

Abstract

There are some technical systems which have emerged nowadays to help the blind persons. To introduce them, the first which can come in our mind is 'Braille'. To print this Braille script which is expensive process. The solution would be a wearable assistive device for the blind which converts the text into acoustic output enabling the user to read any sort of text for that a standalone Raspberry Pi based system with finger mounted camera that can help the visually impaired people in word based reading of the textual data pointed to by the finger. The system consists of a LDR sensor which detect the ambient light condition and turns on the light accordingly, webcam that captures images which are enhanced. Following this the word pointed by the finger is extracted using a novel methodology and given to an Optical Character Recognition (OCR) engine. Subsequently, the textual output is given to a Text to Speech (TTS) converter to obtain audio via an audio output device such as earphones or speaker .

Contents

| | <u>page.no</u> |
|---|--------------------------------|
| 1. Introduction | 3 |
| 1.1 Image Processing | 3 |
| 1.2 AI Enabled IoT | 4 |
| 1.3 Text to speech | 5 |
| | |
| 2.Literature Survey | 7 |
| 2.1 Existing solutions and drawbacks | 7 |
| 2.2 Technique | 7 |
| 2.3 OpticalCharacterRecognition | 8 |
| 2.4 Binarization | 8 |
| 2.5 Text to Speech Conversion | 9 |
| | |
| 3.Objectives | 12 |
| | |
| 4.Methodology | 13 |
| 4.1 OCR Based Facilitator for the Visually Challenged | 13 |
| 4.2 BLIND READER: An Intelligent Assistant for Blind | 14 |
| | |
| 5.Results | 15 |
| | |
| 6. references | 16 |

Chapter – 1

Introduction

1.1 Image Processing

Image processing is a way to convert an image to a digital aspect and perform certain functions on it, in order to get an enhanced image or extract other useful information from it. It is a type of signal time when the input is an image, such as a video frame or image and output can be an image or features associated with that image.

Advantages

- It helps to improve images for human interpretation.
- Information can be processed, extracted from images for machine interpretation.
- The pixels in the image can be manipulated to any desired density and contrast.
- Images can be stored and retrieved easily.

Application:

- Image Enhancement
- X-Ray Imaging
- Medical CT
- UV imaging

Image Enhancement:

Image enhancement is the process of adjusting digital images so that the results are more suitable for display or further image analysis. The image captured through sensor may contain the shadows that will be removed, The result is inverted to get black text on white image.

Exponential transform:

In order to obtain images with distinct edges of the letters, an anti logarithmic transform is applied. The result of the above process is inverted to get white text on black image. The text file is removed and then given to the TTS converter.

1.2 AI Enabled IoT

IoT is about sensors implanted into machines, which offer streams of data through internet connectivity. All IoT related services inevitably follow five basic steps called create, communicate, aggregate, analyse, and act.

While IoT provides data, artificial intelligence acquires the power to unlock responses, offering both creativity and context to drive smart actions. As the data delivered from the sensor can be analyzed with AI, businesses can make informed decisions.

Benefits of AI Enabled IoT

- Boosting Operational Efficiency
- Better Risks Management
- Triggering New and Enhanced Products & Services
- Increase IoT Scalability
- Eliminates Costly Unplanned Downtime

Applications of AI Enabled IOT

- Manufacturing Robots
- Autonomous Vehicles
- Retail Analytic
- Smart Thermostat

1.3 Text to speech

eSpeak is a compact open source software speech synthesizer for English and other languages, and is used to convert the text file to audio output which is heard via an audio output device such as earphones. Python 3 wrapper for eSpeak is utilized to implement the algorithm.

Assistance to blind

In our world information is generally available in the form of books and documents. It is fully usable for the sighted people. When an era has come of printing it facilitates the sighted people partially to acquire knowledge. A major problem for a blind or visually impaired person (BVI) to interact with the world to share knowledge. For them information has to be in a special tactile language or in voice format. They are affected in every works of their daily life. Nowadays technology helps them to overcome this difficulty to some extent. The most difficult task for them is reading text from the books or documents.

