An Intelligent Reading Assistance For Visual Impaired People

Miss. Prajakta S. Udgave[1], Prof. Mr. S. G. Tamhankar[2]
[1, 2] Dept of Electronics Engineering
[1, 2] Walchand College of engineering, Sangli, Maharashtra, India

Abstract- *Usually many people suffered from visual disabilities. Information access to the written transcription is one of the challenge for blind and visual impaired people except it is represented in non-visual form like braille lippy. A intelligent reading assistant is a need of an effective system for blind and visual impaired people. This paper proposes intelligent reading assistant for visual impaired that captures an image through camera, extract only part of image that contains only text and converts that text content into speech. The variety of image processing techniques are performed on captured image to extract only region of image that contain the text. The Pytesseract OCR (Optical Character Recognition) function of OpenCv for converting text image into text. Espeak text-to-speech synthesizer for listen these extracted speech. The aural output is heard through the speakers or earphones.*

Keywords- Raspberry Pi, Camera, speaker, Image processing techniques, Tesseract OCR, TTS engine.

I. INTRODUCTION

In our world, information is usually available in the form of books and documents. It is completely usable for the normal people. And the printed materials only facilitate to sighted people partially to acquire the knowledge. But blind and visual impaired people have to face major problem of sharing any knowledge while interacting with the world. They are affected in every works of their daily life. Nowadays technology helps them to overcome this difficulty to some extent. For them reading text from the book or any document is most difficult. Normally those people uses braille lippy for reading the documents and books which are less reliable to read. Braille materials are complicated to produce and less readily available in braille format for use. There is need for developing device that could bring relief to the excruciating task that the blind and visual impaired people have to face from beginning to end. This paper proposed the system using OCR (Optical Character Recognition) is used to digitize and reproduce the text. Digitizing the text will helps to reduce the storage space. And Espeak text-to-speech synthesizer is useful for blind and visual impaired people to listen the text document without internet. Camera based system will help for people to read any text document.

II. PROPOSED SYSTEM

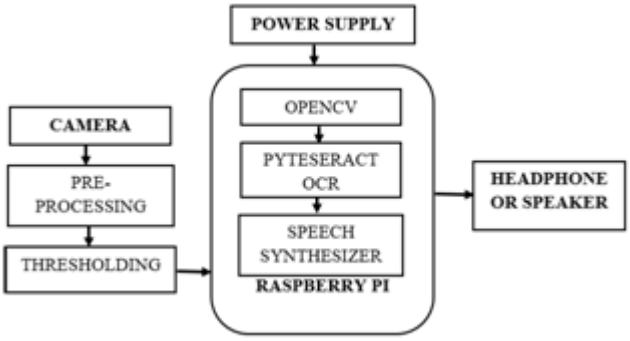


Fig 1: Block diagram of proposed system.

1) Acquisition stage

In the initial phase, device will move over the printed text and image is captured. And image captured from the camera is used as input to the image preprocessing unit. The challenges occurred due to camera fixation/ position problems like blurring or skewness of text due to motion of user, perspective distortion due to different angles of the object formed with camera lens axis. So, care must be taken while fixing camera of the system. To produce desired output, acquisition must be proper.

2) Pre-Processing

Basically the aim of image pre-processing phase is an improve the image data or enhance the image features for further processing because acquired images during image acquisition may not be directly suitable for identification and classification purposes because of some factors, such as noise, weather conditions, and poor resolution of an images etc. Image preprocessing stage has three basic operations which are necessary in each image related operation and these are

a) Grayscale Conversion

The input image is in RGB (colour) form. RGB (colour) image is the combination of three primary colours are red, green and blue. Different proportion of red, green and blue gives you colour image. RGB image is basically an M*N*3 array of colour pixel, where each colour pixel is a triplet which corresponds to red, blue and green colour component of RGB image at a specified spatial location. Gray scale images have one colour which is a shade of gray (shades of white and black) in the range of 0 to 255. A grayscale image is basically M*N array whose scaled values represents the intensities.We must obtain the values of its red, green, and blue (RGB) primaries in linear intensity encoding, by gamma expansion, to convert any color image to a grayscale representation of its luminance.

