IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 8, Issue 1, Feb - March, 2020
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

Extraction and Reading of Line Word Character Segments Directly from Printed Text Document

Pooja S[1], T Akhila[2], Tejaswini J S[3], Tejaswini K R[4], Sharmila Chidaravalli[5]

[1]Dept of ISE,Global Academy Of Technology ,VTU,Bengaluru,India.

[2]Dept of ISE,Global Academy Of Technology ,VTU,Bengaluru,India.

[3]Dept of ISE,Global Academy Of Technology ,VTU,Bengaluru,India.

[4] Dept of ISE,Global Academy Of Technology ,VTU,Bengaluru,India.

[5] Dept of ISE,Global Academy Of Technology ,VTU,Bengaluru,India.

Abstract

The world estimates around 40 million people who suffer from visual disability in which India approximates about 15 million people. Statistically it shows that India has the largest population of people who are suffering from blindness. The proposed system interpolates an effective and worthwhile technique which allows the user to contents present in the image. This paper integrates the notion of Optical Character Recognition and Text to Speech Synthesis. This system permits visually impaired people to interact with computers via the vocal interface. The purpose of the proposed method is to embrace the system that extracts text data from the images by methods like global thresholding, horizontal splitting, and linear smoothing filter and finally, the extracted text will be converted to speech by the usage of Text to Speech synthesis analysis.

Keywords- Optical character recognition (OCR), Text-to-Speech (TTS).

I. INTRODUCTION

Speech is one of the most emphatic media for communication. Visually impaired people can make use of this effective medium which is the only way to know the information. Blindness is basically the biggest curb for human beings. People with blindness face many difficulties in day to day activities such as reading, walking, and socializing. But in today's world, we ourselves keep updated with the information in the same way we have to help them to know more updates about the current world. These can be achieved by our proposed system. Our proposed system will help them to listen instead of reading, which is more difficult for them. The main challenge is to read and know the information. Braille is one of the traditional and existing systems used to read with touch and feel the dots that represent the characters. The drawback of this braille system is that some of the people with blindness will not be able to learn the language. As many technologies that are emerging have put forward to help them but they face difficulty in alignment and in terms of cost which they face expensive to get the information. The human functions like machine reading were the ancient dream. But the current technology of machine reading has made it, in reality, using various technologies.

Optical Character Recognition is a tool that extracts the text under few constraints by considering only images with a smaller number of objects in background which deals with images that contain text as well as ones with objects in the background. It is used to extract the text under few constraints like considering only images with a smaller number of objects in the background. Here we deal with images that contain text as well as ones with objects in the background.

Text to speech synthesizer is the one that generates the speech. This module that consists of a button makes the system ready by setting to scan the image which helps the visually impaired to start the process. While it consists of a camera mounted on top to capture the required and give that as an input to the pi device which helps in the next step to make the image ready to read. The system takes an image as input, processes it and obtains required text from it. This text is later sent to spell check of Microsoft Word and then converted to speech. The above method can be more accurate for images that contain only text and can also give decent output for the images with objects in them.

II. RELATED WORK

Approaching the text which is printed or handwritten is a vital challenge for the person with blindness. The prefacing survey revealed that there are various difficulties with existing devices. We have depicted some existing devices and the drawbacks of it.

Finger Reader is a wearable device that explores the printed text. While accessing the text the pointer to the text would vary and try to recognize different sentences that are not usefully made to the visually impaired, which made a major drawback [1]. The system can read texts to measure video taken from ring-based hardware. Then our system will process images extracted from video and may Zone-based OCR to convert into computer understandable character which could be corrected using Frequent Pattern tree algorithm and would be set to text to speech software.

