International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 3 (2015) pp. 8321-8327

http://www.ripublication.com

Text Reading Machine for the Visually Impared using OCR and TTS

Dr .Josephine Leela and Dr.T.Kalaichelvi

*Associate Professor, *Department of Computer Science and Engineering*
Panimalar Engineering College, Chennai, India
Asscoiate Prof, CSE, Panimalar Engineering College
4ilanleela@yahoo.com

Abstract

Optical Character Recognition (OCR) is a process that converts image to text. In this paper, text is scanned from a book and is processed by the OCR system to extract and display the characters in the text .The displayed text would be difficult for the visually impaired people to perceive it .To overcome this problem TTS system can be used .Text-To-Speech (TTS) is a process that converts the text in the computer to an audible format. The proposed system aids their reading by integrating the functionality of OCR and TTS.

Keywords – Image processing, Optical character recognition, Prosody analysis, Text to speech synthesis ,Visual Impairment.

INTRODUCTION

Visually impaired people are deprived of the sense of vision either partially or completely. In this era people are blessed with books both in printed and electronic format .Books in electronic format can be read with the help of certain software assistive tools like screen readers but what if they want to read books in printed format .They can either seek the help of a volunteer or the books need to be printed in Braille formats to be able to be accessed by the blind .This requirement hinders the visually impaired to read whatever they wish and whenever they want.

The rapid growing technology makes the solution for the above specified problem viable and also visually impaired people have a better hearing and tactile sense that can be used to bridge the gap between reading and printed documents .A seemingly obvious solution to make reading possible with the use of tactile sense is

printing the electronic books and documents as pages of raised dots using Braille printers and embossers but the cost that is inherent in acquiring the resources as well the high quality papers to withstand the inked dots adds up to the total cost .

The paper proposes a software that aids the visually impaired in reading printed textual documents with extended user interface design that assists the visually impaired to easily access and use the reading system with voice control mechanism and the use of keyboard accelerator keys .The proposed system is also economical and has a better cost benefit ratio. Detailed explanation of the work is described in the forthcoming passages section II specifies related work and section III contain project work. Section IV contain result and description and section V contain conclusion

.

RELATED WORK

The survey was carried with the intent to gain knowledge on methods and techniques that are related to the project in order to analyze and compare the techniques that are feasible and to further optimize them to produce better results. Various Papers were collected from speech processing, image processing, human computer interface, text to speech and Optical Character Recognition domains during this study period. The observations from the study are as follows

AimiliosChalamandaris .et al., has proposed the paper [1] to describe about unit-selection, a text-to-speech technology for producing near-natural speech synthesis systems. The overall quality of the synthetic speech achieved by such systems are quite high, however they do not guarantee a high level of user satisfaction. It seems many issues have to be coped with in order to fulfill user expectations when integrating such systems with screen reading tools aiming to assist blind users. The paper describes the design and implementation approaches for the efficient integration of this technology into screen reading environments. In specific, the issues related to natural language processing, multilingual design and overall quality optimization on a whole are addressed in this paper. [2] Presents one approach for OCR named "i" that aims at a high speed, simple, font independent and size independent OCR system based on an unique segment extraction technique. The algorithm is used as a kernel for single alphabet detection in a complete OCR system without the need for complex mathematical operations. The paper differentiates form [1] as in it does not use any libraries or databases of image matrices to recognize alphabets, but it has a unique algorithm to recognize alphabets.

[3] Proposes a standalone Arabic Reading Machine using TTS (Text-to-speech) and OCR (Optical Character Recognition) model built in a user friendly way for Visually Impaired People. The author presents an Arabic standalone reading machine designed specifically for the visually impaired people with convenient control. Also, the author analyzes the system of existing Latin reading machines in detail, along with existing OCR and TTS software. The paper [4] describes Optical character recognition (OCR) in aspects of pattern recognition, machine vision and artificial intelligence which are used in developing algorithms for reading text on images, e.g. reading registration plates, scanned books and documents, etc. This paper presents the system for text extraction on the image taken by grabbing the content of

the TV screen. The preparation steps for OCR are developed which detects the text regions in an image. An open-source algorithm for OCR is then run to read the text regions.

The author presents a novel device developed as an aiding tool in [5] for people having difficulties to read printed text documents. The system enables the user to randomly access the words of a printed text by directing a handheld pointer on the word of intent. The device seems to be very effective and friendly to use. A micro camera is fixed at the tip of the pointer and is used to point the word the user wants to read. A voice synthesizer reproduces the word just framed. A suitable measurement of the device motion is based on real time image processing that determines when the user has stopped pointing and indicates the wish to hear the pointed word(s).

