Printed Text to Audio Converter using OCR

Kalyani Mangale
Department of Electronics and Telecommunication Engineering, Sinhgad Institute of Technology and Science, Narhe, Pune

Hemangi Mhaske
Department of Electronics and Telecommunication Engineering, Sinhgad Institute of Technology and Science, Narhe, Pune

Priyanka Wankhade
Department of Electronics and Telecommunication Engineering, Sinhgad Institute of Technology and Science, Narhe, Pune

Vivek Niwane
Department of Electronics and Telecommunication Engineering, Sinhgad Institute of Technology and Science, Narhe, Pune

ABSTRACT
For many blind users educational choices are made based on which material can be accessed and which cannot. These people are dependent solely on Braille books & audio recordings provided by NGOs.The presented work will provide visually impaired people, an opportunity to have an audio material of their own choice of any printed material. The framework consists of two parts. One is Optical Character Recognition (OCR) which includes operations like grayscaling, thresholding, filtering, thinning, segmentation, cropping, etc. on a character in the image and other part is text to speech conversion using Microsoft's API which will convert the text into speech (audio).

Index Terms
Braille books, OCR, API

1. INTRODUCTION
OCR technology provides reproductive systems. It gives the ability to convert images of characters in a font of machine character that can be understood or recognized by a computer. Thus images of characters in a font of machine are drawn from a bitmap of the image reproduced by the scanner. This presented work is completely based on concept of Image Processing (IP).The OCR process involves several aspects such as segmentation, feature extraction and classification. Image Processing Toolbox for MATLAB provides a feature set that extends the product's capabilities to develop new algorithms and applications in the field of process and image analysis. Image processing is an absolutely crucial area of work for those groups and industries that are working in areas such as medical diagnostics, astronomy, geophysics, environmental science, data analysis in laboratories, industrial inspection, etc. Although not originally developed for users who are visually impaired, Optical Character Recognition (OCR) technology has become an aid for inputting documents quickly by and for users with vision impairments. A complete OCR system consists of a scanner, the recognition component, and OCR software that interact with the other components to store the computerized document in the computer. The process of inputting the material into the computer begins with the scanner taking a picture of the printed material. Then, during the recognition process, the picture is analysed for layout, fonts, text and graphics. Finally, the picture of the document is converted into an electronic format that can be edited with application software. In addition, technological developments will see an increase in accuracy, including the ability of these products to decipher even handwritten materials.

In general, OCR systems work as an external device with the user's existing assistive technology. Once the picture is in electronic format, it is accessed for reading and/or editing through the user's Braille, speech or magnification technology. Since some products work better with certain speech or Braille systems, therefore, it is important to note its compatibility with the other products in the user's computer system. Some products, however, have an adaptive device built in. These are referred to as "stand-alone reading machines."

2. BACKGROUND
Machine replication of human functions, like reading, is an ancient dream. However, over the last five decades, machine reading has grown from a dream to reality. Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. Many commercial systems for performing OCR exist for a variety of applications, although the machines are still not able to compete with human reading capabilities.

3. RELEVANCE
There are many documents in world which are not available in computer comprehensible format i.e. computer cannot read them. Computer vision and Robotics cannot be imagining without OCR. It acts as Benefit to Blind Persons. OCR has enabled scanned documents to become more than just image files, turning into fully searchable documents with text content that is recognized by computers. With the help of OCR, people no longer need to manually retype important documents when entering them into electronic databases. Instead, OCR extracts relevant information and enters it automatically. The result is accurate, efficient information processing in less time.

4. COMPONENTS OF OVERALL SYSTEM
The goal of Optical Character Recognition (OCR) is to classify optical patterns (often contained in a digital image) corresponding to alphanumeric or other characters. The process of OCR involves several steps including segmentation, feature extraction, and classification. Common setup is illustrated in figure1.