The proposed hardware can have a camera but also having buzzer and vibrator to offer physical stimuli to the blind users. But this system again failed while recognizing the text through sequential access which visually impaired would turn into a different line which is not the correct sentence to read the character [2]. The major disadvantage is that visually impaired cannot access the phone to recognize which they fail to make it [3]. An automatic document reader is developed for visually impaired people. In this RaspberryPi is used as the main computational tool. Image sensing sensors will scan the image and give it as input to the RaspberryPi. Raspberrypi makes use of the Tesseract library and python programming. The major disadvantage of this is not a portable device so it can restrict the blind [4].

III. EXISTING METHODOLOGY

According to old technique, the arrangement of text document to speech requires a record with a book on it, which is made compulsory to change over it to speech. There are different innovations that focus on change from the content and speech mostly on unit determination, area union. Different strategies additionally incorporate sine wave blend, Formanant, etc. Next, the front finish of the word reference combination strategy is applied to the English Text amalgamation which has been created with the blend and structure innovation so gave factually made speech union. Next, the strategy utilizes the tesseract library to change over the offered picture to be made to text and Text made to discourse blend for changing over to speech powerfully.

IV. PROPOSED METHODOLOGY

The proposed methodology is especially for the benefit of the blind people overcoming the disadvantages of the existing system. In our system, we have analyzed an efficient and useful approach to resolve the problem of visually impaired people. The blind community needs help and attention as even they have their routine life rather than being isolated. Speech technology. The first step in our proposed method is to capture images from the camera and the captured image is converted to text form using methods like global thresholding, horizontal stripping, and linear smoothing filters. The next stage is to convert the extracted text into speech using Text to speech (TTS) synthesizer.

A. Stages involved in our proposed methodology

1. A Captured image is taken as input to the system.
2. Pre-processing of a scanned image is performed.
3. Conversion of an image into text using the Otsu method, horizontal splitting, and a linear smoothing filter.
4. Increasing the accuracy of the extracted text.
5. At last, the text is converted to speech with the help of a text-to-speech (TTS) synthesizer.

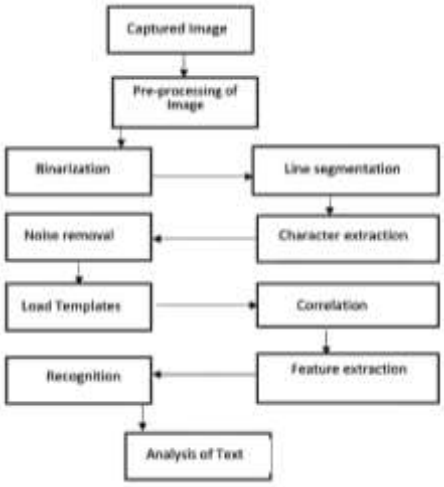

Fig. 1. System Architecture of proposed methodology.

B. Extracting text from the image

Extracting text from the scanned documents using methods like global thresholding, horizontal splitting, and linear smoothing filters is now a popular solution for some text documents. The components of the proposed method are shown in fig.2.

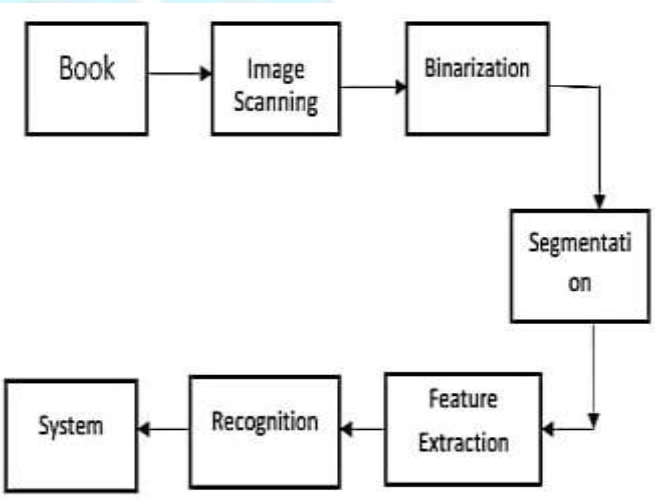

Fig. 2. Text Synthesis

Pre-processing of extracted text is done. After pre-processing, each symbol is extracted from the image. From the previous learning stage, each symbol is compared with extracted features with descriptions of the symbol classes. With the help of a spelling checker, the reconstruction of words and numbers is performed.