[6] Describes a system dedicated to visually impaired people for reading aloud text that makes use of standard optical character recognition (OCR) component. However the inherent difficulty is the OCR component cannot analyze directly photographs taken out of doors because of their complexity. Before starting the OCR the photos are sent for preprocessing that relies on image segmentation and selection of letter-like segments. Its bottleneck is evaluation of image segment similar to letters and digits. This paper also describes different methods and results of their comparison.

[7] The paper presents a novel method using a self-organizing map neural network which can be used for character recognition tasks. The text to be recognized was very much limited to characters typed using the training to recognize 26 handwritten characters in the Verdana font type, bolded with a font size of 18. The program is capable of handling non-ideal images that are even noisy, colored text, rotated image that overcome the problem faced in [6] with a better perspective . Recognition accuracy is almost 100% for ideal images, but however it ranges between 80% - 100% for non-ideal images.

In [8] the author presents two different artificial neural network approaches for phoneme recognition and for text-to-speech applications using Staged Back propagation Neural Networks and Self-Organizing. Applying neural networks for phoneme mapping for text-to-speech conversion creates a fast distributed recognition engine as in [7]. This engine not only supports the mapping of missing words onto the database, but also it mitigates the variations in pronunciations for the same word.[10] presents an artificial neural network approach for Optical Character Recognition System and a simple pattern recognition system is designed and simulated using ANN similar ideology as proposed in [7] [8]. Training the artificial neural network on different sets of noisy data forced the author to learn how to deal with noise.

[11] The author describes an overview of the document processing reader for OS/2 machine. The discussion mainly is centered with the user interface of the system. Since GUI's present serious problem for blind users instead of icons and mouse the system interface is built with voice menus and numeric keypad. This is issue is considered in the model we propose and an alternative approach is described in the forthcoming section.

Proposed work

The proposed system has a front end OCR module and back end TTS system. OCR a branch of Image processing is particularly preferred model for extracting the contents of textual images which is extensively used in most reading machines. We have made use of tesseract OCR engine for character recognition .The input to the OCR module is the photo copy of the book or text document to be read .web camera is used to capture the image of the text document and is sent to the OCR module. Some of the techniques involved in OCR are Pre-Processing .and Character recognition.

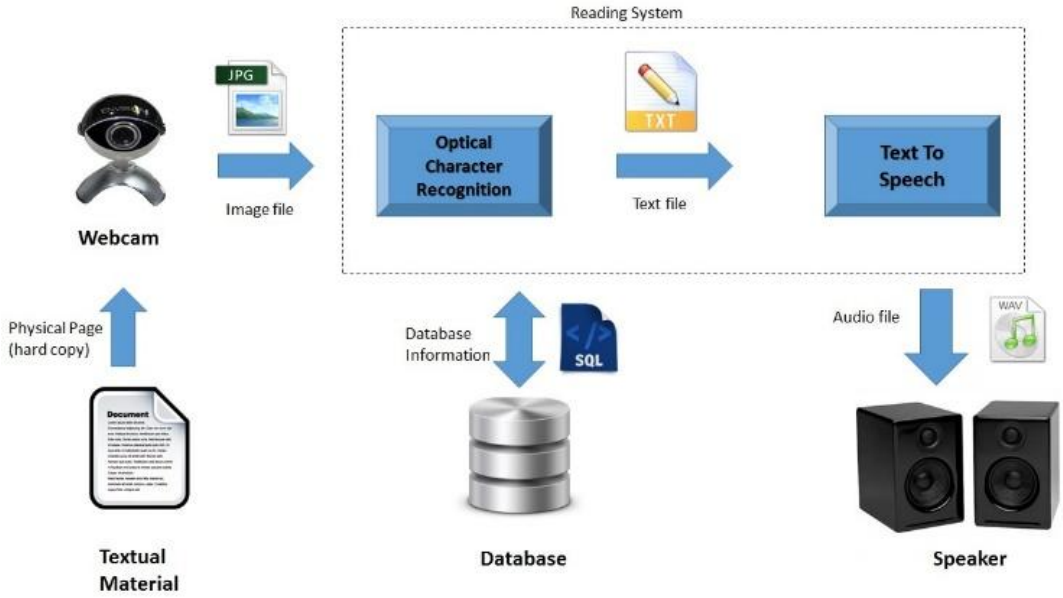

Figure 1 : Architectural Diagram

A) Preprocessing

Preprocessing the image often improves the quality of the image by enhancing the parameters related to the image like brightness, contrast and sharpness .Initially the document is de-skewed to adjust the orientation of the scanned document to make the lines of text to fit either horizontally or vertically. Then de speckle is carried out to remove the unwanted spots or dots that are either present in the paper on which the text is printed or might have been ejected by the printer while printing the document. In order to improve the efficiency or accuracy in recognizing the characters the image is converted from a color or gray scale to a black and white image. Then the line removal is applied to remove the unwanted glyph boxes and lines. Then zoning identifies columns, paragraphs as distinct blocks when the word detection establishes the base lines for words and identifies the word spacing. For advanced fonts or fonts of a relatively bigger size the spacing between the letters might be greater than that of the spacing between the words in which case the algorithms needs to be more sophisticated . The tesseract OCR algorithm uses an inbuilt dictionary in order to facilitate character recognition.