b) Filtering

The filtering is performed to smoothen the image. It reduces or remove the visibility of noise by smoothing the entire image. Here filtering is carried out by gaussian filter since it is used to `blur' images and remove detail and noise. Gaussian filter does better job of retaining the edges of the image.

c) Thresholding

Thresholding is the simplest method of segmenting images. To create binary images from a grayscale image, thresholding can be used.Thresholding is the process of dividing an image into two or more classes of pixels i.e. foreground and background. Black pixels correspond to background and white pixels correspond to foreground. In digital image processing, binary image is used as mask or as a result of some operations like segmentation, thresholding.
Otsu method is used for thresholding. The algorithm of OTSU is

1. Determine histogram and probabilities of each intensity level.
2. Set up initial class probability and initial class means.
3. Step through all possible thresholds maximum intensity.
4. Compute within class variance.
5. Compute between class variance.
6. Desired threshold corresponds to the maximum value of between class variance.

3) Segmentation

Image segmentation is to divide a digital image into many regions with regard to a specific application. The segmentation is based on measurements taken from the image, may be grey level, color, texture, depth or motion. Here edge-based segmentation is properly suitable. As edge detection is an important step in image processing, it is necessary to point out the true edges to get the best results from the matching process. Hence, it is important to choose edge detectors that fit best to the application. In this way canny edge detector is chosen. This edge detectionoperator that uses a multistage algorithm to detect a wide range of edges in images.

Canny Algorithm:

Without affecting any features of an edges in the image canny edge detector is most defined method to find the edges by isolating the noise from the image. The canny has general criteria for optimal edge detection include:

i. Detection of edges with low error rate.
ii. A good localization, which ensures detected edge points should accurately localize on the center of the edges.
iii. A single response, which ensures that given edges in the image should be marked once.

4) Text Recognition

The aim of the processing module is to prepare the image for recognition. Processing involves filtering, grayscale conversion, thresholding, segmentation and skew correction of the text. It helps to recognize the text with accuracy. Here text extraction and recognition is performed by using tesseract OCR. Tesseract OCR is an open source OCR engine that recognizes the text inside the images and converting it into an electronic forms. These images could be of handwritten text or printed text like documents, receipts, name cards etc., or scene images as well as data from screen reader (e.g. smartphone, display screen or computer). OCR has two parts to it. The first part is text detection, in that the textual part inside the image is determined. This localization of text inside the image is vital for the second part of OCR, text recognition, where the text is extracted from the image. Using these techniques, you can extract text from any image. The digitization using OCR obviously has widespread advantages like easy storage and manipulation of the text. Tesseract can detect over 123 languages and read the text from right-to-left. Tesseract 4.0 has added a new OCR engine that uses a neural network system based on LSTM (Long Short-term Memory), one of the most effective solutions for sequence prediction problems. Even thoughit's previous OCR engine using pattern matching is still available.

5) Skew correction of text

While capturing the text image skew is naturally introduced into captured image. And skew effect is bad on character segmentation and recognition task. Therefore detecting and correcting the skew is important. And some affine transformations such as translation, scaling, rotation and retranslation are used to correct skew of the text images. The following is the flowchart of the skew correction of the text.

