



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 9, September 2020

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Implementation of Text Recognition Using Image Processing & Translation

Mayur Pabalkar, Saket Kolte, Meraj Ansari

Dept. of Computer Engineering, Sinhgad Academy of Engineering, Pune, Maharashtra, India

ABSTRACT: Recognition of scene text is a difficult issue contrasted with the recognition of printed documents. An approach is proposed to recognize text in complex background natural scene, word formation from recognized text and word translation into user defined language. The proposed approach is strong to various types of text appearances, including text dimension, text style, shading, and foundation. Joining the particular qualities of various correlative strategies and conquering their deficiencies, the proposed strategy utilizes effective character location and limitation procedure and multiclass classifier to perceive the content precisely. The text will be extracted and will be translated into user defined language. Along with that, the translated text will be converted into speech which will be helpful for the visually impaired people. The proposed approach successfully recognizes text on natural scene images and does not depend on a particular alphabet, text background. It works with a wide assortment in size of characters and can deal with up to 20 degree skewness productively.

KEYWORDS: Image Processing, Optimal Character Recognition, Recurrent Neural Networks, Text Recognition, Text Extraction, Text Extraction

I. INTRODUCTION

Text recognition is a challenging and intricate process due to the often bad quality of images, different backgrounds or different fonts, colors, sizes of texts and hence there is still enough room of improvement in text recognition techniques which plays a vital role in improving the overall performance of the text recognition system. Scene images are characterized by complex background, perspective distortion, low resolution and poor quality. Scene text can appear with any slant, tilt, in any light and upon any surface and hence hard to detect, localize and recognize. There are a number of pre-processing steps followed by the actual recognition in the implementation of text recognition system. The process of character recognition of any script can be broadly broken down into three stages; text extraction, classification, post-processing. Typical text extraction includes a collection of operations that apply successive transformations on an image. It takes in a raw image, removes noise, distortion, skewness and detect text segment by applying various segmentation & connected component analysis and feature extraction techniques. The selection of a stable and representative set of features is the heart of pattern recognition system design. Perhaps the most consequential one is the selection of the type and set of features among the different design issues involved in building a text recognition system. The grouping stage is the fundamental dynamic phase of a book acknowledgment framework and utilizes the highlights separated in the past stage to recognize the content fragment as per preset principles. The post-preparing stage, which is the last stage, improves acknowledgment by refining the choices taken by the past stage and perceives words by utilizing setting. It is at last answerable for yielding the best arrangement and is regularly executed as a lot of procedures that depend on character frequencies, vocabularies, and other setting data.

II. PROBLEM DEFINITION

Recognition of Scene text is a challenging problem as compared to the recognition of printed documents and then translating it and converting it into speech is a tedious task. To overcome these issues, we have built a System that will help in detecting and extracting real time text based image using Optical Character Recognition, translate it into user defined language and convert it into Speech using Android.

III. MOTIVATION

While visiting foreign countries for tourism or for business meetings, people find it difficult to read & understand the local languages. Using this methodology, it will become easier for them to read ,understand the text written in native languages and via speech text can be audible for user if they cant read.

IV. LITERATURE SURVEY

Grover S. et.al [1] state that location of text from archives in which text is installed in complex shaded report pictures is a difficult issue. There are a ton of likely employments of text extraction in picture looking, chronicling reports and so forth. In this paper, we propose a straightforward edge based element to play out this assignment. It targets distinguishing literary areas from the archive and isolating it from the designs divide. The calculation depends on the sharp edges of the characters which are missing in pictures. We discover these edges and use them to arrange text from pictures. This edge data can likewise be utilized for other picture understanding undertakings.