B) Character Recognition

After the preprocessing step is over, character recognition is carried out using matrix matching in the first phase which compares an image to a stored glyph on a pixel-by-pixel basis like pattern matching .The output relies on the input glyph being secluded perfectly from the rest of the image, and the stored glyph in a similar font and at the same scale. However this technique works well with typewritten standard text and is not satisfactory when new fonts are encountered. During the second phase the unrecognized characters are processed using feature extraction technique that decomposes and classifies each characters into features like closed loops, open loops, line intersection. Then the nearest neighbor classifier such as the k-nearest neighbors algorithm are used to compare image features with stored glyph features and choose the nearest match.

The output of the OCR module is a stream of plain text or a text file .But to make the system deployable to the visually impaired it is not enough to just recognize the printed text rather to deliver it in a format that can be perceived by them. Focusing on the posterior part of the Reading system that converts the text to audio format we get into speech synthesis a sub domain of digital signal processing area.

C) TTS system

A TTS system is used to convert normal language text into speech .An intelligible text-to-speech program allows people with visual impairments or reading disabilities to listen to written or printed works on a home computer. A text-to-speech system is composed of two parts. Initially it converts raw text containing symbols like numbers and abbreviations into its equivalent written words this process is called text normalization. It also assigns phonetic transcriptions to each word, divides them and marks the text into prosody units. This process of assigning phonetic transcriptions to words is called grapheme-phoneme conversion. Front end outputs the linguistic representation which contains phonetic transcriptions and prosody information. The posterior processing is often referred to as the synthesizer which converts the symbolic linguistic representations into sound. In some this part also includes the computation of the target prosody, which is then imposed on the output speech

.

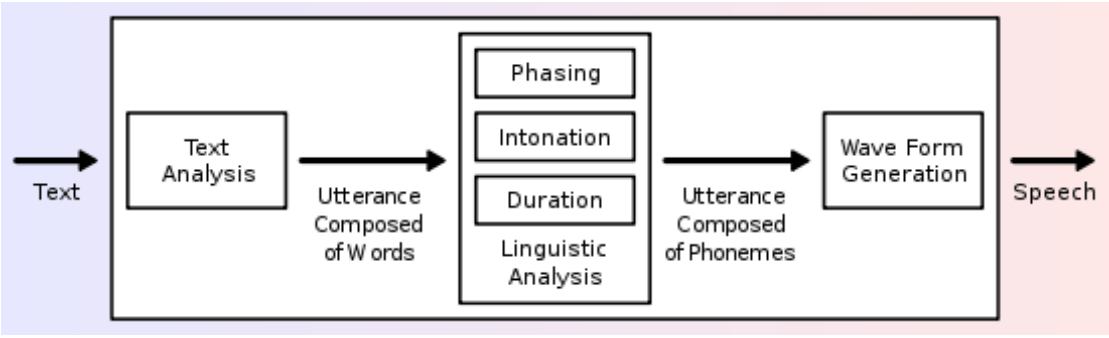

Figure 2: TTS System

D) Simplified User Interface

Since the system is mainly modelled to meet the needs of the visually impaired users the common problems faced by the blind in using the GUI's should be resolved. Thus we have made use of keyboard accelerator keys and voice control mechanism to control the system. Keyboard accelerators are keystroke combinations which sends commands to the application window to perform the required operations. Voice control mechanism can also be used to command the system by which the user finds a comfort in using the system and feels great control in utilizing the system to meet their needs thoroughly.{include a table specifying key combination and the performed function}