For the blind or visual impaired (BVI) person, it is a very difficult job to acquire information from the world. One feasible way in order to perform that job is that someone will help him to read aloud the context. Another way to get the information is by giving a feelings of the information. The later technique is built through a representation of the information on a paper or a substantial surface so that a blind person can feel and recognise the information. But it is a very difficult work to make all the books or documents available for the blind people. An application is developed that read aloud the context of the document which is represented virtually. This later technique reduces the resources to help the blind to acquire knowledge but in this manner he can be assisted only audibly.

Techniques

- The camera is set up in a wearable hand device in such a way that it can capture the image.
- The device consists of an index mounted webcam connected with Raspberry Pi 3 model A+. The image captured by the camera is processed by the computer.
- Then the captured image is enhanced and shadows are removed.
- After enhancement, exponential transform is done on the image.
- The words are distinctly visible in the resulting image. Once the clear image is obtained, it is de-skewed so that it can be accurately converted by the OCR Engine.
- Once the corrected image is obtained the individual words have to be located. Thus contours are found on the image after dilation.

- Bounding boxes are drawn on each of the words. The fingertip is detected by converting the image into HSV model and extracting the extreme point on the isolated contour of the finger. The line from the fingertip to the end of the image is drawn.
- The coordinates of the boxes obtained and the fingertip are compared. The contour closer to the fingertip is extracted. This is provided to the OCR Engine for converting it to text.
- The final image that is given to OCR Engine, and the word isolated is of the same size as that in the input image.
- The algorithm is implemented in Raspberry Pi 3 Model A+. The entire process from capturing the image, processing to extract the intended word and final audio output via device such as earphones.

Chapter – 2

Literature Survey

The review of related literature shows that the system design for wearable assistive device for the blind successfully address the challenges faced by visual impaired or blind person in reading the text without any braille script. In the literature survey, methods and technology are been explained.

2.1 Existing solutions and drawbacks

There are some technical systems which have emerged nowadays to help the blind persons. To introduce them, the first which can come in our mind is ‘Braille’. To print this Braille script, Braille Printer is a kind of embosser which is a technical hardware that can print the hardcopy of braille, which is expensive process. Reading speed is significantly slower than for print reading or recorded materials. -Braille materials are expensive to produce. -Spelling skills require special attention because of the constructions. -Specific elements of Braille present serious problems to a number of multiple handicapped individuals

2.2 Technique

Word Based Text Extraction Algorithm Implementation in Wearable Assistive Device for the Blind

Blindsights Text Detective also reads text from camera in phone. However it needs clear text placed at a particular distance from the camera for optimum results. Such mobile based software thus require proper alignment, lighting and focus for accurate results and helps in converting a block of text in an image as a whole, which may not provide relevant information that the user intents to read

requires a Web Camera to capture the image, which is sent to a single board computer, Raspberry Pi 3 Model A+ and finally the audio output delivered via earphones.

The finger-worn camera helps in obtaining focused images at a fixed distance and helps the user to utilize the sense of touch when scanning the surface of the document

The image captured is re-sized to 800x600 so that the subsequent procedures act on a uniformly sized images and give similar results for all inputs. methods of image

enhancement such as RGB format, histogram equalization, brightness and contrast adjustment using gain and bias parameters, gamma correction have also been implemented.

2.3 OpticalCharacterRecognition

OCR is a reader which recognizes the text character of the computer which may be in printed or written form. the OCR templates of each character are used to recognize the character i.e. scanning process is carried out. After this, the character image is translated in ASCII code which is further used in data processing.

OCR follows some basic architectural framework module as we can see it. It thresholds with the image recognition where each character from the image is being sent for recognition. There are some pre-processes involved to make the image noise free which involves process like binarization, skew correction and normalization . Where we can see the image undergoing some enhancements such as filtering out noise and contrast correction. After this, the real framework of OCR starts with the segmentation process, as the name says everything here the characters is segmented so that they can be texted separately. At the next level of the framework the segmentation is followed by the text lines, words and characters in the image. This level is assisted by some connected component analysis Information and projection analysis to again assist text segmentation. Further the process is followed by feature extraction which is concerned with the representation of the object. All the real- time working on the OCR lies in the Feature extraction which can be titled as the ‘Heart’ of OCR.

