Smart Assistant for Blind Humans using Rashberry PI

Abish Raj. M. S, Manoj Kumar. A. S, Murali. V
Department of Computer Science and Engineering,
Sri Muthukumaran Institute of Technology, Maharashtra, Chennai, Tamil Nadu, India

ABSTRACT

An OCR (Optical Character Recognition) system which is a branch of computer vision and in turn a sub-class of Artificial Intelligence. Optical character recognition is the translation of optically scanned bitmaps of printed or hand-written text into audio output by using of Raspberry pi. OCRs developed for many world languages are already under efficient use.This method extracts moving object region by a mixture-of-Gaussians-based background subtraction method. A text localization and recognition are conducted to acquire text information. To automatically localize the text regions from the object, a text localization and Tesseract algorithm by learning gradient features of stroke orientations and distributions of edge pixels in an Adaboost model. Text characters in the localized text regions are then binaries and recognized by off-the-shelf optical character recognition software. The recognized text codes are output to blind users in speech. Performance of the proposed text localization algorithm. As the recognition process is completed, the character codes in the text file are processed using Raspberry pi device on which recognize character using Tesseract algorithm and python programming, the audio output is listed.

I. INTRODUCTION

In the running world, there is growing demand for the software systems to recognize characters in computer system when information is scanned through paper documents as we know that we have number of newspapers and books which are in printed format related to different subjects. This information by searching process"

One simple way to store information in these paper documents in to computer system is to first scan the documents and then store them as Images. But to reuse this information it is very difficult to read the individual contents and searching the contents form these documents line-by-line and word-by-word. The reason for this difficulty is the font characteristics of the characters in paper documents are different to font of the characters in computer system. As a result, computer is unable to recognize the characters while reading them. This concept of storing the contents of paper documents in computer storage place and then reading and searching the content is called Document Processing.

Sometimes in this document processing we need to process the information that is related to languages other than the English in the world. For this document processing we need a software system called Character Recognition System. This process is also called Document Image Analysis (DIA).

Thus our need is to develop character recognition software system to perform Document Image Analysis which transforms documents in paper format to electronic format. For this process there are various techniques in the world. Among all those techniques we have chosen Optical Character Recognition as main fundamental technique to recognize characters. The conversion of paper documents in to electronic format is an on-going task in many of the organizations particularly in Research and Development (R&D) area, in large business enterprises, in government institutions, so on. From our problem statement we can introduce the necessity of Optical Character Recognition in mobile electronic

devices to acquire images and recognize them as a part of face recognition and validation.

To effectively use Optical Character Recognition for character recognition in-order to perform Document Image Analysis (DIA), we are using the information in Grid format. . This system is thus effective and useful in Virtual Digital Library's design and construction.

II.   CURRENT EXISTING SYSTEM:

In existing approach it is a method to design a Text to Speech conversion module by the use of Mat lab by simple matrix operations. Firstly, by the use of microphone some similar sounding words are recorded using a record program in the Mat lab window and recorded sounds are saved in ".wave" format in the directory. The recorded sounds are then sampled and the sampled values are taken and separated into their constituent phonetics. The separated syllables are then concatenated to reconstruct the desired words. By the use of various Mat lab commands i.e. wave read, subplot etc. the waves are sampled and extracted to get the desired result. This method is simple to implement and involves much lesser use of memory spaces.

The existing navigation systems for the blind people require a precise GPS maps. This make them unusable in region where there are no GPS maps, they are not sufficiently accurate. Algorithm for GPS navigation for the visually impaired along a GPS track, which describe the path as a sequence of waypoints is proposed. The natural voice navigation, adaptive to the velocity and accuracy of the GPS data, start of the navigation from any waypoint, correlation of the direction of movement if it is necessary, return the user to the route if deviation is deviated, work with and without electronic compass, detection of the movement of the user in the opposite direction.