Zhang, Y., & Tan, K. K [2] proposed that system presents the advancement of a human-machine intuitive programming application, explicitly helpful for text extraction from pictures, which are caught utilizing versatile and computerized gadgets with cameras. This product is the central aspect of a full application, which is being created to decipher the text in such a caught picture to another dialect. The full application will involve two phases: a handling stage (which is the piece of worry in this paper) and an acknowledgment stage. In the preparing stage, an insightful and versatile calculation will yield the basic data from the crude picture. This system will zero in mostly on the improvement of the preparing stage. In the acknowledgment stage, the separated content will be deciphered and interpreted through an optical character recognition (OCR) motor, which is broadly accessible. The adequacy of the proposed calculation in meeting the difficulties behind the handling of such pictures will be featured with genuine pictures.

Chidiac N. et.al [3] introducing a hearty calculation that identifies text from regular scene pictures and concentrates them paying little mind to the direction is proposed. All current strategies are intended to work under a certain requirement, such as identifying text just one way. Maximally Stable Extremal Regions (MSER) indicator is picked to remove twofold districts since it has demonstrated to be powerful to lighting conditions. An improvement strategy for MSER pictures is intended to get clear letter limits. Pictures are then taken care of into a Stroke Width Detector and a few heuristics are applied to eliminate non-text pixels. Subsequently, recognized content locales are taken care of into an Optical Character Recognition module and afterward sifted as indicated by their certainty measure. The acknowledgment of characters isn't important for the calculation and the outcomes are as it were about the recognition of text. Our calculation end up being viable on obscured pictures and boisterous pictures also, in light of both emotional and target assessments.

Jie Yang et.al[4] state that A lot of data are implanted in characteristic scenes. Signs are genuine instances of common articles with high data content. In this paper, we examine issues in programmed recognition and interpretation of text from common scenes. We depict the challenges of programmed text discovery and propose techniques to address these challenges. We expand model based machine interpretation innovation for sign interpretation and present a model framework for Chinese sign interpretation. This framework is equipped for catching pictures, naturally identifying and perceiving text, and making an interpretation of the content into English. The interpretation can be shown on a palm size PDA, or blended as a voice yield message over the headphones

Lane, R., & Bansal, A [5] proposed Machine translation(MT) between regular dialects is a scandalously troublesome issue in Natural Language Processing that is still particularly being investigated. This exploration study investigates the viability of building up an versatile interpreter utilizing Lexical Functional Grammars. The fundamental exploration objective is building a machine interpreter generator for multilingual correspondence, for example building up a framework whose information sources are phonetic depictions of an ideal source and target language and whose yield is a program that deciphers between the two common dialects. A bidirectional machine interpreter among English and Hungarian, created as a proof-of-idea contextual investigation, is talked about. The advantages and downsides of this methodology as summed up to MT frameworks are additionally talked about, alongside potential zones of future work.

V. ALGORITHM

1. Optical Character Recognition (OCR)

Optical character recognition is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image.



Input:- Image

Output:- OCR result

1. Loading any image format (bmp, jpg, png) from given source. Then convert the image to grayscale and binarize it using the threshold value (Otsu algorithm).
2. Detecting image features like resolution and inversion. So that we can finally convert it to a straightened image for further processing. (completed the code of rotation of Image but not able to detect Image angle about which we have to rotate the Image, So still working on angle detection part).
3. Lines detection and removing. This step is required to improve page layout analysis, to achieve better recognition quality for underlined text, to detect tables, etc. (Decided To Complete that part in End)
4. Page layout analysis. In this step I am trying to identify the text zones present in the image. So that only that portion is used for recognition and rest of the region is left out.
5. Detection of text lines and words. Here we also need to take care of different font sizes and small spaces between words.
6. Recognition of characters. This is the main algorithm of OCR; an image of every character must be converted to appropriate character code. Sometimes this algorithm produces several character codes for uncertain images. For instance, recognition of the image of "I" character can produce "I", "|", "1", "l" codes and the final character code will be selected later.
7. Saving results and convert to language and so on.

2. Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNN) is a type of Neural Network where the output from previous steps are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence. RNN have a "memory" which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.

