International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 5, May 2020||

Real-Time Assistive Reading for Visually Impaired

Apoorva Kadam[1], Kaivalya Banot[2], Manali Shah[3], Shaunak Umarkhedi[4], Prof.M. R. Apsangi[5]

U.G. Student, Department of Information Technology, PVG's College of Engineering and Technology, Pune, India[1,2,3,4]

Assistant Professor, Department of Information Technology, PVG's College of Engineering and Technology, Pune, India[5]

ABSTRACT: According to the World Health Organization (WHO), 285 million people are visually impaired and 39 million are blind. Keeping these statistics as our motivation we surveyed at "Poona Blind School for Girls" to get a better understanding of the difficulties they face in their day to day lives. Their prominent concern was that they were not able to perform to their full potential because of their disabilities. Students believed that if they would be able to read like a normal child, they would be able to do justice to their talent. In this paper, we have tried to propose a solution that will help visually impaired students to read books and comprehend them easily without being dependent on anyone. We propose a software which would help them to read texts. With the help of this software, we capture images of the book in real-time and read content out loud to the user. The algorithm, used in this software, consists of four phases. The 1st phase is responsible for a calibration procedure which is done to capture the best quality image, 2nd phase preprocesses the input, the 3rd phase converts the image to text with the help of tesseract library, finally, the 4th phase converts the text into speech and provides output to the user in audio format.

KEYWORDS: visually impaired; image processing, text localization; text detection; blur detection; image thresholding; contours; OCR; text to speech;

I. INTRODUCTION

Advancement of technology has led us through times where we can extract the information through the images. The term where we detect text from images is known as text recognition. Through this paper, we are trying to address the issues of the visually impaired community when it comes to reading.

We surveyed "Poona Blind School for Girls" to get a better grasp of the difficulties which they face in their day to day lives. We asked them questions about their difficulties and what they expected from the software which would alleviate their problems. The questionnaire we prepared had questions regarding the existing softwares in the market and how comfortable are those for everyday usage. The students emphasized on the fact that they're uncomfortable with technologies that provide abrupt haptic feedback as it startles them. Inspired from all of these facts we designed software that would help them to read books.

The designed software had four phases- the 1st phase is responsible for a calibration procedure which is done to capture the best quality image, 2nd phase preprocesses the input, the 3rd phase converts the image to text with the help of tesseract library, finally, the 4th phase converts the text into speech and provides output to the user in audio format. This paper will explain all the steps in detail with a proper explanation of the algorithms and methods used.

II. RELATED WORK

To address the issues of visually impaired, many technologies have leveraged text recognition to provide a powerful feature. In the forthcoming section, we present a detailed survey of various such approaches. Some existing devices in the market are- finger reader and Orcam.

The FingerReader [6] is a ring-like device with a small camera that scans and reads the words out loud in real-time, as users follow the text with their finger. The disadvantage of this device is that the blind person has to keep track of the lines while reading. This product can be ideal for a partially blind person but won't work for a completely blind person.

OrCam [5] devices such as OrCam MyEye are portable, artificial vision devices that allow visually impaired people to understand the text and identify objects through audio feedback, describing what they are unable to see. The price of this device is around 3 lakhs which is the device's major disadvantage when it comes to Indian scenarios.

III. PROPOSED ALGORITHM

The software captures the image and checks for the quality of the image. If the image passes the required quality threshold then it is passed on to other modules. The process of converting an image into text happens in four phases (4

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 5, May 2020||

step pipeline)- the 1st phase is responsible for a calibration procedure which is done to capture the image with the best quality, 2nd phase preprocesses the input, the 3rd phase converts the image to text with the help of tesseract library, finally, the 4th phase converts the text into speech and provides output to the user.

A. IMAGE QUALITY CHECK

Phase 1 detects the quality of the image. Every captured image is checked whether it is a blur or not. If the blur value derived for the image is less than the threshold value, then only the image is passed on to further modules. This module uses the Laplace operator to detect the blur value. The image is first converted into grayscale. The output of this module is stored in image format and is utilized by other modules. If the image doesn't qualify as per the threshold, the user is then prompted to capture another picture.