The overall problem statement includes:

- Traditional methods like Braille system.
- The blind people have to trace and read text, which is very slow and not very practical.
- English Braille, also known as grade 2 Braille is the Braille alphabet used for English.
- Existing OCR systems are not automatic.
- Require full-fledged computers to run and hence they are not effective.

This system became inefficient as Braille System is very slower and it's not practical enough. As Pic microcontroller and IR sensor are used it has acquired a quite disadvantage. The OCR system which was used earlier is not automatic.

III.   SYSTEM OVERVIEW:

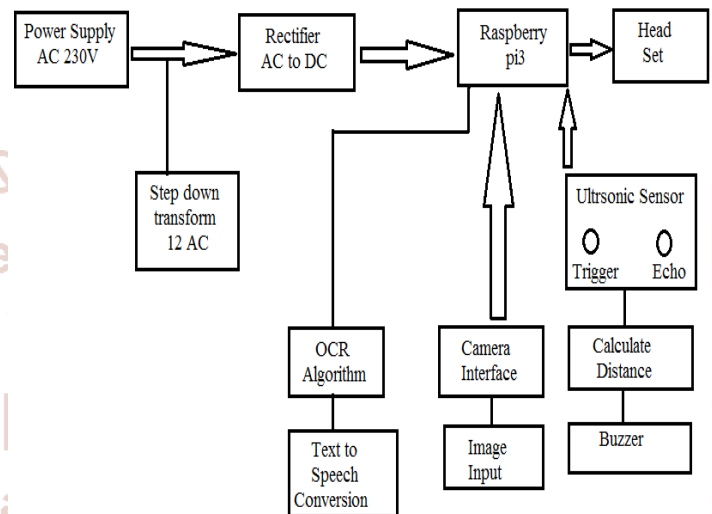### a.   System Architectural Implementation:

Figure.1 Architecture Diagram

We have described a prototype system to read printed text on hand-held objects for assisting blind persons. In order to solve the common aiming problem for blind users, we have proposed a motion-based method to detect the object of interest, while the blind user simply shakes the object for a couple of seconds. The automatic ROI detection and text localization algorithms were independently evaluated as unit tests to ensure effectiveness and robustness of the whole system. We subsequently evaluated this prototype system of assistive text reading using images of hand-held objects captured by ten blind users in person. Two calibrations were applied to prepare for the system test. First, we instructed blind users to place hand-held object within the camera view. Since it is difficult for blind users to aim their held objects, we employed a camera with a reasonably wide angle.In future systems, we will add finger point detection and tracking to adaptively instruct blind users to aim the object. Second, in an applicable blind-assistive system, a text localization algorithm might prefer higher recall by sacrificing some precision.

When our application starts running it first check all the devices and resources which it needs are available or not. After that it checks the connection with the

devices and gives control to the user. The GUI for the user has the following options. An optional label is used for displaying the image taken from the camera. A status box is for representing the detected data from the image. The capture button is to detect the data from the image.

The detect button is to detect the human from the video streaming in front of the camera. The audio jack port is the output port here. The Raspberry board comes with integrated peripherals like USB, ADC and Serial etc. On this board we are installing Linux operating system with necessary drivers for all peripheral devices.

b. Working Principle:

When capture button is clicked this system captures the product image placed in front of the web camera which is connected to ARM microcontroller through USB. After selecting the process button the captured label image undergoes Optical Character Recognition(OCR) Technology. OCR technology allows the conversion of scanned images of printed text or symbols (such as a page from a book) into text or information that can be understood or edited using a computer program.

The most familiar example is the ability to scan a paper document into a computer where it can then be edited in popular word processors such as Microsoft Word. However, there are many other uses for OCR technology, including as a component of larger systems which require recognition capability, such as the number plate recognition systems, or as tools involved in creating resources for SALT develop-ment from print based texts. In our system for OCR technology we are using TESSERACT library. Using Flite library the data will be converted to audio.

Camera acts as main vision in detecting the label image of the product or board then image is processed internally and separates label from image by using open CV library and finally identifies the product and identified product name is pronounced through voice.

