International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Optical Character Recognition - An English Assistant with Digitization, Summarization and Read Aloud Functions

Dhiraj Amin[1], Sitadevi Muthkhod[2], Mebin Philip[2], Meenu Madhu[3], Vivek Menon[4]

[1]*Professor, Dept. of Computer Engineering, MES. Pillai College, Maharashtra, India*
[2,3,4,5]*Student, Dept. of Computer Engineering, MES. Pillai College, Maharashtra, India*

Abstract - *OCREADS (Real-time Optical Character Recognition - an English Assistant with Digitization, Summarization and Read Aloud Functions) is basically an implementation of Intelligent Character Recognition in the field of images, an application that will enable you to convert images of printed or typed characters into digital text. OCR (Optical Character Recognition) is used for preparing a manuscript for submission, digitizing journals and notes to make them easy to search and use or converting old documents to electronic format. This is very helpful for the blind and illiterate people for whom this software can read out loud the text that is detected using the Text-to-Speech feature. Additionally, if the content is too long and the user wishes to have a short summary of the text, the Text Summarization feature comes in handy. For instance, traffic symbols, public hoardings, or instructions on a railway station indicator board can be read out to the user. It can be used for academic purposes for converting printed notes into searchable documents. Government and medical institutions can use this to get data from forms so that it can be converted into searchable digital documents. Thus OCR can be a solution to a wide range of applications.*

Key Words: **OCREADS, Printed OCR, Text-to-Speech, Summarization, English Character Recognition**

1. INTRODUCTION

People with visual impairments face various difficulties while accessing printed text using existing technology that includes problems with alignment, focus, accuracy, mobility and efficiency. We present a software that assists the visually impaired which can read the paper-printed text. The proposed project uses the methodology of a camera-based assistive device that can be used by people to read text documents. The framework is on implementing an image capturing technique in a camera or web-cam enabled system. The design is obtained from studies conducted on visually impaired people, and it is small-scale and mobile, which enables a more manageable operation with little setup. In this project, we have proposed a text readout system for the visually challenged. The proposed fully integrated system has a camera as an input device and speaker as an output device.

2. LITERATURE SURVEY

2.1 Webcam-based Optical Character Recognition

Webcam based Optical Character Recognition is a system used to recognize the character or alphabets in the given text by comparing two images of a certain alphabet. The objective of the OCR system is to develop a program for the Optical Character Recognition (OCR) system by using the Template Matching algorithm. This system has its own scopes which use Template Matching as the algorithm that can be applied to recognize the characters, both in lowercase as well as uppercase, and the digits used with courier new font type, using bitmap image format with 240 x 240 image size and recognizing the alphabets by comparing between images that have been stored. The purpose of this system prototype is to solve the problems of blind people who are not able to read, in recognizing the character which is before that it is difficult to recognize the character without using any techniques and Template Matching is one of the solutions to overcome the problem.

2.2 OCR System for English Language

Optical character recognition from scanned images is a challenging task. But as for record-keeping, we need all the data in digital format to perform manipulative operations such as searching, modifying, adding or deleting some records etc. The main issue in case of character recognition is the different styles and fonts in which the text is present. The authors Honey Mehta, Sanjay Singla, Aarti Mahajan [2] have proposed a new approach by using the concept of Artificial Neural Network and Nearest Neighbour approach for character recognition from scanned images. Three layers are used for the classification purpose. First is the input layer that consists of the input given by the segmented characters. The hidden layer consists of neurons trained by the network and the output layer consists of output neurons to generate Unicode of the input characters.

2.3 Extractive Text Summarization

The authors N. S. Shirwandkar and Samidha Kulkarni [3] develop an approach for Extractive text summarization for single-document summarization. They use a combination

of Restricted Boltzmann Machine and Fuzzy Logic to rank important sentences from the text without losing the meaning and keeping the summary lossless. The text documents used for summarization are in English. Various sentence and word level features are used to generate meaningful sentences. For every document two summaries are generated using Restricted Boltzmann Machine and Fuzzy logic. Both summaries are then combined and processed using a set of operations to get the final summary of the document.

