Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-5, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

Blind Speech Using OCR

[1]Shreejit Achari, [2]Nishkarsha Kotian, [3]Suyog Marde &
[4]Prof Seema Redekar
SIES Graduate School of Technology, India.

Abstract*: The idea of converting written or printed text into digital text is generally called OCR i.e., optical character recognition. Optical character recognition (OCR) can be based on conversion of handwritten text or printed characters as in textbooks into machine editable form. The input handwritten text is scanned and its Digital image form is obtained. The image is handled with the help of Enhancement techniques, segmentation, image recognition and neural networks. The flexibility of this design allows it to extend to other languages easily. The software will be implemented using Neural Networks, which consists of a training phase, a validating phase and a testing phase.*

Optical Character Recognition (OCR) is one of the oldest sub fields of pattern recognition with a rich contribution for the recognition of printed documents. OCR systems scan the documents printed on a paper as an image and recognize the characters present in the document image to form a separate digital text document, which can be edited or processed.

Because of an expanding use of mobile technology, getting information for illiterate humans is made easy. Until now, many of OCR techniques were used just for personal / professional / business purpose. Here we have been taking this project to a new dimension. The inadequacy of language barrier for using various technologies can be eradicated. The OCR Project is also aligned with corporate strategy and objectives since it uses technology to improve the way we do business eliminating problems for language inadequacy. While other alternatives and the status quo were analyzed, the OCR Project was selected for proposal in this business case because it provides the best platform to integrate solutions for various problems faced by (for) personal / professional / business purpose.

Introduction

OCR Stands for Optical Character Recognition. Any OCR application is able to recognize and extract text out of a scanned document, such as PDF, or other formats from various image files. A PDF-to-text Converter can convert any scanned PDF document into an editable text document and this gives the user the ability to make changes of his/her choice.

By Definition:

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic conversion of scanned images of handwritten, typewritten or printed text into machine-encoded text. It is widely used as a form of data entry from some sort of original paper data source, whether documents, sales receipts, mail, or any number of printed records. It is a common method of digitizing printed texts so that they can be electronically searched, stored more compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

If you have a scanned document in formats such as TIFF, PDF, PNG etc, you can't alter, copy or remove any text.

Actually when you scan a paper using a scanner and save it as any supported document formats, the whole content will be captured as an image instead of text and font information. In this case, you'll need to convert the image into editable content before you can make changes to it.

That's exactly what an OCR app does, extracting text from images and then converting them into a PDF document, which can later be converted into an editable text document if needed.

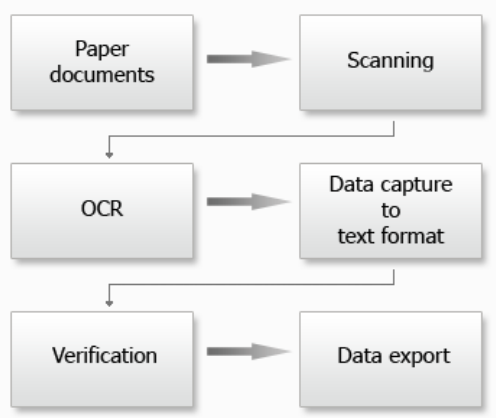

Fig.1 Working of Typical OCR Software

Tesseract

Tesseract [1] is probably the most accurate open source OCR engine available. Combined with the Leptonica Image Processing Library it can read a wide variety of image formats and convert them to text in over 60 languages. Tesseract is an optical character recognition engine for various operating systems. It is free software, released under the Apache License, Version 2.0, and development has been sponsored by Google since 2006.Tesseract is considered one of the most accurate open source OCR engines currently available. Tesseract is suitable for use as a backend, and can be used for more complicated OCR tasks including layout. Tesseract's output will be very poor quality if the input images are not preprocessed to suit it: Images must be scaled up such that the text x-height is at least 20 pixels, any rotation or skew must be corrected or no text will be recognized, low-frequency changes in brightness must be high-pass filtered, or Tesseract's binarization stage will destroy much of the page, and dark borders must be manually removed, or they will be misinterpreted as characters. Apart from tesseract there are many other OCR engines available like ABBYY, aspose etc. which provide better accuracy than tesseract. Tesseract does not come with a GUI and is instead run from the command-line interface.

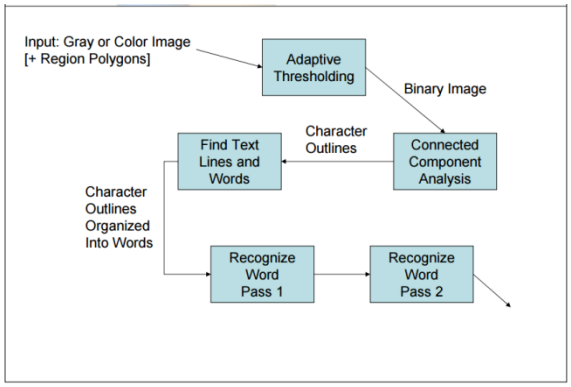

Fig.2 Tesseract Architecture

Need of this software

Because of optical character recognition a user does not need any specific training / teaching to use this software, thus eliminating the problem of people who are not able to read any or specific language and need support for the same.