Now it identifies received label image is converted to text by using tesseract library. Once the identified label name is converted to text and converted text is displayed on display unit connected to controller. Now converted text should be converted to voice to hear label name as voice through ear phones connected to audio jack port using flite library.

IV. MODULE IMPLEMENTATION

a. Image capturing and pre-processing.
b. Automatic text extraction.
c. Text recognition and audio output.

a. Image Capturing and Pre-Processing

The video is captured by using web-cam and the frames from the video is segregated and undergone to the pre-processing. First, get the objects continuously from the camera and adapted to process. Once the object of interest is extracted from the camera image and it converted into gray image. Use haar cascade classifier for recognizing the character from the object. The work with a cascade classifier includes two major stages: training and detection. For training need a set of samples. There are two types of samples: positive and negative.
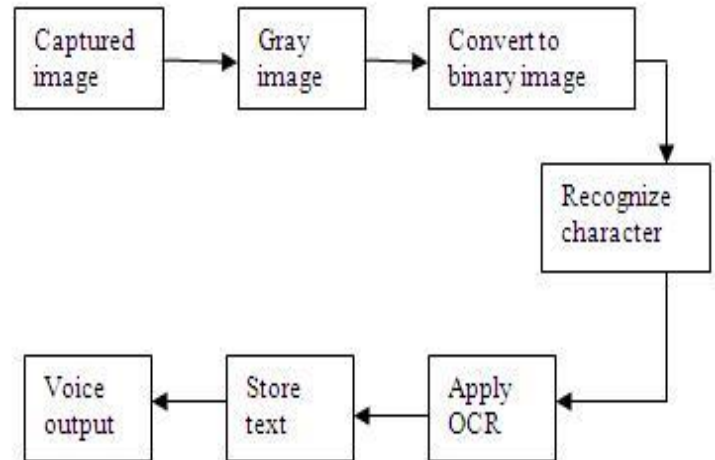

Figure.2 Image Capture and Pre-processing

To extract the hand-held object of interest from other objects in the camera view, ask users to shake the hand-held objects containing the text they wish to identify and then employ a motion-based method to localize objects from cluttered background.

b. Automatic Text Extraction

In order to handle complex backgrounds, two novel feature maps to extracts text features based on stroke orientations and edge distributions, respectively. Here, stroke is defined as a uniform region with bounded width and significant extent. These feature maps are combined to build an Adaboost based text classifier. The extraction information from audio and image source restricted to information execution from text. The actual transduction of audio and image data into text is the processing of OCR output.

c. TextRecognition and Audio Output

Text recognition is performed by off-the-shelf OCR prior to output of informative words from the localized text regions. A text region labels the minimum rectangular area for the accommodation of characters inside it, so the border of the text region contacts the edge boundary of the text characters. However, this experiment show that OCR generates better performance text regions are first assigned proper margin areas and binaries to segments text characters from background. The recognized text codes are recorded in script files. Then, employ the Microsoft Speech Software Development Kit to load these files and display the audio output.

Blind users can adjust speech rate, volume and tone according to their preferences. Are designed to easily interface with dedicated computer systems by using the same USB technology that is found on most computers. Static random-access memory (SRAM) is a type of a semiconductor memory that uses bi-stable latching circuitry to store each bit.

V. HARDWARE REQUIREMENTS AND IT'S IMPLEMENTATIONS

Processor - ARM CORTEX M3 SERIES
Speed   - 1.1GHz
RAM   - 1Gb(min)
Hard Disk - 1.5
Input   - Webcam, Ultra sonic
Output   - Headset, Buzzer

a. Raspberry Pi 2 Model B

Recently, Raspberry Pi 2 Model B has been lunched recently which Broadcom BCM2836 ARM Cortex-A7 Quad Core Processor has powered Single Board Computer running at 900MHz, 1GB RAM and 4 Quad USB ports. It is the advanced version of Model B and is 6 times faster than Model B Raspberry Pi. In addition, it has combined 4-pole jack for connecting your stereo audio out and composite video out and advanced power management.

b. Power supply unit

Apower supply (sometimes known as a power supply unit or PSU) is a device or system that supplies electrical or other types of energy to an output load or group of loads.

