unnecessary noise in picture processing. Image magic is open-source and free software. Image magic consists of a variety of tools in which the proposed technique uses image sharpening and text washing. Image sharpening improves the difference between the image's bright and dark areas. Text cleaning is a technique for cleaning scanned documents in order to make the final picture more readable for OCR.

D. Text Extraction

The image magic program's output image is converted into textual or modifiable data in this stage. For this execution, we deployed the Tesseract OCR tool. The Tesseract OCR program detects textual content in the captured image after it has been analyzed. The results of the extraction of information are saved in the .txt format.

E. Sign detection

Sign classification includes RGB pictures of the traffic signboards. These RGB images are been preprocessed using multiple techniques, particularly shuffling, grey scaling, local histogram equalization, and normalization. To come up with further additional training data transfer image function is being employed which has rotation, sharing, and image translations. Using TensorFlow the information is being trained and tested.

F. Text - Speech converter

A voice synthesizer is employed in the synthesis of retrieved text into speech in this stage. For this procedure, we employed the Google speech synthesizer and the e-speak TTS engine. A speech synthesizer's output is audio or sound.

V. METHODOLOGY

Object Detection

The video has captured with the use of a camera that is then divided into a series of frames. Object detection has been accomplished with the use of CNN classifiers and text to audio conversion is done using pyttsx3.

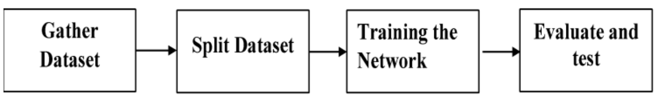

Figure 2. Deep learning steps

For the mobility of the person's body in the internal environment, the procedure picture capture > image processing > acoustic notification is repeated. The total processing time is calculated by adding the three processing cases together, and this sets the acquisition rate for the picture frames input. This strategy must be quick enough to avoid any barriers in a timely manner.

The image processing approach is used to solve an issue with detecting a specific object, in this case, traffic signal recognition. We utilized the cv2 function from the OpenCV library.

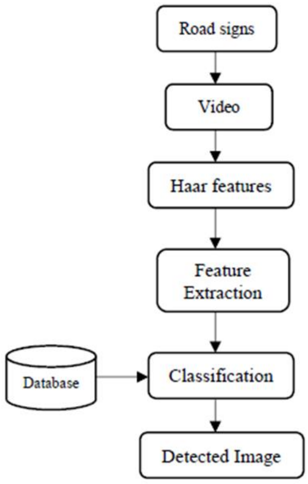

Figure 3. Object detection algorithm workflow

For the Python version, match the template that already exists in the library. The following design standards were addressed by the module:

1. The amount of time that could pass between two subsequent video frames, we strive to maintain every template's processing time as a minimum. The total processing time was increased as we used the method on several scales ought to be sufficiently little to allow real-time decisions.
2. Among each collected video frame is down sampled with diverse methods in the multiscale technique, resolution 5, 3, and 1 are the factors. If the source image's initial resolution is 960x1280 pixels, for example, down sampling with issue 3. As a consequence, three portraits with reduced resolution will be created with the following dimensions: 960 x 1280 pixels, 720 x 960 pixels, and 480x640 pixels. The template is then compared to each image source version that has now been scaled. Another factor to consider is the internal module parameter approach in relation to the size of road signs and source picture resolution. We must account for many processing times at each stage, beginning with image acquisition, module communication, and finishing with the trigger action to emit the auditory alarm, in order to have an overall evaluation. An auditory message is provided through headphones to the user once an object has been spotted. The audio message is, in its most basic form, additional or fewer than 1.2÷1.5s in length. Currently, this time span is used as a benchmark.

