ISSN 2278-3091
Volume 9, No.1, January – February 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse116912020.pdf
https://doi.org/10.30534/ijatcse/2020/116912020

An Approach for Translation of Text in Images using Deep Learning Techniques

Ashwini V. Zadgaonkar [1], Rupali S. Vairagade[2]

[1]Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, 440013, India
zadgaonkarav1@rknec.edu,

[2]Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, 440013, India
vairagaders@rknec.edu

ABSTRACT

Automatically converting text from unknown language to known one is highly desired by frequent travellers moving around the world. In our research work, we tried to build a mobile application for translating text in images from one language to another. Currently, the most common way of translation on mobile devices is typing the words into a dictionary or search engine to find the result. We have designed An Android application for text translation able to extract text from images and perform translation. Our app can automatically recognize texts in simple plain images such as book, sign, and map and gives the extracted text in original language and the translated text in the desired language. We have tested our application on different types of images like signs, book covers, etc.

Key words: Text translation, OCR, Image processing, Deep learning

1. INTRODUCTION

K Satchidandnandan, eminent poet, critic and former Secretary of the Sahitya Academy, India has rightly said: “Translation has helped knit India together as a nation throughout her history. Ideas and concepts like 'Indian literature', 'Indian culture', 'Indian philosophy' and 'Indian knowledge systems' would have been impossible in the absence of translations with their natural integrationist mission.” Translation is an activity, a product, and a process. As an activity, translation is a complex act that requires close reading of a text in the source language, understanding its meaning and creating an equivalent text in the target language. The word “translation” also refers to the product of this activity: the final target language text that will be published or distributed.

Translating the text in an image to a human readable text correctly is one of the interesting research problem. Computer Vision Community had already done a lot of work in this area. The important concept involved in this is OCR – Optical Character Recognition. With the help of the OCR, one can search and recognize the text in electronic documents and convert them into human readable text. It converts electronic documents’ text into corresponding ASCII characters. For the handwritten documents, OCR uses different training models to recognize the characters in it and resolved it to highest accuracy. There are variety of technologies used for text analysis in images like OCR (Optical Character Recognition), Tesseract OCR Engine which is Google’s OCR Engine which is far more powerful, Google Neural Machine Translator which is Google’s Neural Networking bases translator.

OCR stands for Optical Character Recognition [1,2]. It is a mechanism that can convert text in an electrical document or a scanned written document into human readable text. It scans the text of the image character-by-character, analyses the image and then converts into the respectable ASCII Character Code. Most of the OCR devices have optical scanner for scanning the text and then analysis it through the OCR and generating an editable document of the scanned image. This mechanism makes it possible for the people to edit the text, search for different kinds of words or phrases, display or print a copy of the machine generated code and used it to convert text- to-speech or convert the given text to any required language through Google’s Translation API. If the input is an electronic document, then it extracts the text from the image [3,4] and converts it into ASCII and converts it into English Language. But, if the input is a handwritten document, then the OCR takes the help of the database which contains the scanned characters of hand written texts to convert it into specific ASCII Code[5]. There are lots of applications using OCR, especially for handwritten text images [6].

The Tesseract OCR is an OCR Engine using is an Open Source library developed by Google [7] which it uses to study the images and recognize more than100 languages. It uses automatic machine learning technology to understand how words and characters work in day to day life. The Tesseract OCR works in a step-by-step manner where it goes through rigorous process of Adaptive Binarization, Component Analysis, Words and Paragraph Detection followed by a Two Step Recognition, call as Recognition Phase 1 and Recognition Phase 2. From there output will be generated as an Editable Document of the Scanned Text Image. It first takes the original text image as an input and then apply Adaptive Binarization which converts the given image into a Binary. The real power of the Tesseract OCR Engine is to be able to identify any language with 100% accuracy in the electronic document. And if there are few errors in the document, the engine will be able to scan the character to it's highest accuracy possible.