2.4 Binarization

A process of binarization is converting a grayscale image into a binary image using thresholding is known as Binarization. Before making the phenomenon acknowledged taking you the decades back, it was used in faxes now the binarization is really easy but typical to understand in simple words we know that the image contains pixels which are stored bit by bit, now in image there are two colours black(0)and white(1) what does it do is that pixels which are grey it makes them set(accepted) and the pixels with white are made unset further in the process the pixels which are in set mode and are near are combined to make some acceptable character. The important characteristic of the binarization is the distance transformation by which the unset pixel is distanced from another set pixel.

2.5 Text to Speech Conversion

We know that the artificial production of human sound is nothing but the Speech. Some mechanism is used to generate speech this mechanism is known as a speech synthesizer. In that, we can give the sound of human and robot also but generally, we use the robot. The text to speech system is divided into two sections: namely a back-end and front-end. The front-end has two vital tasks. Now the process starts with the conversion of the raw text like abbreviations, symbols, numbers and it a little bit similar to written out words.

2.6 Segmentation

Segmentation is done as the image consists of a number of sentences and each line contains a certain number of the words. Then this word is formed by the number of characters. Hence we can say that the segmentation is a process of partitioning the digital image into the segments. In the segmentation process each line, each word, each character is segmented. There are some inevitable difficulties in segmentation process like image quality is less, every computer system has some different fonts, cursive writing, etc. this affects the efficiency of the segmentation process thus the best practice is done here to eliminate those problems.

Literature Survey Table

| Author and year | Title of the paper | Methodology | Findings | Shortcomings |
|---|---|---|--|--|
| Arunima B Krishna Meghana Hari Dr. Sudheer A.P -2019 | Word Based Text Extraction Algorithm Implementation in Wearable Assistive Device for the Blind. | Capturing the image, processing to extract the intended word and final audio output to the earphone or to the audio jack. | Provides assistance to blind or visually impaired person to read a book. | The device assists the blind through a voice where voice is an intended word pointed by the user and it takes approximately 5s. |
| Trupti Shah Sangeeta -2019 | Efficient Portable Camera Based Text to Speech Converter for Blind Person | Detecting the object region, localization of text on object, extraction of text and text to speech conversion. | The experiment and training are performed on Synth 90k word dataset and using OCR and CRNN a model has been developed. | Combination of OCR and CRNN overcomes drawback of individual and gives better result which successfully recognizes text image and convert it into speech for blind person. |
| Shalini Sonth Jagadish S 2017 | OCR Based Facilitator for the Visually Challenged | Acquisition of image and pre-processing, extracting a text from the image and text-to-speech conversion. | The image is captured after 10 seconds of button is being pressed and sent to processing. | Applying Otsu's threshold and Gaussian blur for denoising the image. |
| Sneha.C. S.B.Gundre -2019 | OCR Based Image Text To Speech Conversion Using MATLAB | Optical Character Recognition,Binari zation,Segmentatio n,Feature Extraction,Recogni tion. | Text is extracted from the input image using OCR then the extracted text is converted into Speech using TTS in MATLAB | The advantage of this system does not need the internet connectivity that is a basic necessity for other TTS system like a Google TTS in Google keep. |

| Author and year | Title of the paper | Methodology | Findings | Shortcomings |
|---|---|---|--|--|
| Thomas Portele, Jiirgen Kramer-2018 | Adapting A Tts System To A Reading Machine For The Blind | Synthesis systems that convert orthographic text into speech usually make assumptions about the input that are no longer valid when used in combination with a scanner and OCR softwaxe | The combination of a scanner and OCR software might sometimes lead to obscun input for the synthesis system; this is especially the case when scanning faxes. In our experience nearly everything is possible | Speech synthesis systems are developed with the main focus on speech quality. But ease of use and robustness are as important for many users as high-quality speech. |
| Akash Singh, Vaishali Kushwaha, Shilpi Mishra -2018 | An Efficient Auxiliary Reading Device for Visually Impaired | Optical Character Recognition (OCR) and finally narration of converted text using Text to Speech synthesizer (TTS). | The device performs its operation through the serial integration of four operations in-line with next to each scanning, image pre-processing, extraction of text using microcontroller using series of algorithms and various engines like OCR | Bottom up design methodology was used in the development of the device. The frame and casing of the goggle were designed using lightweight materials. |
| Prabhakar Manage, Veeresh Ambe Prayag Gokhale -2019 | An Intelligent Text Reader based on Python | The input image is enhanced using Image processing techniques. The Tesseract OCR (Optical Character Recognition) engine embedded in the Raspberry Pi searches for the text in an improved image and converts it into digital document | The main objective of the designed product is to assist the visually impaired people and serve them economically and efficiently | Python based portable virtual text reader is discussed. This reader is used to read- out the text aloud for various applications. |