Text Recognition

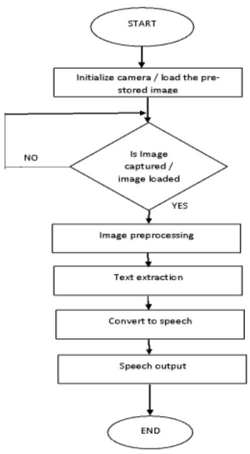

Figure 4. Text Recognition

Image Capturing: The first step is to capture an image from the document or book and thereafter the document or book is placed below the camera that facilitates capturing an image from the document or book. The camera used to capture an image is a PC camera.

Image pre-processing: It is the technique of removing undesirable noise from a photograph, by applying a suitable threshold to the image. It is used for correcting skew angles, sharpening images, thresholding, and segmentation.

Text extraction: In our project, Tesseract OCR is what we are using, an engine that is employed to extract the recognized text.

Text to speech: Once extracting the text, it is converted into a message. The text to speech converter synthesizer is employed to convert text into speech. Ultimately the speech output gets through.

Sign Recognition

Image Capturing: The first step is to capture an image from the road signboard.

Haar features: This may not accurately reflect the strategy, but with a large number of features in multiple stages, this technique will save burden in the later stages because the majority of windows and frames will be rejected only in the initial stages.

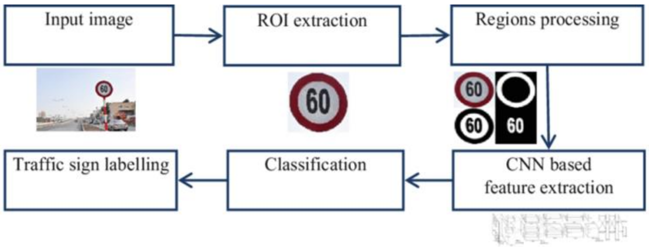

Figure 5. Sign Recognition

ROI extraction (Region Of Interest): It's a region of a picture. that all you have to do now is filter or operate in a way. The required segments from an image convert the RGB image to gray-scale, removes noise from the gray-scale image where a lot of unwanted edges need to be eliminated and a few of the sides have gaps in between that need to be closed.

Once a road sign is detected and identified, the traffic-road sign is labeled and the user receives an auditory message via headphones.

VI. Results and Discussion

The hardware implementation of this system is through using a custom-built spec that will include a camera and separate buttons on the sides for user inputs. The user will hear the output using a headphone set that is linked to the system. The software simulation is created using the same specific buttons. The initial interface shown on the screen when we execute the application is illustrated in Figure 6. Here are two buttons that allow the user to either load a previously shot image or live capture an image in the current environment. The third available option is the quit button, which is used to exit the interface.

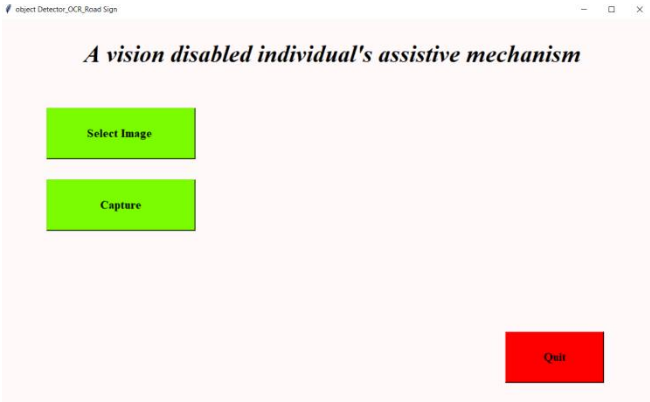

Figure 6. Initial interface

When the select image button is processed, user will be given the option of selecting a signboard image or an image with text from the appropriate folder. The selection interface is depicted in Figure 7.