Though Tesseract OCR Engine can accurately convert any document but the real requirement is to translate the text the Engine has scanned into any language the user has requested. Till date, the best translators present are backed by the Google. It is a free multi- lingual machine translation [8] service developed by Google, to translate text. It offers a web- site interface, mobile apps for Android and iOS, and an API that helps developers build browser extensions and software applications. Google Translate supports over 100 languages at various levels and as of May 2017, serves over 500 million people daily. Rather than translating languages directly, it first translates text to English and then to the target language. During a translation, it looks for patterns in millions of documents to help decide on the best translation. It is actually called Google Neural Machine Translator (GNMT) [9] which is a neural machine translation (NMT) system developed by Google and introduced in November 2016, that uses an artificial neural network to increase fluency and accuracy in Google Translate.

2. PROPOSED SYSTEM

Our objective behind this work is to provide a way to translate textual information in an image from native language to the language understood by the user using image processing and deep learning methods.

Proposed system methodology works as follows.

- Accept textual image from user needed to be translated.
- Perform feature extraction to extract text.
- User will select the language in which the text has to be translated.
- Translate the text in given language provided by user and display the language.
- Display the translated output.

Proposed system for text translation from images goes through following pre-processing phases.

2.1 Scaling

OpenCV comes with a function cv.resize() for scaling. The size of the image can be specified manually or by specifying the scaling factor. Different interpolation methods like cv.INTER_AREA() for shrinking and cv.INTER_CUBIC (slow) cv.INTER_LINEAR() for zooming. By default, interpolation method used is cv.INTER_LINEAR() for all resizing purposes.

2.2 Convert to Gray:

We will be using cvtColor(image, gray_image ,CV_BGR2GRAY) to convert image to grayscale in order to clear unwanted noise in the background of the image

2.3 Binarization

In this cyber-reality where everything eventually boils down to 1's and 0's, converting image to black and white immensely helps Tesseract recognize characters. However, this might fail if input documents lack contrast or have a slightly darker background.

2.4 Median Blurring

Median blurring is a non-linear filter. Unlike linear filters, median blurring replaces the pixel values with the median value available in the neighborhood values. So, median blurring preserves edges as the median value must be the value of one of neighboring pixels. Here, we will use function cv.medianBlur() which takes median of all the pixels under kernel area and central element is replaced with this median value. This is highly effective against salt-and-pepper noise in the images. In contrast linear blurring , in median blurring, central element is always replaced by some pixel value in the image. It reduces the noise effectively.

2.5 Bilateral Filtering

Bilateral filtering is useful for removing the noise without smoothing the edges. Similar to Gaussian blurring, bilateral filtering also uses a Gaussian filter to find the Gaussian weighted average in the neighborhood. However, it also takes pixel difference into account while blurring the nearby pixels. Thus, it ensures only those pixels with similar intensity to the central pixel are an easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it

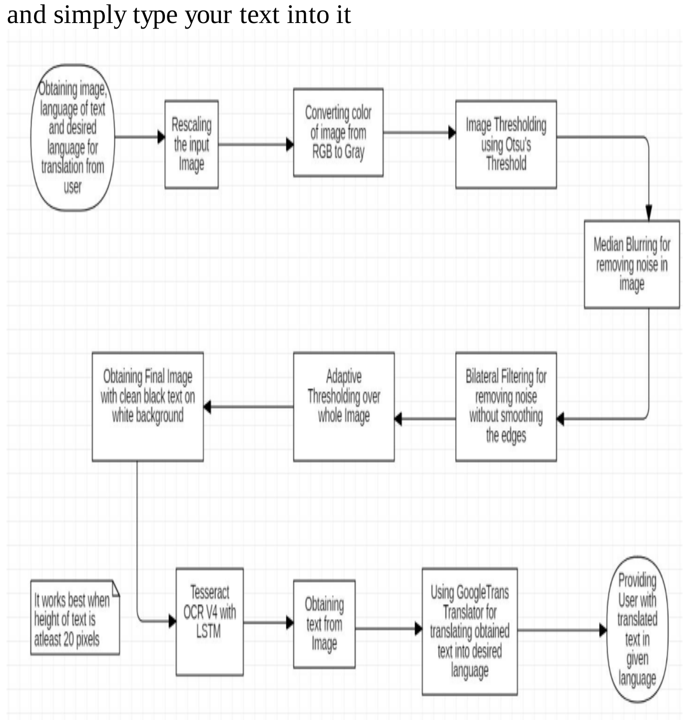

Figure 1: Proposed system architecture

3. IMPLEMENTATION

For labelling the training samples using Tesseract we have taken help of a tool named bbTesseract. To generate the training files for a specific user, we need to prepare the box files for each training images using the following command:

tesseract fontfile.tif fontfile batch.nochop makebox

Figure 2: shows a screenshot of the bbTesseract tool, used for labelling the training set.