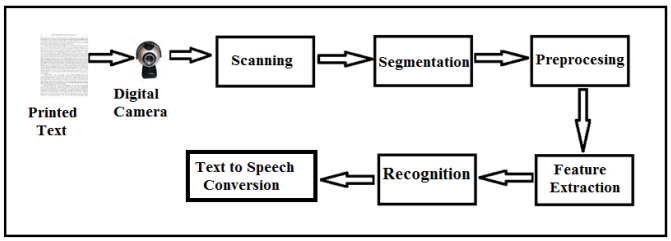

Fig 1: Components of an OCR system

A. Scanning

Through the scanning process a digital image of the original document is captured by the digital camera. In OCR optical scanners are used, which generally consist of a transport mechanism plus a sensing device that converts light intensity into gray-levels. Printed documents usually consist of black print on a white background. Hence, when performing OCR, it is common practice to convert the multilevel image into a bi-level image of black and white. Often this process, known as thresholding, is performed on the scanner to save memory space and computational effort. The thresholding process is important as the results of the following recognition are totally dependent of the quality of the bi-level image. A fixed threshold is used, where gray-levels below this threshold is said to be black and levels above are said to be white.

B. Segmentation

Segmentation is the isolation of characters or words. The majority of optical character recognition algorithms segment the words into isolated characters which are recognized individually. Usually this segmentation is performed by isolating each connected component that is each connected black area. This technique is easy to implement, but problems occurs if characters touch or if characters are fragmented and consist of several parts. The main problems in segmentation may be divided into four groups:

- Extraction of touching and fragmented characters.
- Distinguishing noise from text. Mistaking graphics or geometry for text.
- Mistaking text for graphics or geometry.

C. Pre-processing

Pre-processing techniques are indigence on colours, grey level or binary document images containing text and/or graphics. In character recognition systems usually applications utilize grey or binary images since processing the colour images is computationally high and tedious task. Such type of images may also contain non-uniform background and/or watermarks making it difficult to extract the document text from the image without performing some kind of pre-processing, therefore, the desired result from pre-processing is a binary image comprising text only. Thus, to achieve this, several steps are needed, such as image enhancement techniques to remove noise or improve contrast in the image, and thresholding to remove the background containing any watermarks, scenes or noise, third, page segmentation to isolate graphics from text, fourth, character segmentation to separate characters from each other and, finally, morphological processing to enhance the characters in cases where thresholding and/or other pre-processing techniques eroded parts of the characters or added pixels to them.

A. Gray Scaling

A grayscale or greyscale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information [1]. Grayscale images have many shades of gray in between.

B. Thresholding

It is the process of creating a binary image, that is, the image containing only black and white pixels. Separation of Background and Foreground is a main feature in thresholding.

C. Filtering (Median)

Removal of noise is the crucial task performed by Filtering. Median Filter is used generally to remove salt and pepper noise in an Image.

D. Thinning

This module reduces the thickness of bold letters and remove extra pixels which are not necessary.

Skeleton formation is performed by the process of thinning

Fig2: Original **Fig3: Gray scale** **Fig4: Thresholding**

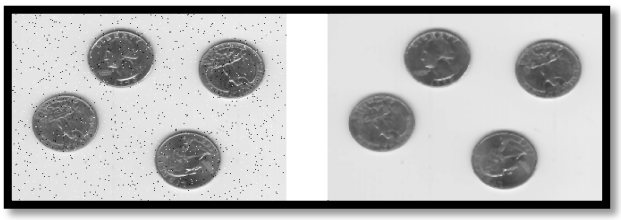

Fig 5: Removal of Noise

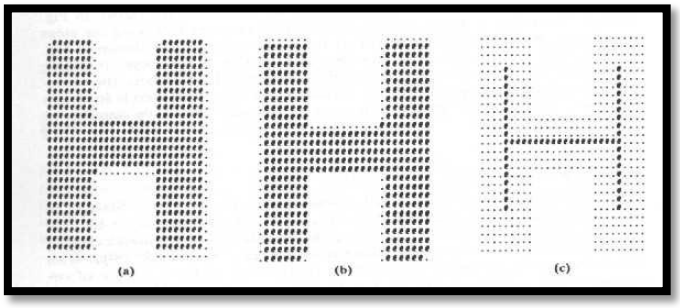

Fig 6: Thinning of an Image

E. Feature Extraction

It is the most important module implemented in OCR system. Feature extraction is a technique to extract the main features of a character to recognize it properly with maximum accuracy. It includes steps like cropping and scans lines. Feature extraction process is followed by Recognition process. It Compares character with the available templates. Result will be the output of comparison. In MATLAB mat2cell command is used for the extraction of image in form of a cell for correlating with the saved templates.