To detect blur, a single channel of an image (grayscale) is taken and is convolved with the following *3 x 3* kernel:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Pseudo Code:

1) Take an image as an input and convert it into grayscale, pass it to a function for calculating blur.
2) The function calculates the Laplacian of an image and returns the focus measure by convolving the 3 x 3 Laplacian operator.
3) The image is considered as "blurry" if the focus measure is less than the fixed threshold.

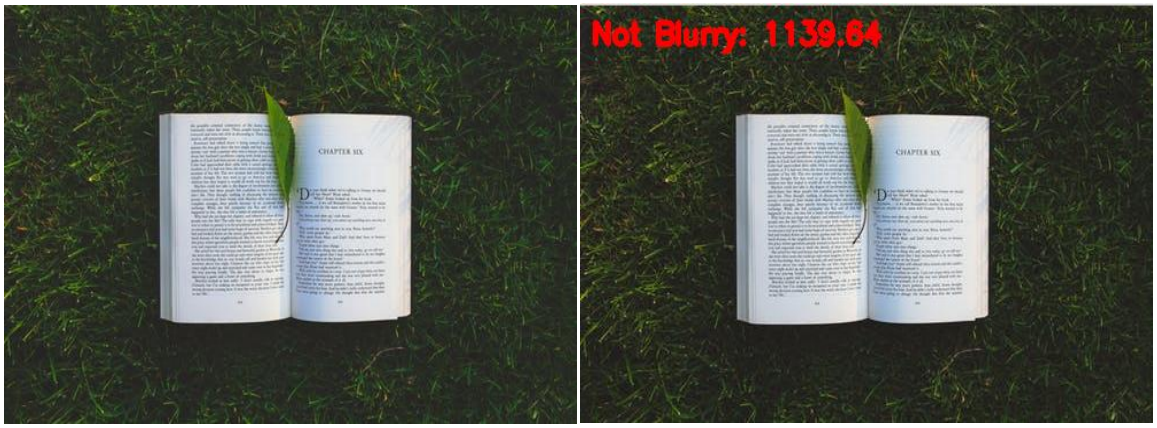

Figure.III.A.1 Non Blurry Image Input, Non Blurry Image Output

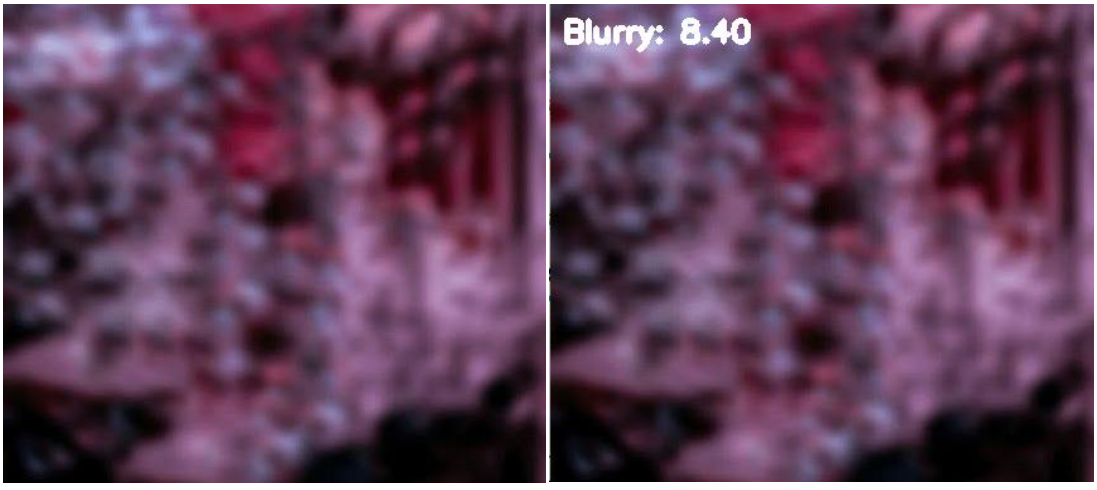

Figure.III.A.2 Blurry Image Input, Blurry Image Output

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 5, May 2020||

B. DOCUMENT LOCALISATION

In this phase, the area where the book is located is extracted from the image. This is achieved by document extraction followed by document localization.
Document Extraction - We used binary thresholding for document extraction which is essentially converting an image to black and white, we kept the thresholding coefficient at 150. Usually, written documents are in a shade of white or yellow, so binary thresholding gives us a complete region of the document in white colour while rest other regions in black colour.
Document Localisation - Finding contours in an image finds closed polynomial shapes in the image. It generates a list of coordinates of the recognized shape in the image.