[1] Brief description of operation: Gives out well-regulated +5V output, output current capability of 100 mA

[2] Circuit performance: Very stable +5V output voltage, reliable operation

[3] Power supply voltage: Unregulated DC 8-18V power supply.

[4] Power supply current: Needed output current + 5 mA.

c. Webcam

A **webcam** is a video camera which feeds its images in real time to a computer or computer network, often via USB, Ethernet or Wi-Fi.

Features (LOGITECH WEBCAM C100):-
- Plug-and-play setup (UVC)
- Video capture: Up to 640 x 480 pixels
- Photos: Up to 1.3 megapixels (software enhanced)
- Frame rate: Up to 30 frames per second (with recommended system)
- Hi-Speed USB 2.0 certified

d. Ultrasonic sensor

Ultrasonic sensors are devices that use electrical–mechanical energy transformation, the mechanical energy being in the form of ultrasonic waves, to measure distance from the sensor to the target object. Ultrasonic waves are longitudinal mechanical waves which travel as a succession of compressions and rarefactions along the direction of wave propagation through the medium. Any sound wave above the human auditory range of 20,000 Hz is called ultrasound.

e. Buzzer

The PS series are high-performance buzzers that employ unimorph piezoelectric elements and are designed for easy incorporation into various circuits. They feature extremely low power consumption in comparison to electromagnetic units. Because these buzzers are designed for external excitation, the same part can serve as both a musical tone oscillator and a buzzer. They can be used with automated inserters. Moisture-resistant models are also available. The lead wire type(PS1550L40N) with both-sided adhesive tape installed easily is prepared. Two wires, red & black. Polarity matters: black=ground. The buzzer case supports the piezo element and has resonant cavity for sound

f. Optical Character Recognition (OCR)

Optical Character Recognition or OCR is the text recognition system that allows hard copies of written or printed text to be rendered into editable, soft copy

versions. It is the translation of optically scanned bitmaps of printed or written text into digitally editable data files.

The OCR process begins with the scanning and subsequent digital reproduction of the text in the image. It involves the following discrete sub-processes.

A flat-bed scanner is usually used at 300dpi which converts the printed material on the page being scanned into a bitmap image.
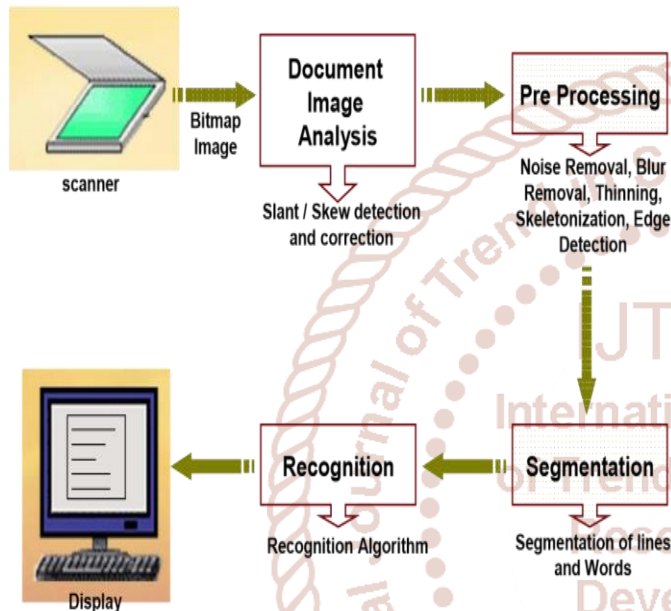

Figure.3 OCRProcesses

The bitmap image of the text is analyzed for the presence of skew or slant and consequently these are removed. Quite a lot of printed literature has combinations of text and tables, graphs and other forms of illustrations. It is therefore important that the text area is identified separately from the other images and could be localized and extracted.