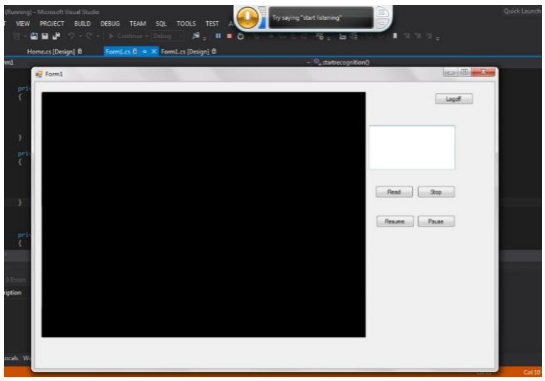

Figure 3: System waiting for reading operation

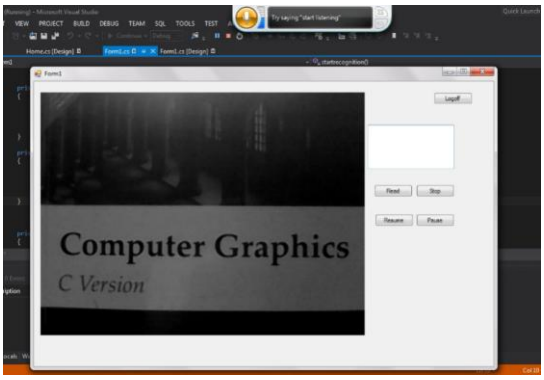

Figure 4 : User displays printed material to read

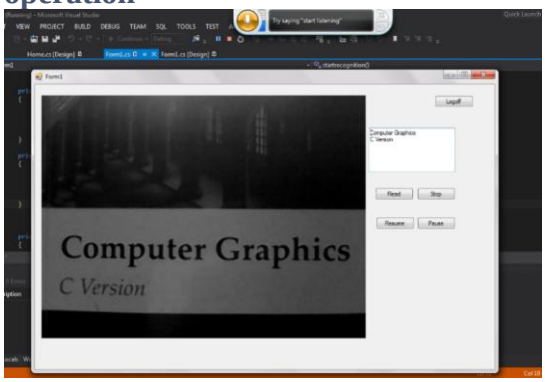

Figure 5 : The Image from the webcam is processed into text and converted into audio which is sent to the speaker.

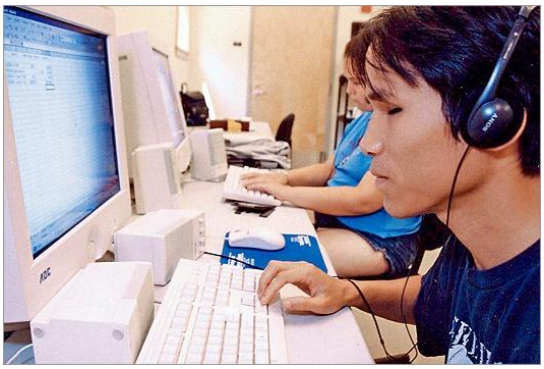

Figure 6 : Visually Impaired person listens to converted audio

References

[1] AimiliosChalamandaris, " Unit Selection of text to speech system", vol.56, IEEE Transaction 2010.

[2] Yi , C.Tian , " Portable Camera Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind People", IEEE Transaction 2013.

[3] I. Bouazizi, F.Bouriss, "Arabic Reading Machine for Visually Impaired People using TTS and OCR" ,IEEE Transaction 2013.

[4] Shastry.S, Gunasheela,G ,""i" A novel algorithm for optical character recognition (OCR)",IEEE Transaction 2013.

[5] Kastelan, I, Kukolj. S ," Extraction of Text on TV Screen using Optical Character Recognition", IEEE Transaction 2012.

[6] Minoni. U , Bianchi. M., "Trebeschi, V. A handheld real-time text reader" , IEEE Transaction 2011.

[7] JacekLebiedź , " Letters Detection in the Photograph", IEEE Transaction 2008.

[8] Ganapathy,V, "Optical Character Recognition Program for Images of Printed Text using a Neural Network", IEEE Transaction 2006.

[9] Embrechts,M.J. ; Arciniegas, F., "Neural networks for text-to-speech phoneme recognition", IEEE Transaction 2000.

[10] Fabio violaro and Olivier Boeffard., "A hybrid model for text to speech synthesis" , IEEE Transaction 1998.

[11] Mani Nallasamy, Paul Voumard., "An optical character recognition using artificial neural network", IEEE Transaction 1996.

[12] Portele,T,Kramer,J., "Adapting a TTS system to a reading machine for the blind" , IEEE Transaction 1996.

[13] Chieko Asakawa.,"Userinterface of document processing for the blinds", IEEE Transaction 1996.

[14] Jiang Liu, Chung-Mong Lee andRen-Ben Shu , " An Efficient Method for the skew normalization of a Document Image ", IEEE Transaction 1992.

[15] Michael H. O'Malley.,"Text To Speech Conversion Technology ", IEEE Transaction 1990.

[16] Chieko Asakawa.," User Interface of a Document-Processing Reader for the Blind ", IEEE Transaction 1996.

[17] *Mark Tathamand Eric*Lewis., "Improving TEXT-TO- SPEECH Synthesis" ,IEEE Transaction 1991.

[18] L. Leija, **S.** Santiago , "A system of text reading and translation to voice for blind persons" , IEEE Transaction 1996.

[19] Kazuhide Sugawara," Document Reading System for the Visually Disabled ",IEEE Transaction 1999.

[20] *C. Garcia and X. Apostolidis ,"*Text detection and segmentation in complex color images" , IEEE Transaction 2000.

[21] IvanKastelan, Sandra Kukolj, "Extraction of text on TV screen using optical character recognition", IEEE Transaction 2012.