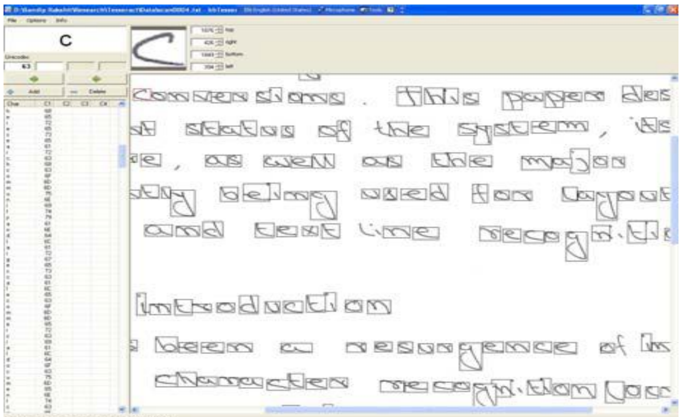

Figure 3: JTesserbox Example

The training process, will create a new language set. Then Tesseract will again use the newly created language set to label the rest of the box files corresponding to the remaining training images. The output of this step is fontfile.tr which contains the features of each character of the training page. The character shape features can be clustered using the mftraining and cntraining programs:

mftraining fontfile_1.tr fontfile_2.tr ...

Tesseract uses 3 dictionary files for each language. Two of the files are coded as a Directed Acyclic Word Graph (DAWG), and the other is a plain UTF-8 text file. To make the DAWG dictionary files a wordlist is required for our language. The wordlist is formatted as a UTF-8 text file with one word per line.

The corresponding command is:

wordlist2dawg frequent_words_list freq-dawg

wordlist2dawg words_list word-dawg

Now we have to collect all the 8 files and rename them with a lang. prefix, where lang is the 3-letter code for our language and put them in our tessdata directory. Tesseract can then recognize text in our language using the command:

tesseract image.tif output -l lang

These are some of the sample outputs obtained from the system extracting textual information from the input images given to the system, we have achieved the accuracy 97.9% after extracting text from images.

