International Journal of Latest Engineering and Management Research (IJLEMR)
ISSN: 2455-4847
www.ijlemr.com || Volume 03 - Issue 04 || April 2018 || PP. 64-67

Reading text extracted from an image using OCR and android Text to Speech

Nilesh Jondhale[1], Dr. Sudha Gupta[2]
[1](*Department of Electronics Engineering, K.J. Somaiya College of Engineering, India)*
[2](*Department of Electronics Engineering, K.J. Somaiya College of Engineering, India)*

Abstract: Now-a-days Machine learning has become one of the peak of technology. Previously it was not possible to compute data at higher or faster rate, with the help of leading technology it is now possible to process data at higher rate to get optimized hence better result. Pattern recognition, a branch in machine learning is/can be helpful in many different ways. OCR is used to recognition of character with high accuracy. Using handheld mobile device camera for capturing an image of a printed or handwritten document to generate text from the same. On global scale there are billions of android devices running. With the help of android device and android text to speech we can convert text into an effective & accurate speech optimally.
Keywords: Android, Machine Learning, OCR, Text-to-Speech

I. INTRODUCTION

There is a lot research work has done on Pattern Recognition which comes under Machine Learning, Artificial Intelligence. OCR well known as Optical Character Recognition is one of the leading branch of the Pattern Recognition [1]. The system reduces human efforts along with time. It can also be helpful for the person who doesn't know the language / pronunciation of particular words. The system can also be helpful for visually impaired or person with weak visual ability. Optical character recognition is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document or a photo of document [2]. It is widely use as form as a form of information entry from printed paper data records, whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static data or any suitable document. OCR is a field of research in pattern recognition, artificial intelligence, computer vision [3]. The app uses a camera of an Android mobile device to take an input. Input is a binary image scanned by the camera. The OCR engine processes the image data and converts it into a text [4]. The respective text is then sent to Android Text-to-Speech. Android text-to-speech is an engine which has ability to convert the text into a speech. The system uses machine learning, it takes a training data and learns from it, hence the accuracy of the output grows down the pages, pass by pass [5].

II. METHODOLOGY

1. Functional Use of a System – There are stages and steps in which the user can use the system and functionalities where the system can perform its own tasks. The figure mentioned represents the use case of the system. An actor is nothing but the use of the system / application. There are two easy ways by which user can interact with the system, also there are two main functionalities by which system comes into a picture. Initially the user will start the app. The rear camera is kept open by default in the system, so that it is easy for users to use the system without much hurdle. The actor will scan hardcopy document by the camera of this android device phone. The scanned image can be anything like printed or handwritten document. Once there is image in front of the screen the system functionalitycome into picture. The system will try to identify and generate the text from the image that is provided to the OCR engine.

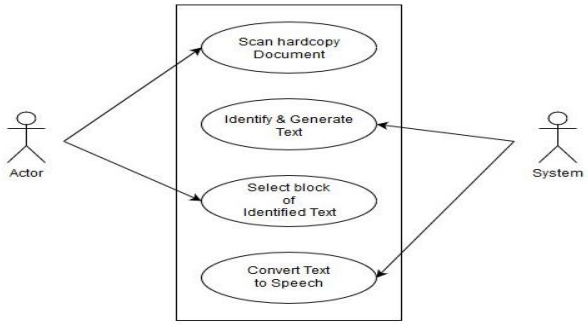

Fig. 1. Functional use of system

Once the user sees the detected text and is converted into a machine identified and generated text, the actor has choice to select the text he desired. Once the actor selects the block of identified text it is considered as an input to the text-to-speech. The another functionality of the system comes into picture and the system will convert the text into an effective speech.

2. System Working – The user scans the image at the first stage. The user will be holding an android device, the rear camera is considered as by default open and the user will scan an image. The binary image is then sent to the OCR engine for the pre-processing over the image. In the pre-processing the text is recognized by different terms.