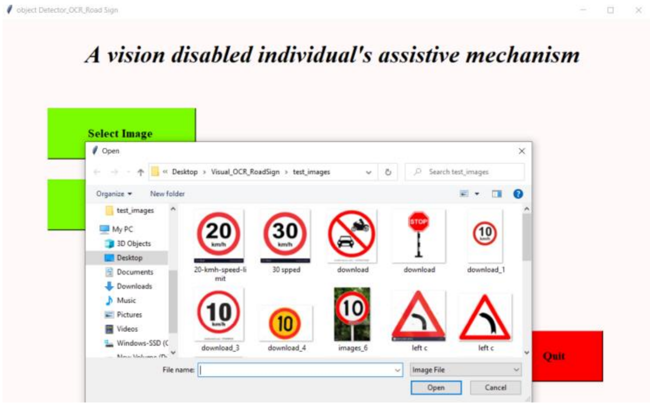

Figure 7. Selection layout

Following the selection of the needed image, three buttons will appear on the screen, each of which must be selected in relation to the image chosen by the user.

After clicking the relevant button, the image is pre-processed, and we see a textual output at the bottom of the window, as well as an audio output of the same text. Figures 8 and 9 show the analyzed text and sign, respectively.

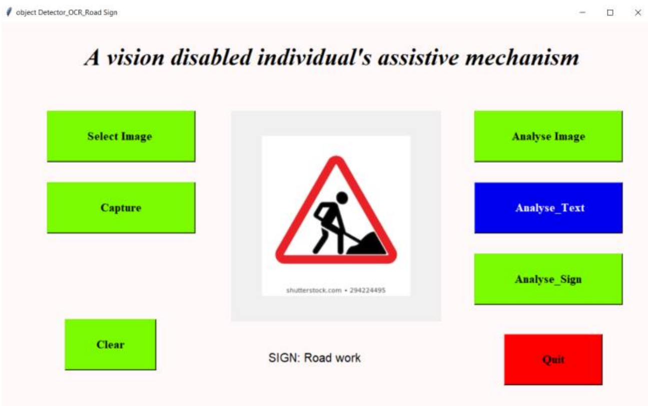

Figure 8. Sign detection window

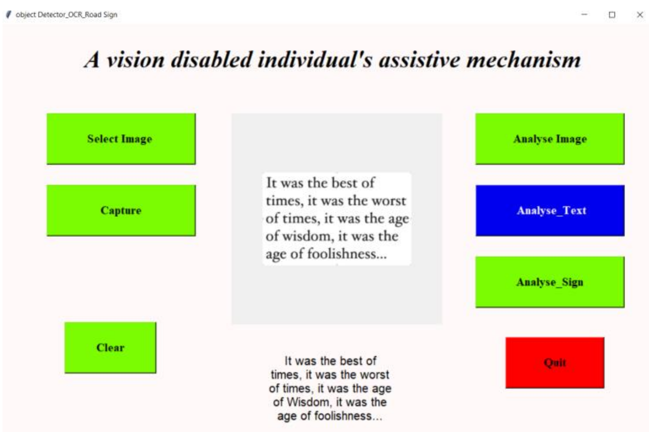

Figure 9. Text detection window

If the user wants to process a real-time image, this option is included in the live capture button. When the capture option is selected, the system's camera is automatically activated, and an image is captured in the GUI's center as shown in Figure 10. We get an output window after analyzing the image, where the object is recognized including its location in the frame. For audio assistance, this output is obtained through the system's speakers.