2.4 Python Based Portable Virtual Text Reader

The author Hasan U. Zaman, Saif Mahmood, Sadat Hossain and Iftekharul Islam Shovon [3] implemented a product based on a Raspberry Pi module that also has a camera connected to it which is used to take pictures. The whole bodywork is also integrated with Optical Character Recognition (OCR), Text-To-Speech (TTS) and a speaker. A Graphical User Interface (GUI) helps the users to take pictures in a couple of clicks. A button is integrated to feed the image to the system which helps initiate the program and carry out all the functions. Once the image is successfully captured, the system carries out some basic image processing like binarization and de-noising. The processed image is then supplied to the Neural Network which understands the characters and gTTS is used to convert the text to audio format and read out using software called eSpeak.

Table -1: Summary of Literature Survey

Paper	Description
"Webcam-based Optical Character Recognition using MATLAB" - Dhiraj Kumar Jasrotia, Aarti Malik	Advantages: 1. A text readout system for the visually challenged with a camera that scans the given document converts it to digital text and reads out the text. It works on raspberry pi. 2. 90% success rate achieved. Disadvantages: 1. Does not perform text summarization.
"Optical Character Recognition (OCR) System for Roman Script & English Language using Artificial Neural Network (ANN) Classifier" - Honey Mehta, Sanjay Singla, Aarti Mahajan	Advantages: 1. The non-linear nature of ANN helps to deal with the complex nature of text recognition from scanned images. 2. 98.89% accurate recognition rate is obtained for three different font styles. Disadvantages: 1. Consumes time since many processes are applied and ANN uses many hidden layers. 2. Accuracy specific only to limited font styles
"Extractive Text Summarization using Deep Learning" - Nikhil S. Shirwandkar, Dr Samidha Kulkarni	Advantages: 1. The proposed approach generates short and precise summaries without any irrelevant text. Using features like Sentence-Centroid similarity and thematic words has improved the connectivity of the sentences. 2. Using the proposed method, on an average 88% precision, 80% recall and 84% F measure is obtained
"Python Based Portable Virtual Text Reader" - Hasan U. Zaman, Saif Mahmood, Sadat Hossain, Iftekharul Islam Shovon	Advantages: 1. The results came out positive for most of the sample because TesseractOCR which is a Google API has an accuracy rate of 99% for the English language 2. The Tesseract OCR is able to read font size 12 and above Disadvantages: 1. Font size below 12 and underlined text does not give any output. 2. For certain colours, it is unable to understand the imageand shows no output

3. PROPOSED MODEL

3.1 Model architecture

In order to achieve better domain results, in addition to the existing technique, the proposed model uses Convolutional Neural Network which seeks to inherit the advantages and eliminate the disadvantages.

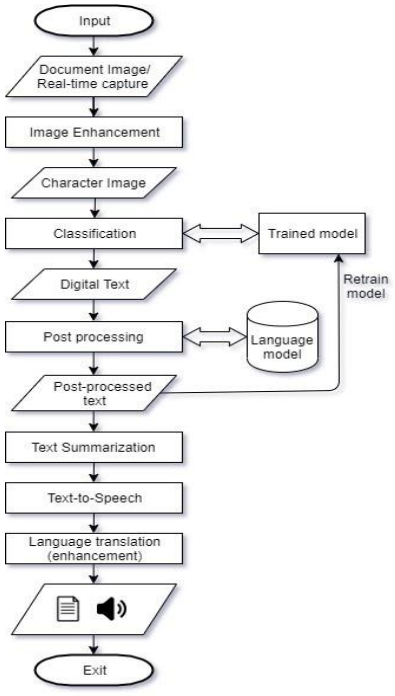

Fig -1: Proposed system

3.2 Algorithm

3.2.1. Image Processing

a. Binarization

Image binarization is the process of taking a grayscale image and converting it to black-and-white, essentially reducing the information contained within the image from 256 shades of grey to 2: black and white.

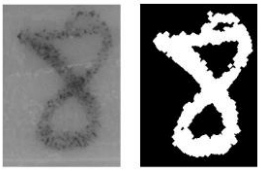

Fig -2: Binarization

b. Skew Correction

Skew detection and correction is one of the first operations to be applied to scanned documents when converting data to a digital format. Its aim is to align an image before processing because text segmentation and recognition methods require properly aligned next lines.