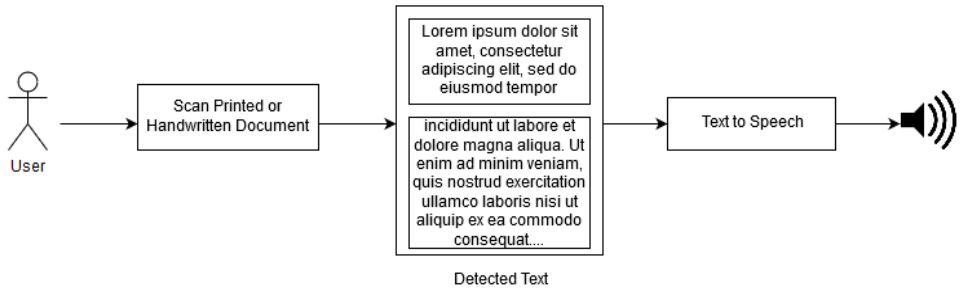

Fig. 2. System Working

The system follows step by step pipeline architecture. Below figure shows the pipeline processing the output of one stage is given to another stage as its input. In stage II of the system the system will pass the connected component analysis over the text. The connected component analysis is a computationally expensive process, but as smartphones contains higher end processors it does not take much processing to get a desired output. The main advantage of this process is it can recognize the text no matter whether the text is black-on-white or white-on-black [3]. Connected component analysis performs the inspection of the nesting of outlines & the number of child & grandchild outlines. Once we successfully recognized the child & grandchild outlines it is easy to recognize whether the text is black-on-white or white-on-black. KNN and Naive Bayes algorithms use search history database, User's Metadata and search engine results database to classify data according to user's need and gives optimized result [7].

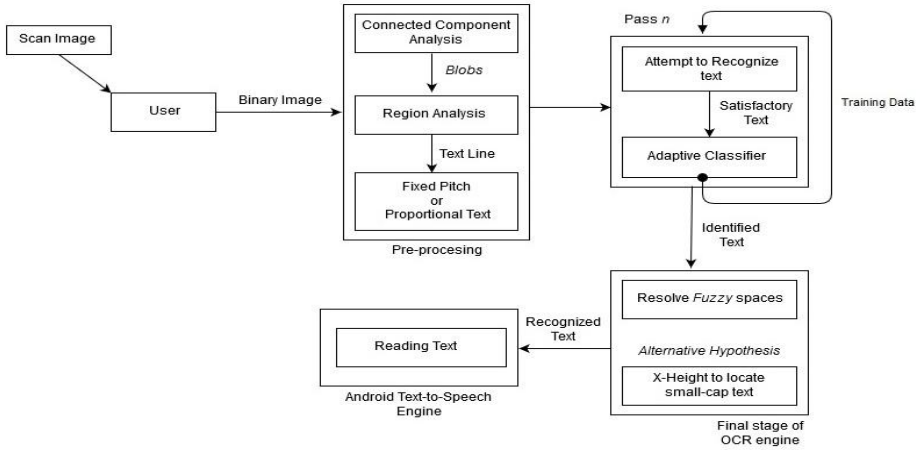

Fig. 3. System Architecture

Once we successfully detected the pattern of the text. At this step outlines are gathered together, those outlines are nested into Blobs. Blobs are nothing but the connected but not completely organized text. Later nested blobs are organized into text lines and regions are analyzed for fixed pitch or proportional text. The fixed pitch is nothing but the text in which all characters uses the same space. Text lines are broken into words differently according to the kind of character spacing. The equal character spacing like monospace is an example of fixed pitch. On contrary to this the proportional text does not have equal spacing between all the characters. Hence both type of text should be treated separately. Fixed pitch text is chopped immediately by character cells, whereas proportional text is broken into words using definite spaces and fuzzy spaces [3]. In the