In Preprocessing phase several processes are applied to the text image like noise and blur removal, banalization, thinning, skeletonization, edge detection and some morphological processes, so as to get an OCR ready image of the text region which is free from noise and blur.

If the whole image consists of text only, the image is first segmented into separate lines of text. These lines are then segmented into words and finally words into individual letters. Once the individual letters are identified, localized and segmented out in a text image it becomes a matter of choice of recognition algorithm to get the text in the image into a text processor.

Recognition is the most vital phase in which recognition algorithm is applied to the images present in the text image segmented at the character level. As a result of recognition character code corresponding to its image is returned by the system which is then passed to a word processor to be displayed on the screen where it can be edited, modified and saved in a new file format.

VI. SOFTWARE REQIREMENTS AND ITS IMPLEMENTATION

[1] Python IDE
IDE stands for Integrated Development Environment. It's a coding tool which allows you to write, test and debug your code in an easier way, as they typically offer code completion or code insight by highlighting, resource management, debugging tools, and even though the IDE is a strictly defined concept, it's starting to be redefined as other tools such as notebooks start gaining more and more features that traditionally belong to IDEs.
Because of all the features that IDEs have to offer, they are extremely useful for development: they make your coding more comfortable and this is no different for data science. However, given the fact that there aren't only the traditional IDEs to consider, but also new tools, such as notebooks, you might be wondering which development environment to use when you're just starting out with data science.

[2] Raspbian OS
Raspbian is a free and open operating system based on Debian which has been optimized for a Raspberry Pi hardware. Raspbian makes use of PIXEL(Pi Improved Xwindows Environment, Lightweight) as it's main desktop environment. It is composed of a modified LXDE desktop environment and the Openbox stacking window manager with a new theme and few other changes.

Installation of operating system on Raspberry Pi
Raspberry Pi is a small computer; hence operating system (OS) should be installed. As the Raspberry doesn't have hard drive, OS is installed in the external memory. For that, memory card (SD card) is used for the installation of operating system and all the required software and supporting files are stored in the same SD card.

Installation of required applications on Raspberry Pi

There are many applications that are needed to install in the Raspberry Pi for the completion of the thesis. For data logging, MySQL apache5 and php my admin are needed to install whereas for the web-page development, PHP is needed to install. Web page is used for the monitoring and managing purpose.

VII. MERITS OF THE PROPOSED SYSTEM:

- Flexible for blind people where this becomes an subsitute and a far better than Braille
- The printed books can be easily converted to digital text rather than typing the entire content.
- Effective to handle.
- Text to audio output can be achieved through this implementation.
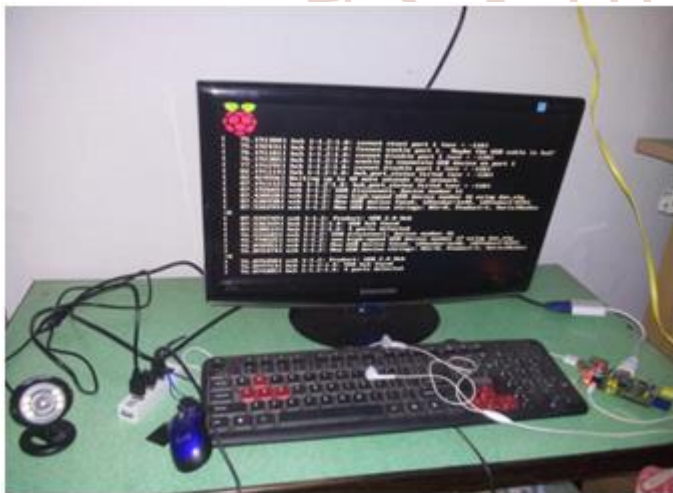- Blind people get courage to be more independent without expecting others help.