Visually impaired people are dependent solely on Braille books & audio recordings provided by NGOs. Owing to many constraints in above two approaches blind people can't have book of their choice. The presented work will provide them an opportunity to have an audio-book of their choice in English or Marathi language of any printed book having English, Marathi or Braille script. Printed text from textbook having English, Marathi or Braille script will be taken as input in the form of an image which will be converted into plain editable text with the help of Optical Character Recognition (OCR). This plain text will be then fed to Text to Speech (TTS) converter which will generate the audio output file in English or Marathi language corresponding to the input text image script. People with various reading disabilities will be benefitted by these audio books. Even people without a disability could benefit from this as it will provide an alternative to physical reading material; for example, while riding a bus, instead of reading a book, people could listen to the audio book.

Issues in Existing System:

- They are not 100% Accurate.
- They do not have proper Graphical User Interface.
- Text and Audio conversion is not provided under single platform.
- The Characters in Existing System are trained by the developer.

Objectives

Since its inception, this project has relied upon application of various technologies to manage language barrier problems and other personal / professional / business problems that come along with OCR software. As the number of connection grows, so does the level of inadequacy for proper information transmission due to various linguistic manners. In the last couple of years, a diverse country like India faces the problem of understanding various language and/or their written transcripts. Literacy rate is not at top even though it is rising each year. So the main problem lies for the working society of higher age.

The Project directly supports several of the personal / professional / business goals and objectives.

Table 1. Goals and objectives.

Objective	Description
Text Object to transcript	With the help of optical hardware in use, any text can be converted into transcript.
Word to Speech	People, who cannot read the formatted transcript by the software, can have a listening support for the same.
Speech to Audio	Transcripts will be converted into an audio book having the preferred format of the user.
Language translate	These audio books can then be translated based on requirements of the user and the options available.

Implementation results

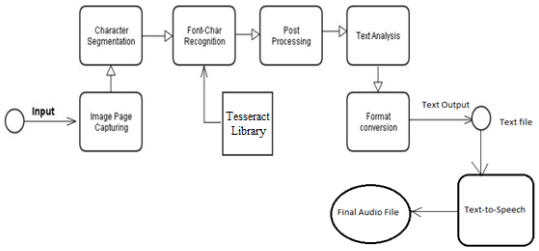

Fig3. Flowchart

The input given in the above flow chart is the image one would like to process. After processing of the image, i.e doing character segmentation and Font-Char Recognition, Post processing of the image is ready.
The output that is obtained after this stage is a editable text file which is then fed to the text-to-speech converter.
After which the final output is obtained that is the audio file, ready to be heard.

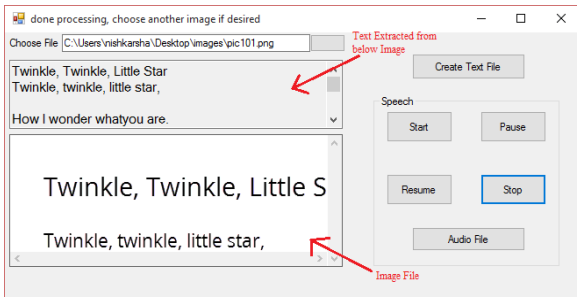

Fig.4 Screenshot

In the above screenshot, the image file is displayed below.
After processing the image, text extracted from the image can be saved as a text file by clicking on the "Create Text File" button.
It is further processed and converted into a audio file which can be paused, resumed or stopped according to the convenience of the user.
By clicking on the "Audio File" button, an audio file is created and saved in a location of the user's choice.

The above Project was made in Visual basic. Tesseract Libraries where used for processing and extracting text from the image file. The project was divided into two modules, namely

1. Text Extraction using OCR
2. Text to Speech

Create Text File Button: Creates a Text File which can be edited and used for further processing.

Speech Module: Here we are making use of speech synthesis engine; The Speech Synthesizer class provides access to the functionality of a speech synthesis engine that is installed on the host computer. Installed speech synthesis engines are represented by a voice, for example Microsoft Anna. A Speech Synthesizer instance initializes to the default voice.

Audio File Button: Creates a .mp3/.wav audio file, which can be copied into any device and can be listened without the need of an actual book.

Benefits

- We intend to perform multiple functions in addition to OCR, such as PDF conversion and compression. OCR products have an accuracy rate above 99% and can process up to 20 pages per second.
- It increases the efficiency and effectiveness of office work.
- Workflow is increased since we no longer have to waste time on manual labour and can work quicker and more efficiently.

Limitations

- Text from a source with a font size of less than 12 points may result in errors.
- The output from a finished text scan might require spellchecking and proofreading as well as reformatting to desired final layout.
- Language jargons may differ from library functions.

Future scope

- OCR application will also display the signatures and the other symbols as it is in the document.
- It will also update its features including the translation of one language to another, so that it will helpful for people from other countries who can't understand the local language.
- This OCR Software will be modified into a mobile application, so that the user can directly use his Smartphone to perform the above activities using the mobile camera.
- In case of wrong words detected during Text to Speech, it will automatically add a beep instead of the wrong word being pronounced.

Conclusion

This project, i.e. "Blind Speech using OCR" is software that converts an image to text document which later converts it into an audio file, can be implemented with the translation of language feature and the benefits of this being implemented successfully could affect a lot of people.

This Software can also be made in such a way that it trains according to each user, uniquely, as opposed to the usual developer trained handwriting system.

References

[1] Tesseract Software. Available: https://en.wikipedia.org/wiki/Tesseract_(software)

[2] Speech Synthesizer. Available: https://msdn.microsoft.com/en-IN/library/system.speech.synthesis.speechsynthesizer(v=vs.110).aspx

[3] Ravina Mithe, Supriya Indalkar, Nilam Divekar, "International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-1, March 2013"