Input:- Parameter of bias

Output:- Recurrent layer

1. RNN converts the independent activations into dependent activations by providing the same weights and biases to all the layers, thus reducing the complexity of increasing parameters and memorizing each previous outputs by giving each output as input to the next hidden layer.
2. Hence these three layers can be joined together such that the weights and bias of all the hidden layers is the same, into a single recurrent layer.

3. Image Processing

Read and Write Images. Detection of images and its features. Detection of shapes like Circle, rectangle etc in a image, Detection of coin in images. Text recognition in images. e.g. Reading Number Plates. Modifying image quality and colors.

Input:- Image

Output:- Text recognition in images

Image processing mainly include the following steps:

1. Importing the image via image acquisition tools;
2. Analyzing and manipulating the image;
3. Output in which result can be altered image or a report which is based on analyzing that image.



PHASES OF IMAGE PROCESSING:-

- 1.ACQUISITION**– It could be as simple as being given an image which is in digital form. The main work involves: a) Scaling b) Color conversion(RGB to Gray or vice-versa)
- 2.IMAGE ENHANCEMENT**– It is amongst the simplest and most appealing in areas of Image Processing it is also used to extract some hidden details from an image and is subjective.
- 3.IMAGE RESTORATION**– It also deals with appealing of an image but it is objective(Restoration is based on mathematical or probabilistic model or image degradation).
- 4.COLOR IMAGE PROCESSING**– It deals with pseudo color and full color image processing color models are applicable to digital image processing.
- 5. WAVELETS AND MULTI-RESOLUTION PROCESSING**– It is foundation of representing images in various degrees.
- 6. IMAGE COMPRESSION**-It involves in developing some functions to perform this operation. It mainly deals with image size or resolution.
- 7. MORPHOLOGICAL PROCESSING**-It deals with tools for extracting image components that are useful in the representation & description of shape.
- 8. SEGMENTATION PROCEDURE**-It includes partitioning an image into its constituent parts or objects. Autonomous segmentation is the most difficult task in Image Processing.
- 9. REPRESENTATION & DESCRIPTION**-It follows output of segmentation stage, choosing a representation is only the part of solution for transforming raw data into processed data.
- 10.OBJECT DETECTION AND RECOGNITION**-It is a process that assigns a label to an object based on its descriptor.

VI. PROPOSED SYSTEM APPROACH

In a proposed system, we are proposing real time images of any natural text by text recognition. Proposed system consists of mainly following modules Text recognition, extraction and translation. Image has been used as one of the mainstream manners for text recognition. However, with the popularity of this system, it takes only a few seconds to extract the text and translate. In our system we translated text into user preferred language by using Text Recognition, Text Extraction, Text Translation and Text to Speech.

In our system mainly consist of 3 basic module. i.e. OCR, Text Translation and Text To Speech

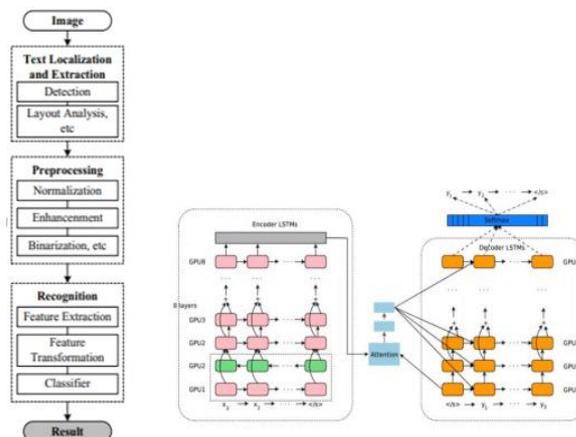


Fig.1 Block Architecture Diagram of Proposed System



We are going to develop following modules in our system:

1. OCR:

1.1 Text Recognition : - Text recognition can automate tedious data entry for credit cards, receipts, and business cards. With the Cloud-based API, you can also extract text from pictures of documents, which you can use to increase accessibility or translate documents. Apps can even keep track of real-world objects, such as by reading the numbers on trains.