Chapter – 3

Objectives

1. To build a wearable device
2. To Check Ambient Light Conditions
3. To finger Tip Detection and Word Extraction:
4. Image To Text conversion

Chapter – 4

Methodology

4.1 OCR Based Facilitator for the Visually Challenged

- a camera based framework built on the Raspberry Pi, integrated with Image processing algorithms, OCR and Text-to-Speech (TTS) synthesis module.
- The preprocessing stage includes binarization, de-noising, deskewing, segmentation and feature extraction
- OCR used in this project is Google Tesseract
- TTS employed is Pico.

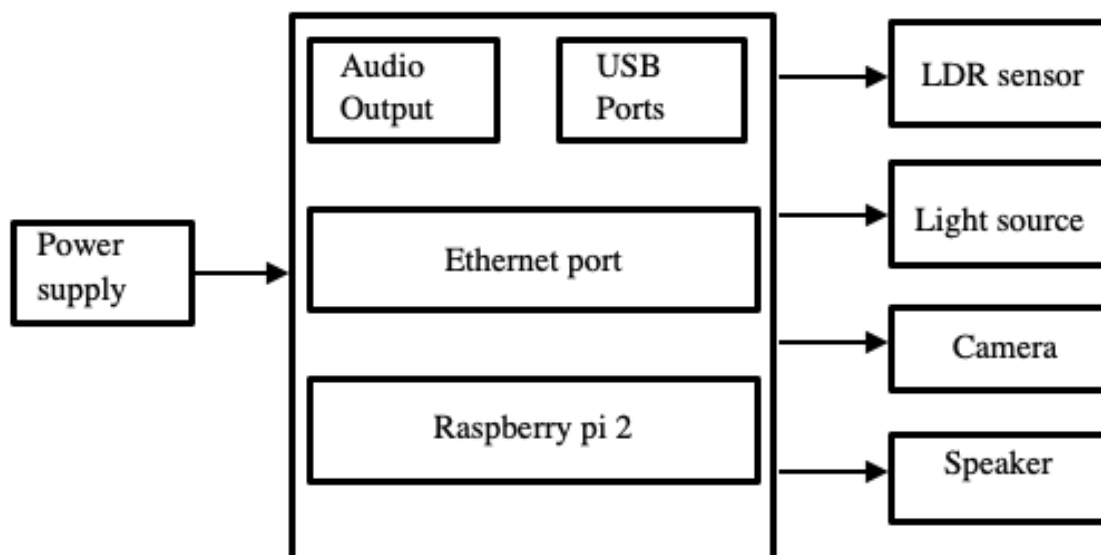
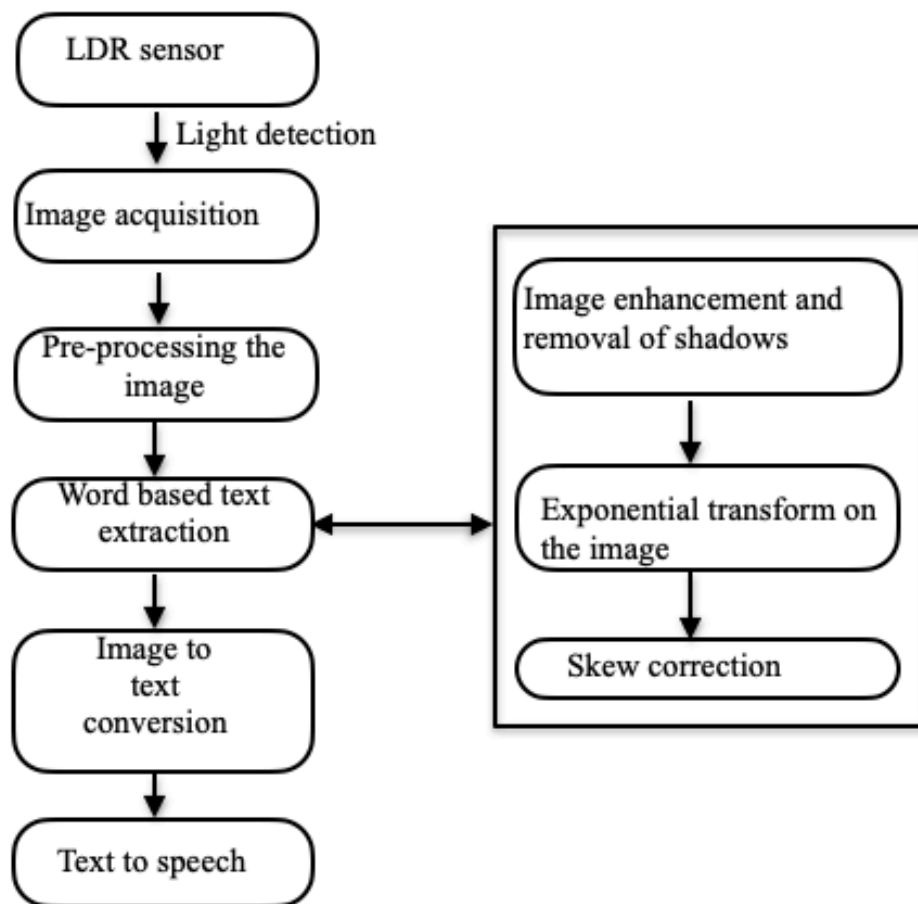