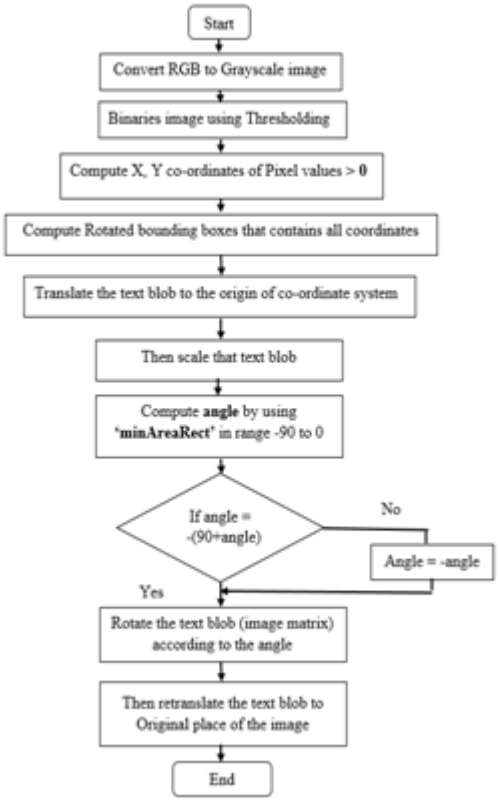


Fig 2: Flowchart of Skew Correction of the Text.

1) Text-to-Speech Synthesis

Speech synthesis is the artificial production of human speech. A computer system used for this purpose and can be implemented in software or hardware products is called a speech synthesizer.

A text-to-speech (TTS) system or software is used to convert word from a computer document into audible speech, spoken through the computer speaker. The quality of a speech synthesizer is judged by two characteristics. The speech should be similar to the human voice and it should be clearly understood. An intelligible text-to-speech synthesizer allows people with visual impairments or reading disabilities to listen to written words on a home computer. Here we are using Espeak text-to-speech synthesizer to generate the speech. Espeak text-to-speech synthesizer is a compact open source speech synthesizer for English and other languages. eSpeak uses a "formant synthesis" method. Formant synthesis produces artificial, robotic sounding speech and it does not use human speech at runtime. Especially Formant synthesizers are usually smaller programs therefore it can be used in embedded systems, where memory and microcontroller powers are limited.

III. FLOWCHART

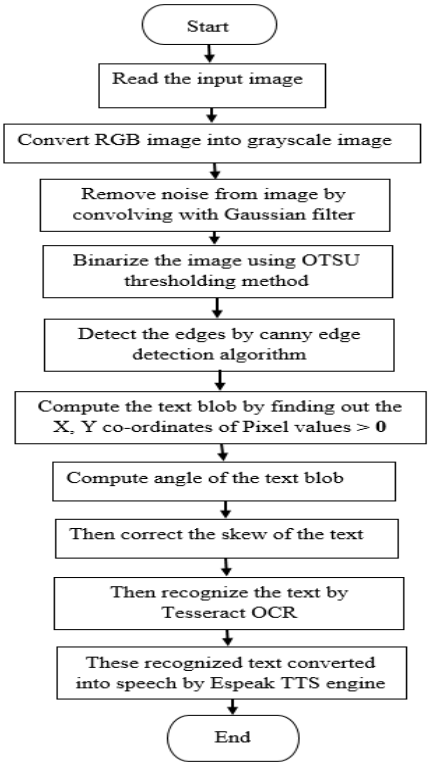
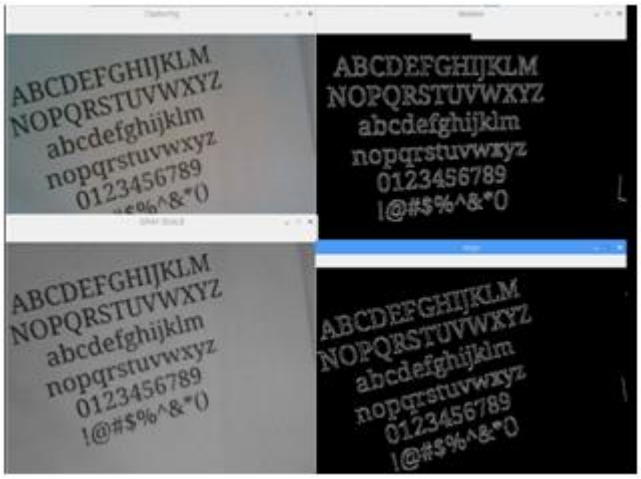


Fig 3: Flowchart of Proposed system.