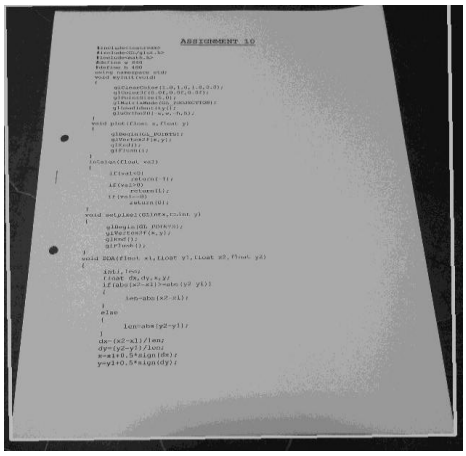

Figure.III.B.1 Output of contour finding algorithm, A rectangle drawn localising the document can be seen

The problem of a bounded book - The problem with a bound book is if we take a photo of it, there are two pages in sight separated by the spine(the middle part where pages are attached) of the book. In image processing, there is a persistent shadow in that middle spine region.

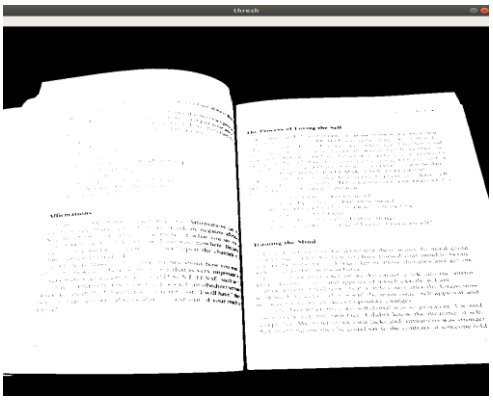

Figure.III.B.2 Thresholded image of a book,dark pixel intensity is visible in the middle part.

Using the above mentioned method we cropped the image in two separate parts for the left side page and the right side page.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 5, May 2020||

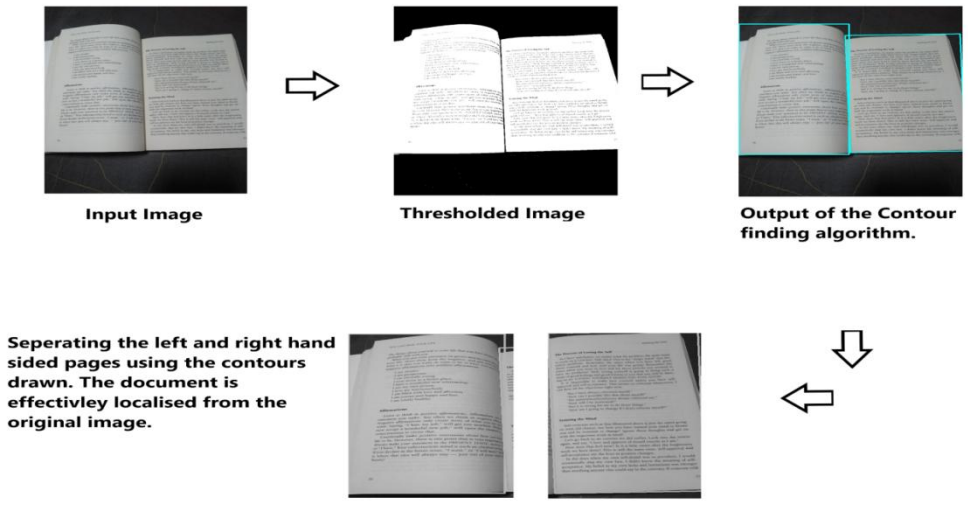

Figure.III.B.3 A pictorial representation of the entire Document extraction-localisation process.