next stage the recognition of text and machine learning processing is applied. This stage gets the input from the previous stages output. Which is a recognized text, it might be either fixed pitch or proportional text. On that input an attempt is made to recognize each word in the text. Each word that is satisfactory is passed to an adaptive classifier as a training data. The adaptive classifier then gets chance to more accurately recognize the text lower down the page. The adaptive classifier checks the data, and the detected text passes for further processing. Till now the system has detected and successfully recognized much of the text. At the start of the text there might be a chance that there is some undetected text, hence the trained data from the adaptive learner is sent over for the pass II of character recognition. In both the passes the system has detected maximum detectable text hence considered it as an output. Since the system again sorts small and capital letters for better accuracy, the generated text considered as an identified text. This text is then sent to the final stage of theOCR engine. To resolve fuzzy spaces that lies in the proportional text. The alternative hypothesis may be applied to recognize the X-height of the text so that system can locate the small-cap text at the final stage of the OCR engine [3]. Hence we successfully fetched the text from an image at this stage. The text is then further sent to android text-to-speak engine. This engine gets its input as a text and converts it into a speech. The engine comes in build in most of the android devices. The final output as a speech is provided though the speaker of the smartphone device.

III. RESULTS

The proposed system named as 'Digital Eye – Reading a text extracted from an image using Optical Character Recognition and Android Text-to-speech'. The proposed system extracts the text from an input image and converts it into a speech. The figure below shows the proposed / possible outcome of the system. It is observable from the figure below, the input image (left) represents the input image. It will scan the document. The system identifies the text from that image and generate an output as a text, as shown in the figure (right). The blocks represent an expected output as detected text divided into blocks.

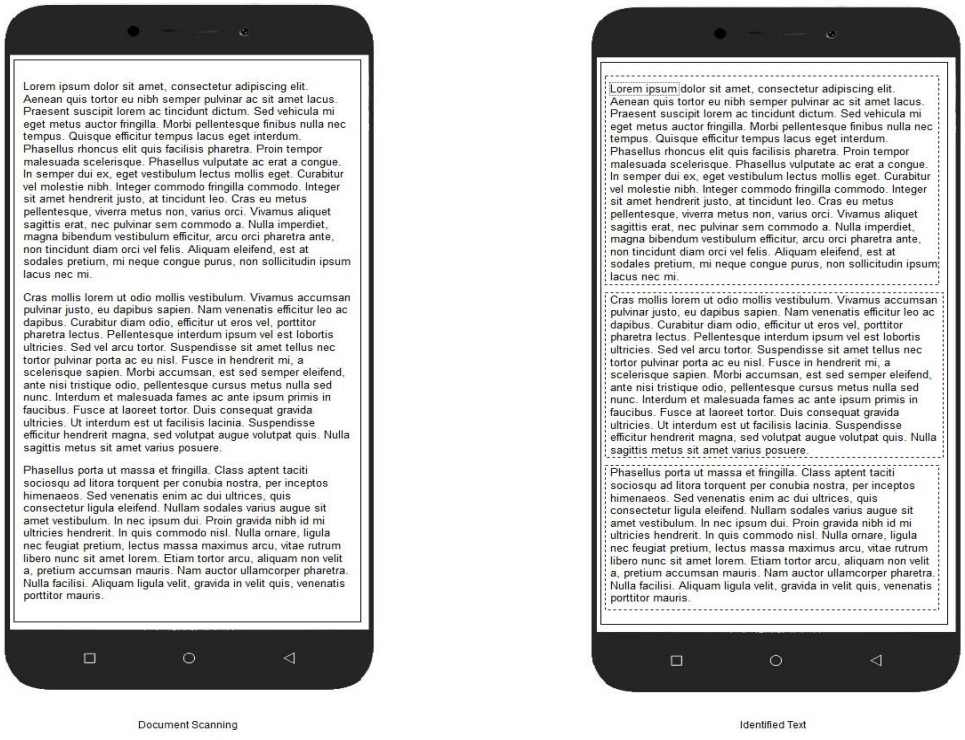

Fig. 4. Output of Digital Eye Application

IV. CONCLUSION

The proposed system contains mobile application named as 'Digital Eye'. This application uses rear camera of mobile device as a default input device. It gathers data in binary image format, those images are processed for the Optical Character Recognition. The OCR recognizes the text from a binary image, and converts it into machine generated text. The proposed system is also able to recognize the printed as well as handwritten text with higher accuracy. The OCR text detection is based on the connected component analysis