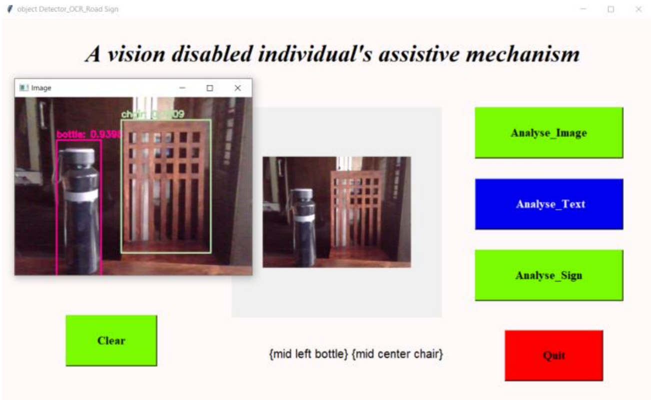

Figure 10. Realtime objects detection

VII. CONCLUSION

In our platform, we have created a word-based system with voice recognition software as well as object detection, recognition, and localization. Our proposed method decodes the required picture and identifies it aloud. The device can be utilized in a variety of ways because its result is already in the form of speech. This is an easy-hearing setup for people who are visually handicapped. This technology is both efficient and cost-effective solution for blind individuals. This equipment is advantageous for blind students at schools and universities. This can also be viewed as an example of artificial intelligence in action. It is beneficial to illiterate people and is also a smart technology that is highly important and feasible to society.

REFERENCES

[1] M.P. Arakeri, N.S. Keerthana, M. Madhura, A. Sankar, T. Munnavar, “Assistive Technology for the Visually Impaired Using Computer Vision”, International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, pp. 1725-1730, sept. 2018.

[2] R. Ani, E. Maria, J.J. Joyce, V. Sakkaravarthy, M.A. Raja, “Smart Specs: Voice Assisted Text Reading System for Visually Impaired Persons Using TTS Method”, IEEE International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT), Coimbatore, India, Mar. 2017.

[3] L. Ţepelea, A. Gacsádi, I. Gavriluţ, V. Tiponuţ, “A CNN Based Correlation Algorithm to Assist Visually Impaired Persons”, IEEE Proceedings of the International Symposium on Signals Circuits and Systems (ISSCS 2011), pp.169-172, Iasi, Romania,2011

[4] P. Szolgay,L. Ţepelea, V. Tiponuţ, A. Gacsádi, “Multicore Portable System for Assisting Visually Impaired People”, 14th International Workshop on Cellular Nanoscale Networks and their Applications, pp. 1-2, University of Notre Dame, USA, July 29-31, 2014.

[5] E.A. Hassan, T.B. Tang, “Smart Glasses for the Visually Impaired People”, 15th International Conference on Computers Helping People with Special Needs (ICCHP), pp. 579-582, Linz, Austria, 2016.

[6] Sanket Khade, Yogesh H. Dandawate, “Hardware Implementation of Obstacle Detection for Assisting Visually Impaired People in an Unfamiliar Environment by Using Raspberry Pi. ”, Smart Trends in Information Technology and Computer Communication, Pune, India, 2016.

[7] V. Tiponuţ, D. Ianchis, Z. Haraszy, “Assisted Movement of Visually Impaired in Outdoor Environments”, Proceedings of the WSEAS International Conference on Systems, Rodos, Greece, pp.386-391, 2009.

[8] Mallapa D. Gaurav, Shruti S. Salimath, Shruti B. Hatti, Vijayalaxmi I. Byakod, Shivaleela Kanede, ”B-Light: A reading aid for the blind people using OCR and OpenCV”, International Journal of Scientific Research Engineering & Technology, May 2017, vol. 6, issue 5, pp. 546-548.

[9] Nikhil Mishra, “Image Text to Speech Conversion using Raspberry Pi & OCR Techniques”, International Journal for Scientific Research and Development, vol. 5, issue 08, 2017.

[10] Ahmed, S. T., Singh, D. K., Basha, S. M., Nasr, E. A., Kamrani, A. K., & Aboudaif, M. K. (2021). Neural Network Based Mental Depression Identification and Sentiments Classification Technique From Speech Signals: A COVID-19 Focused Pandemic Study. Frontiers in public health, 9

[11] Zhiming Liu, Yudong Luo, Jose Cordero, “Finger-eye: A wearable text reading assistive system for the blind and visually impaired”, IEEE International Conference on Real-time Computing and Robotics, 6-10 June 2016.

[12] Govardanam V, Vignesh Babu T.N, Sri Hari Kavin N, “Automated Read-Write kit for Blind using Hidden Markov Model and Optical Character Recognition”, International Conference on Applied and Theoretical Computing and Communication Technology, 29-31 Oct. 2015.