Figure.4 System Implementation

Figure.5 Final Intimation of an object for a blind people

FUTURE SCOPE:

The future enhancement of this project can be to fetch all the possible solutions for every disabled people to make them feel as normal humans without making them realize about their defects. This project can be made affordable to all with cost efficient. The range for identifying the obstacles can be increased further so that the blind people can be pre-notified and they can become more independent while they are in mobility.

CONCLUSION:

In this paper, we have described a prototype system to read printed text on hand-held objects for assisting blind persons. In order to solve the common aiming problem for blind users, we have proposed a motion-based method to detect the object of interest, while the blind user simply shakes the object for a couple of seconds. This method can effectively distinguish the object of interest from background or other objects in the camera view. To extract text regions from complex backgrounds, we have proposed a novel text localization algorithm based on models of stroke orientation and edge distributions. The corresponding feature maps estimate the global structural feature of text at every pixel. Block patterns project the proposed feature maps of an image patch into a feature vector. Adjacent character grouping is performed to calculate candidates of text patches prepared for text classification. An Adaboost learning model is employed to localize text in camera-based images. Off-the-shelf OCR is used to perform word recognition on the localized text regions and transform into audio output for blind users. Our future work will extend our localization algorithm to process text strings with characters fewer than three and to design more robust block patterns for text feature extraction. This project will also extend our algorithm to handle no horizontal text strings. Furthermore, we will address the significant human interface issues associated with reading text by blind users. It has been developed by integrating features of all the hardware components and software used. In this paper, the camera acts as input for the paper. As the Raspberry Pi board is powered the camera starts streaming. The streaming data will be displayed on the screen using GUI application. When the object for label reading is placed in front of the camera then the capture button is clicked to provide image to the board. Using Tesseract library the image will be converted into data and the data detected from the image will be shown on the status bar. The obtained data will be

pronounced through the ear phones using Flite library. Presence of every module has been reasoned out and placed carefully thus contributing to the best working of the unit. In this by using highly advanced ARM11 board this paper has been implemented.

REFERENCES:

1. X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proc. Comput. Vision Pattern Recognit., 2004, vol. 2, pp. II-366–II-373.

2. S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text extraction and document image segmentation using matched wavelets and MRF model," IEEE Trans Image Process., vol. 16, no. 8, pp. 2117– 2128,Aug. 2007.

3. K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12,pp. 1631–1639, Dec. 2003.

4. N. Giudice and G. Legge, "Blind navigation and the role of technology,"in The Engineering Handbook of Smart Technology for Aging, Disability,and Independence, A. A. Helal, M. Mokhtari, and B. Abdulrazak, Eds. Hoboken, NJ, USA: Wiley, 2008.

5. World Health Organization. (2009). 10 facts about blindness and visual impairment.

6. Advance Data Reports from the National Health Interview Survey (2008).

7. International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011).

8. X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process ,vol. 13, no. 1, pp. 87–99, Jan. 2004.

9. D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," IEEE Trans. Syst., Man, Cybern.,vol. 40, no. 1, pp. 25–35, Jan. 2010.

10. B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. Comput. Vision Pattern Recognit.,2010, pp. 2963–2970.

11. Y. Freund and R. Schapire, "Experiments with a new boosting algorithm,"in Proc. Int. Conf. Machine Learning, 1996, pp. 148–156.

12. An overview of the Tesseract OCR (optical character recognition) engine, and its possible enhancement for use in Wales in a pre-competitive research stage Prepared by the Language Technologies Unit (Canol-fan Bedwyr), Bangor University April 2008.

13. A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition:ICDAR Robust Reading Competition Challenge 2: Reading text in scene images," in Proc. Int. Conf. Document Anal. Recognit., 2011, pp. 1491–1496.

14. KReader Mobile User Guide, knfb Reading Technology Inc. (2008).

15. S. M. Lucas, "ICDAR 2005 text locating competition results," in Proc.Int. Conf. Document Anal. Recognit., 2005, vol. 1, pp. 80–84