C. Speech synthesis

Speech synthesis is the contrived rendering of natural speech. It is the one that is used to translate text into audio information. Mainly in voice-enabled services it plays an important role. Synthesized speech is created by grouping pieces of recorded information of speech which is present in the database.

D. Text to speech conversion

Text to Speech conversion can be done by the usage of Text-to-Speech (TTS) which is read aloud technique. It is capable of reading the text characters. It takes the words on digital device. Finally, it converts it into audio.

It consists of the front end which assigns phonetic transcriptions to word and the back end that converts symbolic representation to audio. words are named text-to-phoneme conversion. The backend often mentioned as the synthesizer—then converts the symbolic linguistic representation into sound.

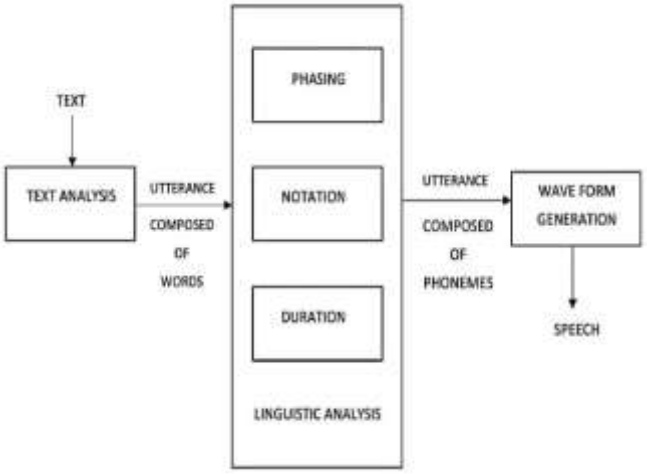

Fig. 3. Text-To-Speech Synthesis

V. RESULT

A. Converting RGB image to grayscale image

Red, blue and green color are put in a separate matrix and represented by 8bit to 32bit. Grayscale Image is a combination of black and white colors. It eliminates hue and saturation information while retaining the luminance. By forming a weighted sum of the R, G, and B components, it converts the RGB values to grayscale values using, 0.2989 * R + 0.5870 * G +0.1140 * B

Fig. 4(a). Input image

Fig. 4(b). Conversion of RGB to gray scale

B. Adaptive Threshold

Adaptive thresholding is a thresholding method that takes into account spatial variations in illumination. In out proposed method, we present a way for real-time adaptive thresholding using the integral image of the input. It separates the foreground from the background with nonuniform illumination. Adaptive threshold is the method of choosing the threshold value to minimize the intra class variance of the black and white pixels.
Gray thresh – Global image threshold using Otsu method. Using LEVEL = graythresh(I), the global threshold value is found to convert an intensity image to a binary image with IM2BW. LEVEL is a normalized intensity value which lies in between the range [0, 1]. im2bw – which is used to convert image to binary image by thresholding. im2bw outputs binary images from indexed, intensity, or RGB images. To perform this, it converts the input image to grayscale format (if it is not already an intensity image), and then converts this grayscale image to binary by thresholding. The resultant binary image BW has values of 1 (white) for all pixels in the input image with luminance greater than LEVEL and 0 (black) for all other pixels.

Fig. 5. Adaptive thresholding

C. Noise Removal

Noise in the image originates from within the camera which has a few main causes. The three main causes of noise in the image are electricity, heat, and sensor illumination levels. In low-light situations, each pixel has very little light wave fluctuation to report before being amplified. To remove noise present in the image, linear smoothing filters are used. Linear smoothing filters is a method of recovering image by convolving the actual image with a mask that represents the low-pass filter or smoothing operation.