Fig -3: Skew correction

3.2.2 Segmentation

Projection Profiles

Line segmentation consists of slicing a page of text or a zone of interest into its different lines. The main objective of Line level segmentation is to determine the coordinates of lines in an image, which can divide the image into lines. The technique used is Horizontal Projection.

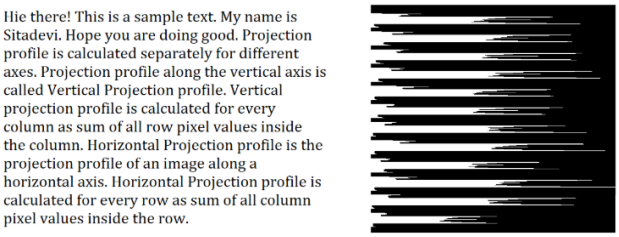

Fig -4: Horizontal Projection Profile

Word segmentation is performed to determine where we have to segment the image to separate out words. This is done using the same logic as for lines but the only difference being Vertical Projection.

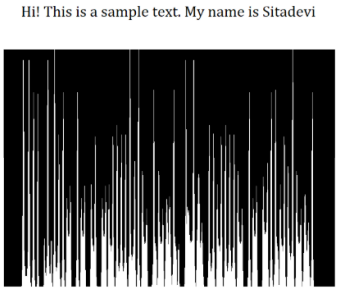

Fig -5: Vertical Projection profile

3.1.3 Training

We have implemented the OCR System using a Convolutional Neural Network consisting of following layers.

1. 2D Convolutional Layer with a 5x5 kernel

2. 2D Max Pooling Layer

3. Dropout with a factor of 20 percent

4. Flatten

5. Dense with l2 regularizer

6. Dense output layer

3.1.4 Summarization

Text summarization is the technique for generating a concise summary of voluminous texts without losing the overall meaning. In extraction-based summarization, following techniques are applied to paraphrase and shorten the original document.
a) Convert the paragraph into sentences: Use the delimiter as a full stop(.)
b) Generate similarity matrix across sentences using cosine distance between word vectors
c) Rank sentences in the similarity matrix
d) Calculate the threshold of the sentences: Usually, the average score is the threshold
e) Sort the rank and pick top sentences that are above the threshold
f) Display the corresponding sentences and save as text document

3.1.5 Text to speech

The gTTS API supports several languages including English, Hindi, Tamil, French, German and many more. The speech can be delivered in any one of the two available audio speeds, fast or slow.

4. PERFORMANCE EVALUATION

4.1 Dataset

The model was trained on a combination of datasets gathered from UCI Machine learning repository and auto generated images using python code. It consists of over 60000 images of dimension 32x32. These were split into training validation and testing data in the ratio 6:1:1.

4.2 Performance

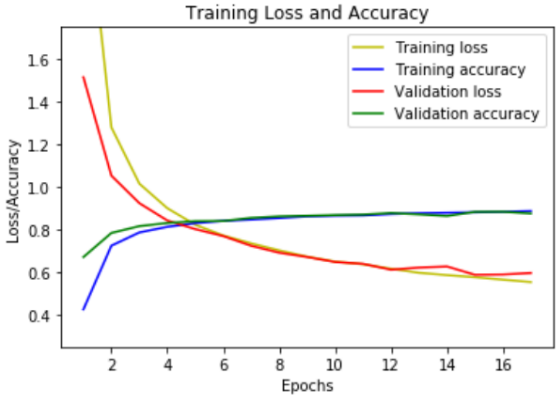

Chart -1: Loss and accuracy curves

The accuracy of training data as well as the validation set has converged and stabilised into a plateau structure. Though there was a difference between the losses during the initial epochs, they reduced in the subsequent epochs and their values were pretty much the same. This stability was also maintained for further epochs.

The accuracy of the character recognition model was 72.48 per cent when the convolutional layer was not added. Adding dense hidden layers led to a spike in the accuracy due to overfitting. This was reduced by introducing a Dropout layer which is used to force the network to drop some features and find other paths. A heavy dropout can cause validation accuracy to be higher than the training accuracy. A regularisation layer was also added to avoid overfitting that had reached 98.02 per cent. After adjusting several parameters such as batch_size, number of epochs, number of filters in the Dense layer, Dropout parameter etc we trained the model towards a decent accuracy of 92.27 per cent and an f1 score of 92.3 per cent. Precision and Recall were each found to be 92 per cent. However, there were four characters that could not be classified properly. We plotted the confusion matrix to explore further.