1.2 Image Processing :- After getting face images by using frontal face XML only faces get cropped and further used to process or deform.

1.3 Text Extraction :- The Text Extraction process generates text from electronic documents that have associated text. Its functionality is similar to using the "Generate Searchable Text" command on an electronic document in the Laser fiche Client. Once text has been extracted from a document, it can be sent along with the document to the repository, making the document full-text searchable in Laser fiche.

2. Text Translation:-

Translation is the communication of the meaning of a source language text by means of an equivalent target-language text.[1] The English language draws a terminological distinction (not all languages do) between translating (a written text) and interpreting (oral or sign-language communication between users of different languages); under this distinction, translation can begin only after the appearance of writing within a language community.

3. Text To Speech:-

Text-to-speech (TTS) is a type of assistive technology that reads digital text aloud. With a click of a button or the touch of a finger, TTS can take words on a computer or other digital device and convert them into audio.

VII. COMPARATIVE RESULTS

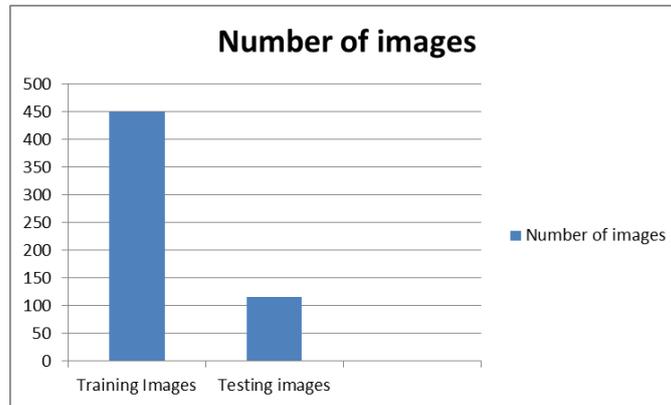
In our experimental setup, as shown in table 6.1, shows dataset of training images and testing images. below table 6.1 shows that 450 training images and 115 testing images.

Sr. No	Number of training images	Number of testing images
1	450	115

Table1 6.1: Number of images

VIII. RESULTS

From above data, as shown shows dataset of training images and testing images. Below graph 7.1 shows that 450 training images and 115 testing images.



Graph 7.1: Number of images

IX. CONCLUSION

We are going to develop real time text recognition and translation based on Recurrent Neural Networks and Optimal Character Recognition. We are going to generate real time text translation and the translated output will be converted into Speech as the final output using android libraries. While there are certainly still some minor issues to overcome, we feel we made important progress on the challenging limitations of neural network depends on image processing.

ACKNOWLEDGMENT

This work is supported in a text recognition using image processing & translation system of any state in India. Authors are thankful to Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, Pune for providing the facility to carry out the research work.

REFERENCES

- [1] Grover, S., Arora, K., & Mitra, S. K. (2009). "Text Extraction from Document Images Using Edge Information." Annual IEEE India Conference 2009
- [2] Zhang, Y., & Tan, K. K. "Text extraction from images captured via mobile and digital devices." International Journal of Computational Vision and Robotics, 2009
- [3] Chidiac, N.-M., Damien, P., & Yaacoub, C. "A robust algorithm for text extraction from images" 39th International Conference on Telecommunications and Signal Processing (TSP), 2016
- [4] Jie Yang ; Xilin Chen ; Jing Zhang ; Ying Zhang ; Alex Waibel "Automatic detection and translation of text from natural scenes" IEEE International Conference on Acoustics, Speech, and Signal Processing 2002
- [5] Lane, R., & Bansal, A. "An Adaptive Machine Translator for Multilingual Communication." 2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). 2017



INNO SPACE
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details