F. Recognition

Recognition process Compare character with the available templates. Result will be the output of comparison. Template matching techniques are different from the others in that no features are actually extracted. Instead the matrix containing the image of the input character is directly matched with a set of prototype characters representing each possible class. The distance between the pattern and each prototype is computed, and the class of the prototype giving the best matches assigned to the pattern.

The technique is simple and easy to implement in hardware and has been used in many commercial OCR machines. However, this technique is sensitive to noise and style variations and has no way of handling rotated characters.

G. Text to Speech conversion

Using speech application program interface (SAPI) the given text is converted into audio. The Speech Application Programming Interface is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. To date, a number of versions of the API have been released, which have shipped either as part of a Speech SDK, or as part of the Windows OS itself. Applications that use SAPI include Microsoft Office, Microsoft Agent and Microsoft Speech Server.

5. OCR PERFORMANCE EVALUATION

No standardized test sets exist for character recognition, and as the performance of an OCR system is highly dependent on the quality of the input, this makes it difficult to evaluate and compare different systems. Still, recognition rates are often given, and usually presented as the percentage of characters correctly classified. However, this does not say anything about the errors committed. Therefore in evaluation of OCR system, three different performance rates are investigated:

Recognition rate
The proportion of correctly classified characters

Rejection rates
The proportion of characters which the system was unable to recognize.

Rejected characters can be flagged by the OCR-system, and are therefore easily retraceable for manual correction

Error rate
The proportion of characters erroneously classified. Misclassified characters go by undetected by the system, and manual inspection of the recognized text is necessary to detect and correct these errors. There is usually a trade-off between the different recognition rates. A low error rate may lead to a higher rejection rate and a lower recognition rate. Because of the time required to detect and correct OCR errors, the error rate is the most important when evaluating whether an OCR system is cost-effective or not. The rejection rate is less critical. An example from barcode reading may illustrate this. Here a rejection while reading a barcoded price tag will only lead to rescanning of the code or manual entry, while a miscoded price tag might result in the customer being charged for the wrong amount. In the barcode industry the error rates are therefore as low as one in a million labels, while a rejection rate of one in a hundred is acceptable. In view of this, it is apparent that it is not sufficient to look solely on the recognition rates of a system. A correct recognition rate of 99% might imply an error rate of 1%. In the case of text recognition on a printed page, which on average contains about 2000 characters, an error rate of 1% means 20 undetected errors per page. In postal applications for mail sorting, where an address contains about 50 characters, an error rate of 1% implies an error on every other piece of mail.