Fig. 6. Noise removal (object removal)

D. Line Extraction

Extraction of lines can be made by horizontal Splitting is applied to the pure black and white image which we got after the above two processes. Start by splitting the above obtained image into horizontal segments. These segments are then divided into horizontal rows containing no trace of white (i.e., the row which is completely black). In this way all the lines are obtained.

Fig. 7. Line segmentation

E. Character Extraction

Useful text from horizontal strip are obtained by applying the above method. Character Extraction can be done by further splitting it vertically to separate characters by means of similar technique which was used while splitting horizontal lines. The characters are divided on basis of vertical rows which has no trace of white. It also keeps track of the average space between each character. It is later used to detect separation between two words.

F. Separating two closely spaced characters (extra space)

When we must separate two closely spaced characters it can lead to such a case that can be found where it is difficult to extract characters from the given above approach. In that case, where there are no separable column or row, it uses the concept of connected components. The set bits that belong to the first character are first identified and will be used as source nodes for the algorithm. The indices of the set bits with the first character are added to a queue. Then the front-most set bit present in the queue is taken and all neighbors to set bit is added in the queue. The process is repeated until there are no unvisited set bits of the first character. The first character has been completely spanned by now and that can be stored in a separate matrix. To obtain only the first character, it subtracts the above-obtained matrix from the original image.

G. Template Matching (dataset)

Over 6500+ images that can contain both numerals and alphabets were collected from MNIST. It contains about 1071 samples of 65 characters. This dataset is pre-processed and it is already saved in a template (.mat file form). This template is loaded while running the code. Then, each character obtained after during Character Extraction method undergoes comparison with dataset using r = corr2(A, B) returns the correlation coefficient between A and B, where A and B are matrices or vectors of the same size. r is a scalar double. For further optimizations, if sure that the character we are looking for, is the most probably known, then we break then and there to avoid further unnecessary comparisons in our runtime. To space them properly we calculate the average space between words by finding the number of columns that are completely black and then compare them with our pre-defined value.

H. Spell Check

Finally, when character extraction is completed, the sentences formed by the words are obtained, there may be a chance where the words contain some other character because the accuracy got till now is only 95%. To increase the accuracy level, each word is sent to the spell check function. The Spell Check function is not an inbuilt function. However, using ActiveX, the words are passed to the spell checker of Microsoft Word.

I. Conversion of text to speech

The text to speech conversion function TTS(TXT) synthesizes the speech from the obtained string and spells it out. The audio output is mono, 16bits, and 16k Hz by default. The text to speech conversion function synthesizes the speech from the obtained string and spells it out. The audio output is mono, 16bits, and 16k Hz by default.

Input image:

Output image:

d
e
p
a
r
t
m
e
n
t
o
f
i
s
e
department of ise

VI. CONCLUSION

A technique is provided for the issue of vision for visually impaired in this paper. The proposed model is a multipurpose model that is designed for a reading purpose which can carry out the reading operations such as reading books, newspapers, magazines, resulting in a reduction of human intervention. The Proposed method works in two fields. Firstly, the conversion of an input image to text format and secondly converting the given text format as an input to speech using Text to speech synthesizer (TTS).

By this proposed methodology, it is able to read the text from a given document, Web page series, or even with e-Book which is available and can generate speech through some external speakers. This is made to save time than reading a book manually.

The proposed methodology made to show an accuracy that is about 90-92%. The accuracy was totally obtained from the spelling check method. Hence the system which is used to read the sentences can be more helpful for the people with vision disabilities, and the person who needs the image to be read can place an image and would be able to hear the contents of their choice, Children's who would like to hear their choice of rhymes, Magazines, books and so on can be made with the help of the proposed method.

VII. FUTURE ENHANCEMENT

With the helping hand for visually impaired with the proposed technology it can be enhanced with an automatic page-turner which would help them to further flip the page without them turning the page. It is also enhanced with the CNN algorithm to classify with each character using a dataset.