Fig 4.1 : System hardware design

The system consists of a LDR sensor which detect the ambient light condition and turns on the light accordingly, webcam that captures images which are enhanced. Following this the word pointed by the finger is extracted using a novel methodology and given to an Optical Character Recognition (OCR) engine. Subsequently, the textual output is given to a Text to Speech (TTS) converter to obtain audio via an audio output device such as earphones or speaker .

4.2 BLIND READER: An Intelligent Assistant for Blind

He can use his fingers to move around the screen over a virtual document and get the information audibly. In this case, he gets the access to all the information of the world as most of it is stored virtually. Besides it is a human cognizable method through which human can interact with the system to acquire knowledge.



Finger Reader. It is a wearable device in finger. It helps the BVI to access the plain printed text. People who ware this device, scan a text line with their finger and in a result they get an audio feedback of the words and also a haptic sense of the layout. These senses may be the start or end of the line, new line and so on. It also alerts the reader if he moves away from the baseline thus it helps him maintain straight canning.

Chapter – 5

Results

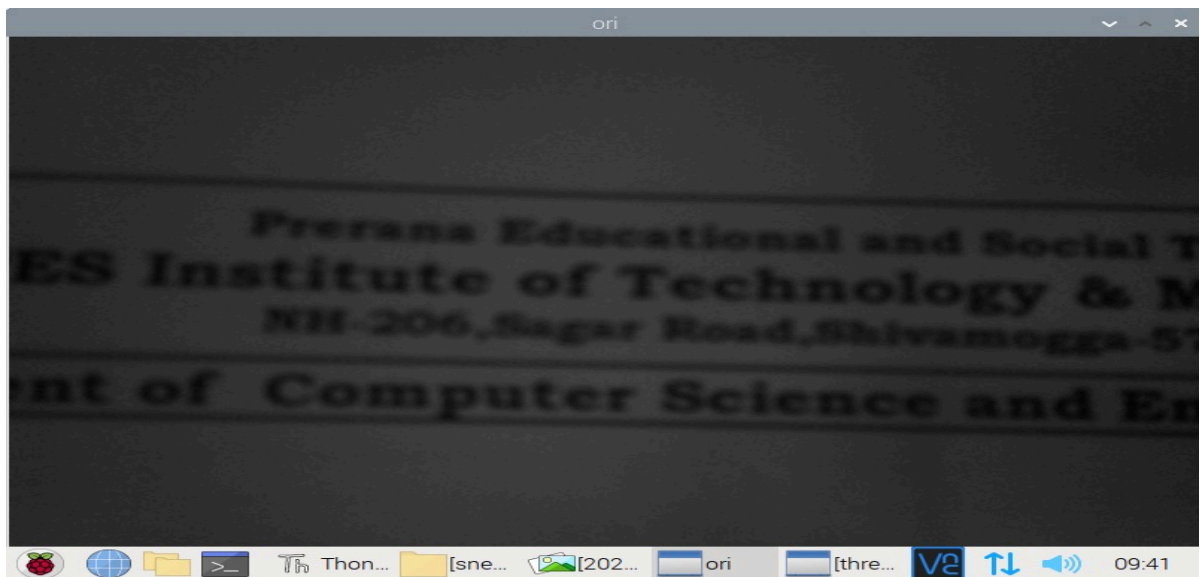


Fig 5.1 : Image captured

The above figure depicts image captured by the raspberry PI camera of a plain text from a printed document

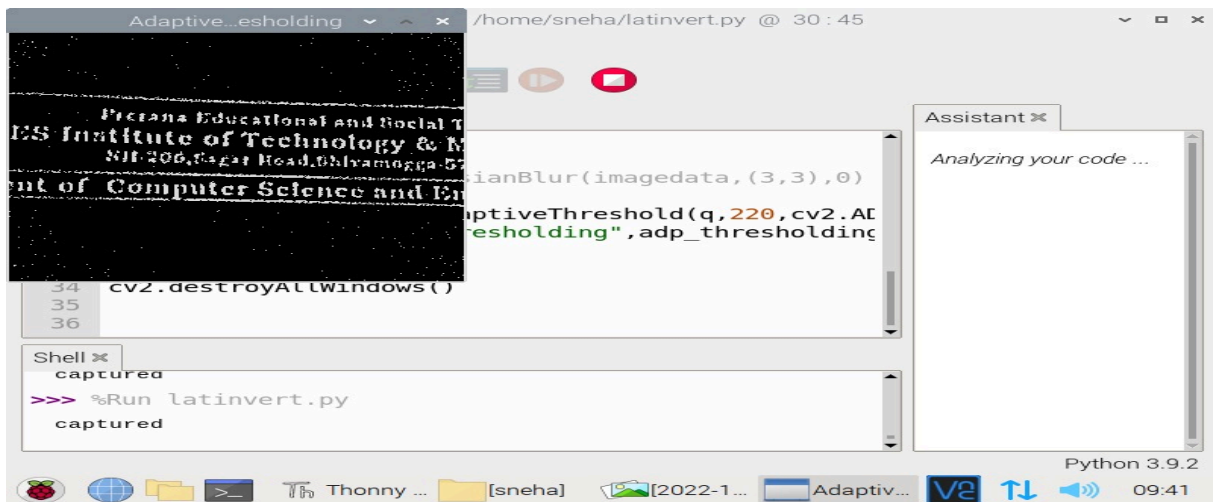


Fig 5.2 : Image after binarization

The above figure depicts resulting of binirization of the image captured from the raspberry PI camera

References:

1. Visual impairment and blindness 2010," 2013. [Online]. Available: [http:// www.who.int/blindness/data_maps/](http://www.who.int/blindness/data_maps/)
2. Trupti Shah and Sangeeta Parshionikar :Efficient Portable Camera Based Text to Speech Converter for Blind Person,
3. Arunima B Krishna, Meghana Hari and Dr. Sudheer A.P : Word Based Text Extraction Algorithm Implementation in Wearable Assistive Device for the Blind,
4. Shalini Sonth and Jagadish S. Kallimani : OCR based facilitator for the visually challenged,
5. Prabhakar Manage,Veeresh Ambe, Prayag Gokhale, Vaishnavi Patil, Rajamani M.Kulkarni, Preetam R. Kalburgimath. : An Intelligent Text Reader based on Python,
6. M. A. Hersh and M. A. Johnson, Eds. Berlin : D. Keating, Assistive technology for visually impaired and blind people : Springer London, 2008.