6. RESULTS

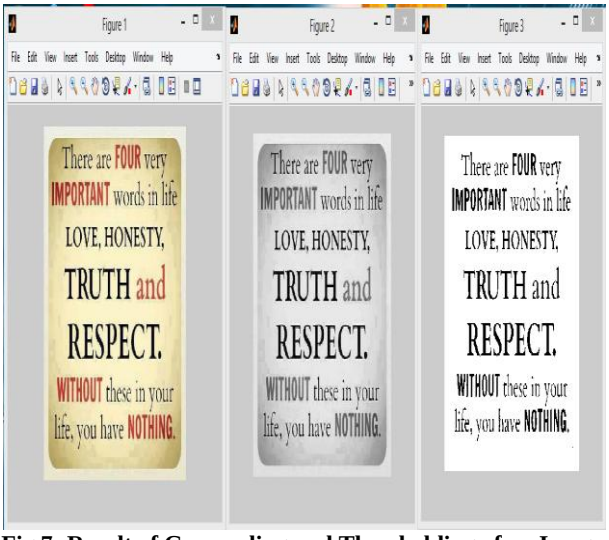

Fig 7: Result of Grayscaling and Thresholding of an Image

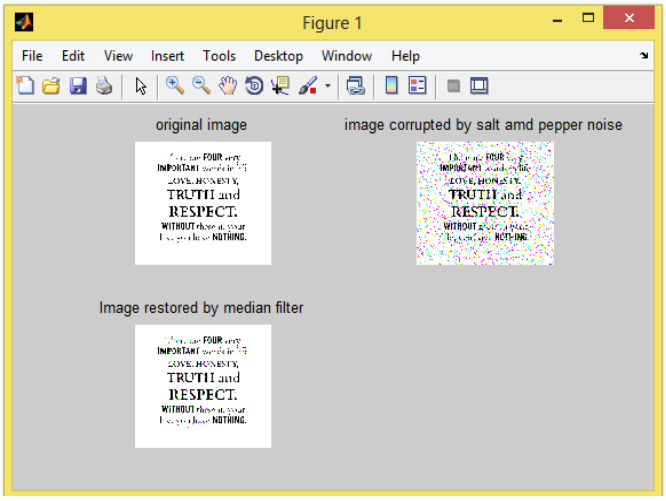

Fig 8: Result for Removal of Salt and Pepper Noise

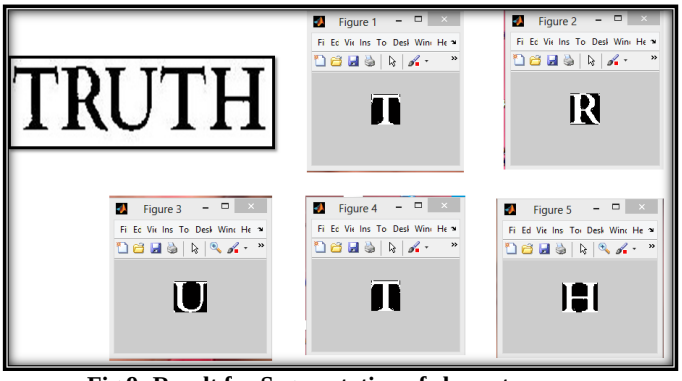

Fig 9: Result for Segmentation of characters

7. CONCLUSION

The project presents complete Optical Character Recognition system followed by text to speech conversion. The basics of image processing and MATLAB are taken into consideration. Because of the computing constraints of handheld devices, study is being kept limited to light-weight and computationally efficient techniques. Various algorithms for optical character recognition have been studied. Modules related to OCR are implemented which determines the efficiency and accuracy of project. Different text to speech conversion techniques and versions of SAPI are taken into consideration. Among those SAPI 5 version for text to speech implementation is preferred. The methods are kept as simple as possible so that economic and operational feasibility will be there.

8. REFERENCES

[1] Gonzalez Rafael .C "Digital Image Processing" Pearson Education Second Edition, Upper Saddle River New Jersey, USA, 2002

[2] Chapman Stephen .J "MATLAB programming for Engineering" second edition, 2002

[3] Forsyth David .A, Ponce Jean "Computer Vision-A Modern Approach" Pearson Education, First Edition, Upper Saddle River New Jersey, USA, 2003

[4] Sandeep Tiwari, Shivangi Mishra, Priyank Bhatia, Praveen Km. Yadav, "Optical Character Recognition using MATLAB" International Journal of Advanced Research (2013), Volume 1, Issue 9, 757-767

[5] Oi-Mean Foong and Nurul Safwanah Bt Mohd Razali, "Signage Recognition Framework for Visually Impaired People", In.proc.of International Conference on Computer Communication and Management, Vol.5, 2011.

[6] Azadeh Nazemi and Iain Murray,"An open source reading system for print disabilities", International Conference on Computing, E-Learning and Emerging Technology & International Conference on Advances in Computer , Electrical and Electronic Engineering - Sydney, Australia, (ISSN : 2091-1610) , Volume No : 12 Issue No : 2