Image Reader Using Optical Character Recognition and Text-To-Speech Technique

Shaishavkumar A Patel — Metatree Systems Pvt. Ltd., Surat, Gujarat, India.

Dr. Kamaljit I. Lakhtaria — Associate Professor, Of Computer Science, Gujarat University

ABSTRACT The Tesseract is Open Source OCR engine, It was initially developed and maintain by Hewlett Packard labs, and since 2006 Google is working on it.[1] Optical character recognition (OCR) is process for Converting printed text into editable text. OCR is widely used in Machine translation, Text-To-Speech, Text Mining etc.[2] This Paper combines two technology Optical Character recognition (OCR) and Text-To-Speech (TTS). This System will take image as input and extract text from given image and convert generated text into speech.

INTRUDUCTION
OCR

Optical Character Recognition is conversion of images of handwritten, printed or typed text into editable text. OCR technology is capable for recognizing both handwritten and printed text. Result of OCR is mainly depends on quality of image. As OCR extract text out of image, this technology enables you to convert most type of documents into editable and searchable data. Images captured using digital camera differ from scanned document, as it may have distortion at the edges or can have dimmed light, this cause difficult for most OCR applications to get higher result. To perform OCR, our system involves three important steps. First is segmentation, Second is feature extraction, and Final task is classification.[3] In 1955, the first commercial system was installed, which used OCR to input sales report into a computer and then after OCR has become helpful in computerizing the physical office documents. There are many applications that use OCR, which includes: License plate recognition [4-11], OCR is designed to process different format of images that consist text as a major part with very little non-text part obtain from image captured by mobile phone camera. This application is for Android Operating System that also uses Google's Open-source OCR engine, Tesseract OCR.[12] Google's language translation service and the Android text-to-speech synthesizer, which allow phone users to take picture of text using mobile camera and have the text read aloud.

Most of OCR application using input image with a scanner or a digital camera and a Software. If you do not have a scanner or a digital camera, a hardware problem occurs. In order to overcome this limitation of computer occupying a larger space, a system based on android phone is proposed [13].
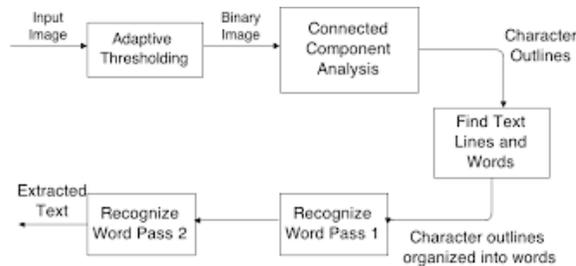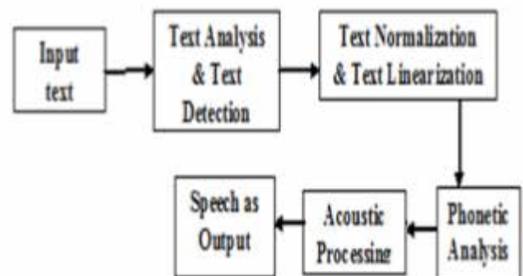

Figure 1 Architecture of Tesseract
Tesseract takes image as input and using thresholding it converts into binary format. Outlines of components are stored on connected Component Analysis. Nesting of outlines is done which gathers the outlines together to form a Blob. Text lines are analyzed for fixed pitch and proportional text. Then the lines are broken into words by analysis according to the character spacing. Fixed pitch is chopped in character cells and proportional text is broken into words by definite spaces and fuzzy spaces. Tesseract performs activity to recognize words. This recognition activity is mainly consists of two passes. The first pass tries to recognize the words. Then satisfactory word is passed to Adaptive Classifier as training data, which recognizes the text more accurately. During second pass, the words which were not recognized well in first pass are recognized again through run over the page. Finally Tesseract resolves fuzzy spaces. To locate small and capital text Tesseract checks al-

a) Text Analysis & Detection:
Text analysis is a part of pre-processing. It analyzes the text and organizes it into manageable list of words. Then these words are transforms into full text. It localizes the text areas from printed documents.

b) Text Normalization & Linearization:
Text Normalization is the conversion of text to pronounceable form. It is often performed before text is processed in some way, such as generating automated language translation or synthesized speech. The main objective of this task is to identify punctuation marks and pauses between words. Usually the text normalization process is done for converting all letters of lowercase or uppercase, stop words or too common words, to remove punctuations, accent marks and other diacritics from letters.

c) Phonetic Analysis
Phonetic Analysis provides phonetic alphabets. The grapheme to phoneme conversion is done in this step. It is actually a conversion of orthographical symbols into phonological symbols.

d) Acoustic Processing.

It performs formant synthesis. It works intelligently and thus does not require any kind of database of speech samples. For speak out the text, it uses voice characteristics of a person.