IV. SYSTEM DEVELOPMENT

Fig 4: Implementation of proposed system.

a. Simulation results



Above figure shows the captured image, is in the RGB form and this image is converted into grayscale image with 255 gray levels.

Then by Otsu thresholding, it is converted into binary image which has two assigned pixel values are black and white. Then canny edge detection is used for fine edge detection from the image.

If skew is detected in the image then deskew the image by using skew correction of the text.

b. Simulation output result

Figure below shows the output of tesseract OCR which is obtained after processing. The Output should be in the text format of the captured image.

```
==================== RESTART: /home/pi/Project/
[INFO] angle: 13.981
ABCDEFGHIJKLM
NOPQRSTUVWXYZ
abcdefghijklm
nopqrstuvwxyz
0123456789
1@#$%A&*0
>>>
```

V. QUANTIFICATION OF RESULTS

The system was tested for the different kinds of text like printed, handwritten as well as compact blob and combination of picture with text in one image also different font style in a word. It accurately recognize and its accuracy is given in below table**.**

Table 1: Quantification of Results

Sr. No.	Type of Text	Recognition Accuracy
1.	Printed text	98%
2.	Handwritten	90%
3.	Running lippy	70-80%

VI. CONCLUSIONS

In this paper, design and implementation of text recognition and speech generation system has discussed. For conversion of text image to text and that text into speech, a simple and efficient algorithm we have proposed. Text extraction and recognition is performed by using tesseract OCR that improves accuracy as well as increases the speed of the system. After the successful text recognition, these text is converted into audio output by using Espeak TTS synthesizer without using an internet.

VII. FUTURE SCOPE

Database can be formed by capturing the face image and face recognition can be done every time the person is detected to identify his identity. Also the locations can be added and the path to reach there for blind people. This whole module can be converted into smart glasses so it could be easier for blind people to carry along. Even barcodes of shelf products and money bills, different colours can be recognized.

REFERENCES

[1] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang, “EAST: An Efficient and Accurate Scene Text Detector”,arXiv: 1704.03155v2 [cs.CV] 10 Jul 2017

[2] Akhilesh A. Panchal, ShrugalVarde, M.S. Panse, “Character detection and recognition System for visually impaired people”, 978-1- 5090-0774-5/16/© 2016 IEEE.

[3] Sandeep Musale, VikramGhiye, “Smart reader for visually impaired”, 978-1-5386-0807- 4/18/ ©2018 IEEE.

[4] ShahedAnzarusSabab, Md. Hamjajul Ashmafee, “Blind reader: An intelligent assistancefor blind”, 978-1-5090-4090-2/16/ ©2016 IEEE.

[5] Sanjana. B, J. RejinaParvin, “Voice assisted text reading system for visually impaired person using TTS method”, IOSR Journal of VLSI and Signal Processing (IOSR-JVSP).

[6] Asha G. Hagargund, SharshaVanriaThota, MitadruBera, Eram Fatima Shaik, “Image to speech conversion for

visual impaired", International Journal of Latest Research in Engineering and Technology (IJLRET)

[7] K NirmalaKumari, Meghana Reddy J, "Image text to speech conversion using OCR technique in Raspberry Pi", IJAREEIE 2016, vol. 5,

[8] Liu, D., & Yu, J. (2009). Otsu Method and K- means. 2009 Ninth International Conference onHybrid Intelligent Systems. doi:10.1109/his.2009.74

AUTHOR'S PROFILE

Miss. Prajakta S. Udgave. She is an Mtech student of Walchand College of Engineering, Sangli. She completed her Graduation in Electronics engineering from DKTE, Ichalkaranji. She has a keen interest in the field of digital image processing.

Mr. S. G. Tamhankar. He received his ME degree in Electronics. He is Assistant Professor in department of Electronics Engineering at Walchand College of Engineering, Sangli. He has more than 20 years of teaching experience. Area-Computer Network, IOT, Image Processing