4.3 Evaluation

It was found from the confusion matrix [8] that classes 0 and 50 i.e digit 0 and the capital letter 'O' were significantly misclassified as the other. Similarly, classes 21 and 44 i.e lowercase letter 'l' and capital letter 'I' were significantly misclassified too. Of all the characters of the Latin Script the most ambiguous ones are zero and ‘O’; ‘l’, ‘1’ and ‘I’; and sometimes ‘g’ and ‘9’ in some font types. However, these misclassifications can be rectified while extracting the text from a word or a sentence with reference to the context by using a language model such as NLTK (python package) or a dictionary.

The summarization module performs well. It calculates the cosine distance between sentences and ranks them accordingly. Hence the generated summary is reproducible. Google’s Text-To-Speech API is a very efficient reader for the English language. It can also translate the text to a few prominent international languages quite efficiently. However, the language translation feature has not been employed in the current implementation of the project.

5. CONCLUSION

The computer vision and digital image processing are fast-growing fields that are essential in many aspects of other areas like multimedia, artificial intelligence, robotics and much more. Image analysis involves the study of segmentation, feature extraction, and classification techniques. Humans interact quite naturally with each other over writing and speech, similarly, human-computer interaction would make things exciting and easier to the user. From the study, it is found that using CNN can significantly improve the efficiency of the model as opposed to using in-built modules in Python such as the widely used KNN algorithm or using the Tesseract OCR tool. The performance measures are described in the previous section. The different standard datasets or variable inputs defined may be used in the implementation for building OCR Systems. We have generated our own database using python for this project. The evaluation metrics to gauge the performance of our model are identified and presented. Using a larger database and language models for post-processing can increase the efficiency of the model.

ACKNOWLEDGEMENT

We would like to take this opportunity to thank our senior friends Rohit Shamdasani and Roshni Ram from IIIT Vadodara for helping us throughout the project. We would also like to thank our Class Coordinator Prof. Rupali Nikhare, our Project Coordinator Prof. K. S. Charumathi, and our HOD Dr Sharvari Govilkar for guiding us towards the successful completion of project work within the given time frame. We would like to express our special thanks of gratitude to our principal Dr Sandeep Joshi who gave us an opportunity to work on this project which helped us learn new concepts and enhance our skills. We extend our sincere appreciation to all our Professors from Pillai College of Engineerirng, Navi Mumbai for their valuable insight and tips during the designing of the project.

REFERENCES

[1] Dhiraj Kumar Jasrotia & Aarti Malik, "Webcam-based Optical Character Recognition using MATLAB," International Journal of Engineering Sciences & Research Technology, 7(8), 216–222. doi: 10.5281/zenodo.1336727

[2] H. Mehta, S. Singla and A. Mahajan, "Optical character recognition (OCR) system for Roman script & English language using Artificial Neural Network (ANN) classifier," 2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS), Bangalore, 2016, pp. 1-5, doi: 10.1109/RAINS.2016.7764379.

[3] N. S. Shirwandkar and S. Kulkarni, "Extractive Text Summarization Using Deep Learning," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697465.

[4] H. U. Zaman, S. Mahmood, S. Hossain and I. I. Shovon, "Python Based Portable Virtual Text Reader," 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), Subang Jaya, Malaysia, 2018, pp. 1-6, doi: 10.1109/ICACCAF.2018.8776778.

[5] N. Ezaki, M. Bulacu and L. Schomaker, "Text detection from natural scene images: towards a system for visually impaired persons," Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., Cambridge, 2004, pp. 683-686 Vol.2.

[6] D. Andrews et al., "A parallel architecture for performing real-time multi-line optical character recognition," 1993 (25th) Southeastern Symposium on System Theory, Tuscaloosa, AL, USA, 1993, pp. 533-536, doi: 10.1109/SSST.1993.522837.

[7] Manliguez, Cinmayii. (2016). Generalized Confusion Matrix for Multiple Classes. 10.13140/RG.2.2.31150.51523 for Confusion Matrix

[8] https://towardsdatascience.com

[9] https://stackoverflow.com