II. PROPOSED SYSTEM

OCR technology allows the conversion of scanned images of printed text into text or information that can be understood. Open source OCR software called Tesseract as a basis for Optical Recognition project, which is considered as the most accurate free OCR engine in existence. In this system it takes the Image as the input and extract the text from it, this steps involve the use of OCR technology, we have used Tesseract to do OCR. This extracted text is used as a input for next step, which involve converting text into audio (Text to Speech).

III. CONCLUSION

The proposed system internally used Optical Character recognition and Text To Speech. We assumed that input images are only document images, and designed the proposed system based on that.

A. LIMITATIONS

1). Accuracy of this system directly depend on Optical Character Recognition and Input image
2). Final output is Audio which also depend on the TTS result.

B. FUTURE SCOPE

This system can be enhanced and connect to the web system which will interact with user and read the document requested by the user. By doing this, the front of this system will not depend on any operating system, So it will become plate-form independent.

REFERENCES

[1] Wikipedia Tesseract OCR Software [https://en.wikipedia.org/wiki/Tesseract_(software)] access on May 2016.

[2] Wikipedia Optical Character recognition [https://en.wikipedia.org/wiki/Optical_character_recognition] access on May 2016.

[3] Smith, Raymond W. *The Extraction and Recognition of Text from Multimedia Document Images*. Diss. University of Bristol, 1987.

[4] Anagnostopoulos, Christos Nikolaos E., et al. "A license plate-recognition algorithm for intelligent transportation system applications." *Intelligent Transportation Systems, IEEE Transactions on* 7.3 (2006): 377-392.

[5] Wu, Hui, and Bing Li. "License plate recognition system." *Multimedia Technology (ICMT), 2011 International Conference on*. IEEE, 2011.

[6] Fan, Xin, and Guoliang Fan. "Graphical models for joint segmentation and recognition of license plate characters." *Signal Processing Letters, IEEE* 16.1 (2009): 10-13.

[7] Wen, Ying, et al. "An algorithm for license plate recognition applied to intelligent transportation system." *Intelligent Transportation Systems, IEEE Transactions on* 12.3 (2011): 830-845.

[8] Zheng, Lihong, et al. "Accuracy enhancement for license plate recognition." *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*. IEEE, 2010.

[9] Jiao, Jianbin, Qixiang Ye, and Qingming Huang. "A configurable method for multi-style license plate recognition." *Pattern Recognition* 42.3 (2009): 358-369.

[10] Kocer, H. Erdinc, and K. Kursat Cevik. "Artificial neural networks based vehicle license plate recognition." *Procedia Computer Science* 3 (2011): 1033-1037.

[11] Roy, Ankush, and Debarshi Patanjali Ghoshal. "Number Plate Recognition for use in different countries using an improved segmentation." *Emerging Trends and Applications in Computer Science (NCETACS), 2011 2nd National Conference on*. IEEE, 2011.

[12] Bhaskar, Sonia, Nicholas Lavassar, and Scott Green. "Implementing Optical Character Recognition on the Android Operating System for Business Cards." *Radical Eye Software., Palo Alto, CA* (2007).

[13] Lian, Wing-Soon Wilson, and Diane P. Pozefsky. *Heuristic-Based OCR Post-Correction for Smart Phone Applications*. Diss. Thesis. Chapel Hill: The University of North Crolina, 2009.

[14] Sasirekha, D., and E. Chandra. "Text to speech: A simple tutorial." *International Journal of Soft Computing and Engineering* 2.1 (2012): 275-278.

[15] Lakhtaria, Kamaljit I. "Analyzing Zone Routing Protocol in MANET Applying Authentic Parameter." arXiv preprint arXiv:1012.2510 (2010).

[16] Lakhtaria, Kamaljit I. "Enhancing QOS and QOE in IMS enabled next generation networks." Networks and Communications, 2009. NETCOM'09. First International Conference on. IEEE, 2009.

[17] Lakhtaria, Kamaljit I., and Dhinaharan Nagamalai. "Analyzing Web 2.0 Integration with Next Generation Networks for Services Rendering." Recent Trends in Networks and Communications. Springer Berlin Heidelberg, 2010. 581-591.

[18] Kamaljit, Lakhtaria. "Analyzing Zone Routing Protocol in MANET Applying Authentic Parameter."

[19] Lakhtaria, Mr Kamaljit, et al. "Securing AODV for MANETs using Message Digest with Secret Key." Nettork Security & Its Applications (IJNSA) 1 (2009): 111-116.

[20] Hidoussi, Faouzi, et al. "Centralized IDS based on misuse detection for cluster-based wireless sensors networks." Wireless Personal Communications 85.1 (2015): 207-224.

[21] Bhavsar, Krunal S., and Kamaljit I. Lakhtaria. "Black Hole Problem in AODV Routing Protocol in MANET."