C. TEXT LOCALIZATION AND RECOGNITION

The images processed so far are sent to an open-source OCR engine[1] and the recognized text is converted to speech for giving audio feedback. This step is divided into two prominent parts:

1. Text detection and recognition[2]
2. Converting text to speech. In the first part, i.e text detection and recognition[3,4].

We have used tesseract for implementing OCR applications. Tesseract [7] is an open-source text recognition (OCR) Engine, It works using LSTM technology. For this project, we have implemented tesseract with the help of python and OpenCV which gives a better hand on covering test conditions. Currently, the software can detect the English language. In the second part, i.e. converting text into speech, we have used **pyttsx3** . It is a Python library for text-to-speech conversion. It works offline and is compatible with Python 2 & 3, unlike alternative libraries. The application invokes the pyttsx3.init() factory function to get a reference to a pyttsx3. All the processed results are passed through this module to get the speech output that the user can listen to.

IV. SIMULATION RESULTS

A. *DOCUMENT EXTRACTION AND LOCALISATION*

IV.A.1 A thresholding coefficient of 150 was found suitable when dealing with document color ranging between white to beige and normal external lighting of a common household tubelight.

IV.A.2 The document area coefficient for contouring of 50000 was found suitable for documents covering the majority of the image.

Document type	Color	Thresholding Coefficient	Outcome
1. Book	White	150	Correct identification of left and right side pages.
2. Book	Yellow	100	Correct identification of left and right side pages.
3. Single page document	White	150	Correct identification of the document.

Table.IV.A.1 Results of Document extraction-localisation algorithm

IV.A.3 The above tests were carried out at a camera-to-document distance of 25 cm.
IV.A.4 The above module can be tuned to read documents of any color, shape and font.

B. *TEXT RECOGNITION AND TEXT-TO-SPEECH:*

Type of Document	Total words	Correctly Recognized Words	%	Correct Text-to-Speech	%
Single page Document With Special characters (e.g. !$%?)	135	119	88.14	116	85.92
Single page Document (only text)	155	140	90.32	135	87.09
Book (Both pages)	293	272	92.83	266	90.78
Average	583	531	**91.08**	517	**88.67**

Table.IV.B.1 Results of Text Recognition,Text-to-Speech Algorithm

IV.B.1 After running the algorithm on multiple pages having different font-sizes, background color, illumination and shape of the document following results were found
IV.B.1.1 Average Text recognition accuracy - 91.08 %
IV.B.1.2 Average Text-to-Speech accuracy - 88.67 %

V. CONCLUSION AND FUTURE WORK

In this paper, we discussed the development of the software with the help of which visually impaired people can read text present on the documents. The algorithm used to detect texts firstly checks for the quality of the image. If it passes a quality check then it converts the text into speech and provides output to the user.

However, there are a few remarkable modifications that can be used to extend this product:

1. Reading of traffic signs
2. Reading the boards of public transport
3. Reading of complex languages
4. Reading of mathematical equations
5. Describing a picture/diagram present in the book

REFERENCES

1. Ramkishor V and Mr.Rajesh L, “Artificial Vision for Blind Peoples Using OCR Technology”, 2017 International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 6, Issue 3.
2. P. M. Manwatkar and S. H. Yadav, "Text recognition from images," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2015, pp. 1-6.
3. Boiangiu, Costin-Anton & Ioanitescu, Radu & Dragomir, Razvan-Costin. (2016). VOTING-BASED OCR SYSTEM. Journal of Information Systems & Operations Management. 10. 470-486.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

||Volume 8, Issue 5, May 2020||

4. Hamad, Karez & Kaya, Mehmet. (2016). A Detailed Analysis of Optical Character Recognition Technology. International Journal of Applied Mathematics, Electronics and Computers. 4. 244-244. 10.18100/ijamec.270374.
5. Orcam website: https://www.orcam.com/en/
6. Shilkrot, Roy & Huber, Jochen & Wong, Meng Ee & Maes, Pattie & Nanayakkara, Suranga. (2015). FingerReader: A Wearable Device to Explore Printed Text on the Go. 2363-2372. 10.1145/2702123.2702421.
7. R. Smith, "An Overview of the Tesseract OCR Engine," Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Parana, 2007, pp. 629-633, doi: 10.1109/ICDAR.2007.4376991.