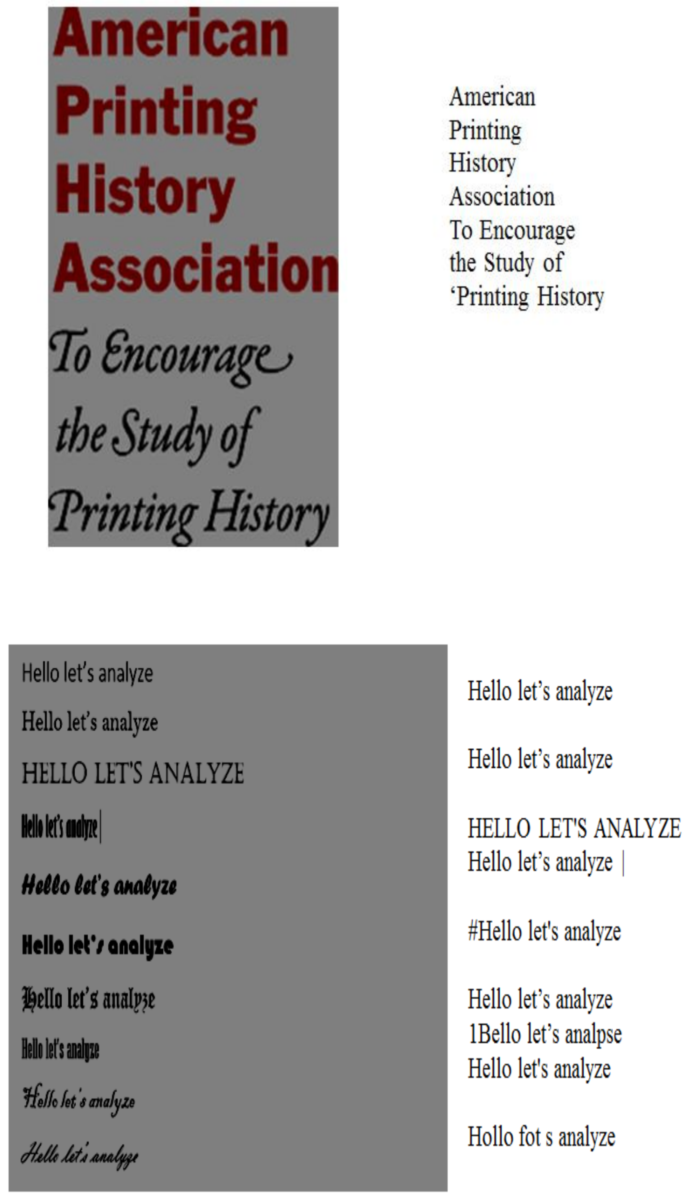

Figure 4: Extraction of Text in English languages

Figure 5: Sample text extractions in different languages

4. CONCLUSION & FUTURE SCOPE

By combining the different technologies, we try to propose an approach to overcome communication barrier using mobile application for text translation in images. The main purpose of work is to identify and translate text in just a snap of a finger. We have succeeded in obtaining textual images from user for any language and translating to user desired language. Though Tesseract identifies just printed text we have tried to train it for handwritten characters. We have trained the Tesseract for different local languages which improves the overall accuracy of the system. Using Text-to-Speech powered by the Tech Giant GOOGLE INC., we can extend our work to make the text speak out after the translation has been done. This method will help lot of people with no vision to overcome communication barriers worldwide.

There is a need for translation in diverse fields such as education, mass communication, politics, science and technology, literature, tourism, religion, trade and business, etc. Today, the translation industry includes a multitude of companies providing services such as Translating written material and paper-based documents ,Interpreting services and Sign-language services , Digital documentation translation, Software translation and website translation (localization)

Some of the potential applications of text translation are discussed below.

- Cultural Interchange
- Accurate News Translation
- In Tourism Promotion
- As a Business communication tool
- Efficient handling of Nation's External Affairs

REFERENCES

1. Karez Abdulwahhab Hamad and Mehmet Kaya. **A Detailed Analysis of Optical Character Recognition Technology**, *IJAMEC, 2016, 4(Special Issue),* pp 244–249 https://doi.org/10.18100/ijamec.270374
2. George Nagy, Thomas A. Nartkerb, Stephen V. Rice. **Optical Character Recognition: An illustrated guide to the frontier**, *Proceedings of SPIE - The International Society for Optical Engineer- ing1999,*pp 3967:58-69
3. Ayatullah Faruk Mollah, Nabamita Majumder, Subhadip Basu and Mita Nasipuri. **Design of an Optical Character Recognition System for Camera based Handheld Devices.** *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 4, No 1, July 2011 ,pp 283289
4. Ali Mir Arif Mir Asif, Shaikh Abdul Hannan, Yusuf Perwej, Mane Arjun Vithalrao. **An Overview And Applications Of Optical Character Recognition**,. *International Journal of Advance Research In Science And Engineering, IJARSE*, Vol. No.3,No.7, July 2014, pp 261-274
5. Jacob George, Varghese George, A. Deepan, Joshua Ninan Valson .**Optical Character Recognition Based Hand-Held Device for Printed Text to Braille**

Conversion, *2011 3rd International Conference on Machine Learning and Computing (ICMLC 2011),*pp 172-175

6. G. Nagy, Yihong Xu. **Automatic prototype extraction for adaptive OCR.** *Conference: Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on,* Volume: 1
7. Ray Smith. **An Overview of the Tesseract OCR Engine** . *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)* https://doi.org/10.1109/ICDAR.2007.4376991
8. Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. **Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation**. *https://arxiv.org/abs/1609.08144*
9. Melvin Johnson, Mike Schuster, Quoc. **Google's Multilingual Neural Machine Translation System**: EnablingZero-ShotTranslation. *https://arxiv.org/abs/1611.04558*
10. Fredrick Lundh. **Python Imaging Library Hand Book.** http://effbot.org/imagingbook/pilindex.htm
11. G. Nagy, Yihong Xu. **Automatic prototype extraction for adaptive OCR.** *Conference on Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on,* Volume: 1.