VIII. REFERENCES

[1] Sunitha S, Afroz B, G Tejaswini : *A Text Recognizing Device for Visually Impaired*, International Journal of Engineering and Computer Science (IJECS); ISSN:2319-7242, March 2018.

[2] Nuno Fonsecaa, RobertoNeto: *Camera Reading For Blind People* International Conference on Health and Social Care Information Systems and TechnologiesTechnology of Procedia 1200 – 1209, 2014.

[3] Supriya Khatavkar , AkashSehrawat, Anush Goel, Ankush Patil, Prashant Chougule: *Raspberry Pi Based Reader for Blind People* International Research Journal of Engineering and Technology (IRJET)Issue 6, Volume 5, ISSN: 2395-0056, June 2018)

[4] Ashish Jadhav, Urmila Shinde, Bhagyashri Sawal, Ashanta waghmar, Mohini Pangare: *Raspberry Pi Based Reader for Blind People* International Research Journal of Engineering and Technology (IRJET), Issue 4, Volume 6, ISSN: 2395-0056, April 2019)

[5] Jayaprakash.V, Jothimani.A, Joe Benhur Xavier.F Shirly Edward.A: *Text-To-Speech Device for Visually Impaired People;* International Journal of Pure and Applied Mathematics (IJPAM) No. 15, 1061-1067, ISSN: 1314-3395, Volume 119, 2018.

[6] Ms.Anupama Pandey, Ms.Vrunal Patil, Ms.Athira Panicker : *Smart Shopping Assistant Label Reading System with Voice Output for Blind using Raspberry Pi and its use;* International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Issue 10, Volume 5 ISSN: 2278 – 1323, October 2016.

[7] S.K.Katti , M.A.Anusuya,: *Speech Recognition by Machine: A Review* International Journal of Computer Science and Information Security (IJCSIS); No. 3, 2009 Vol. 6.

[8] J.Rejina Parvin, Sanjana : *Voice Assisted Text Reading System for Visually Impaired Persons Using TTS Method;* IOSR Journal of VLSI and Signal Processing (IOSR-JVSP), Ver. III, Volume 6, Issue 3, ISSN: 2319 – 4200, May-Jun 2016.

[9] Narendiran K, Joshuva J, Dr. B. Muthusenthil, Kishore S: *Smart Assistance for Blind People using Raspberry Pi;* International Journal of Advance Research, Ideas and Innovations in Technology; Issue 2Volume 4 ISSN: 2454-132.

[10] Ramesh Babu and Vasanthi: *Vision Based Assistive System for Label Detection with Voice Output;* International Conference on Engineering Technology and Science (ICETS - 14), Special Issue 1, Volume 3, 2014 January)

[11] Stavan Kodolikar, Dhananjay Laghate, Amarsinh Khavale, Prachi Rajarapollu: *FPGA Based Braille to Text & Speech For Blind Persons;* International Journal of Scientific & Engineering Research, , 348 ISSN 2229-5518, Volume 4, 2013, Issue 4.

[12] Jochen Huber, Meng Ee Wong, Roy Shilkrot, Pattie Maes, Suranga Nanayakkara: *A Wearable Device to Explore Printed Text on the Go*, Conference on Human Factors in Computing Systems April 23 2015 , CHI Crossings.

[13] Prof. Priyank Shah, Rutvi Prajapat: *Design and Testing Algorithm for Real Time Text Images: Rehabilitation Aid for Blind* IJSTE: International Journal of Science Technology & Engineering, Issue 11 Volume 2, 2016, ISSN 2349784X).

Divyalakshmi m k , daniel dinesh , vijaya kumar A sandhiya r, dharmarajan g: *an advent approach of tactile based entry with speak support for visually impaired:* International Journal of Advances in Science Engineering and Technology, Issue-3, Volume- 2, , ISSN- 2321-9009, 2014.