and feature detection. It gathers outline and converts it into Text lines, those lines are then analyzed for fixed pitch or proportional Text. An attempt is made to recognize each word, and successfully detected data will be sent to an adaptive learner. Further the system goes on training providing more accurate results. The Digital Eye Application work on Offline mode and it requires less memory size on RAM. Initially the scope is limited to a particular language such as English. Considering the part of development, the system can import an image from a gallery and read aloud the text in it. With addition to that PDF scanning and reading functionality can also be provided in future. User should be able to access the scanned document later in time, hence the save functionality can be provided. The system can be more improved to take the app globally such as it is possible to make such system for regional languages and improve localization & globalization. Along with it multilingual output can be implemented so that a person can scan document in any language and gets its output in his desired / local language. The recognized data is sent further for converting it into speech using Android text to Speech. The system reduces human efforts along with time. It can also be helpful for the person who doesn't know the language / pronunciation of particular words. The system can also be helpful for visually impaired or person with weak visual ability.

ACKNOWLDGEMENT

I would primarily like to thank my mentor Dr. Sudha Gupta, for guiding us throughout the entire project work, motivating us to face the encountered challenges and helping us at all stages to take the project to higher standards of refinement. Without her immense support and guidance, this project would not have been possible.

REFERENCES

[1] Recognition of Offline Handwritten Characters using the Tesseract open source OCR engine, Qi Li, Weihua An, Anmi Zhou, Lehui Ma, 2016
[2] Optical Character Recognition using Artificial Neural Network, T. K. Das, Asis Kumar Tripathy and Alekha Kumar Mishra, 2017
[3] An overview of the Tesseract OCR engine. Ray smith, Google Inc.
[4] Offline Character recognition system using artificial neural network. Nisha Vasudeva, Hem Jvotsana Parashar and Singh Vijendra, 2012
[5] Optical character recognition (OCR) performance in server based mobile environment, Teddy Mantoro, Abdul Muis Sobri, Wendi Usino, 2013
[6] Designing mobile application for retrieving book information using optical character recognition. Nana Ramadijanti, Achmad Basuki, AgrippinaG.J, W. 2016.
[7] Optical Character Recognition using back propagation neural network. Shyla Afroge, Boshir Ahmed, Firoz Mahmud, 2016.
[8] Image text to speech conversion in desired language by Translating with Raspberry Pi. Rithika H. Nithiya Santhoshi. 2016.
[9] J. Pradeep, E. Srinivasan and S. Himavathi (2011). "Diagonalbased feature extraction for handwritten alphabets recognitionsystem using neural network". International Journal ofComputer Science & Information Technology (IJCSIT),Volume 3, No 1, pp.27-38
[10] Detecting Text Based Image with Optical Character Recognition for English Translation and Speech using Android, Sathiapriya Ramiah, Tan Yu Liong, Manoj Jayabalan, 2015
[11] A. Shaik, G. Hossain and M. Yeasin, 'Design, development and performance evaluation of reconfigured mobile Android phone for people who are blind or visually impaired', *Proceedings of the 28th ACM International Conference on Design of Communication – SIGDOC10*, 2010.
[12] Sherine Sebastian and Priya S., "Text Detection and Recognition from Images as an Aid to Blind Persons Accessing Unfamiliar Environments," Asian Research Publishing Network (ARPN) Journal of Engineering and Applied Sciences, ISSN 1819-6608, Vol. 10, No. 17, September 2015.
[13] Text Detection and Recognition with Speech Output for Visually Challenged Person. Rupali D. Dharmale, Dr. P. V. Ingole, 2016
[14] Design and Implementation of Konkani Text to Speech Generation System using OCR Technique, John Colaco, 2016
[15] Computer Vision and Text Recognition for Assisting Visually Impaired People using Android Smartphone Hao Jiang, Thomas Gonnot, Won-Jae Yi and Jafar Saniie 2017