ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse100922020.pdf
https://doi.org/10.30534/ijatcse/2020/100922020

Text Detection and Recognition Techniques from Images using React Native

Rutuja Kherde[1], Jignyasa Sanghavi[2]
[1]M. Tech Scholar, Shri Ramdeobaba College of Engineering & Management, Nagpur, India,
kherderk@rknec.edu
[2]Assistant Professor, Shri Ramdeobaba College of Engineering & Management, Nagpur, India,
sanghavijb1@rknec.edu

ABSTRACT

Text detection and recognition from images becomes very crucial part in many of the Computer Vision problems. It becomes one of the modes for concluding on some knowledge-based decision while working in domains of natural language processing, robotics and artificial intelligence. Due to the usage of modern tools for image acquisition like surveillance cameras, mobile cameras, etc. and social website posts has gathered heaps of image data. Textual information can be extracted from images and used for further classifications or decision making. However, research on Text detection and recognition has been carried out from two decades. In this paper we tried to give brief of the research work done in this field. The paper shows various comparisons between the methods used by researchers for text detection and recognition. It also highlights the various applications in combination with GPRS, hardware, etc. used now a days using text detection and classification.

Key words: React Native, tesseract-ocr, text detection, text recognition.

1. INTRODUCTION

In era of 21st century, technology has changed drastically, world is now working on finger tips with the help of technical gadgets and internet. By the shake of technology, documents such as personal documents, newspaper, books, official documents etc. can be access anywhere and anytime with the help of technology. On visit to other countries, communication problem or language problem becomes the major problem. If the hoardings and sign boards are in their language then it becomes difficult to travel in that place. Translation medium is required to convert their language to understandable language. Google Translate have tried to solve this problem by taking the input image and translating to understandable language. The main area of research is to detect the text from image and recognize it for translation. Some of the major issues involved in this work can be complex background, multi-orientation, resolutions of captured image, noise, etc.

The basic working model for Text detection and recognition from the given image is described in figure 1. The work flow of the given model starts taking image as input, applying some preprocessing techniques to enhance the quality of image and then finding the region of interest, it is nothing but text region by applying some methodologies. After finding text region, segmentation of that region is performed for character extraction and then recognition. Characters are extracted separately which is given as input to character recognition algorithm and then consolidated to form a word.

Steps of text detection and recognition process:

1. Preprocessing: The input image may contain some disturbance or noise, to eliminate the noise and to enhance the quality of image we apply preprocessing techniques. Some of the techniques used for this are min filters, max filters, mean filter, median filter, Gaussian filter etc. For adjusting the alignments slant and rotation normalization is used to obtain characters of uniform size.
2. Text region detection: In this step region of interest is detected using various techniques and filters. The text areas are detected from the image for further processing.
3. Segmentation: The text region detected in previous step is segmented from the image so as to perform extraction and recognition operation on text areas only.
4. Character extraction: After segmentation all text characters are extracted from image.
5. Character recognition: After performing character extraction, characters are recognized using different methods which can further be used for the whole word recognition.

2. RELEATED WORK

Text detection and recognition is trending topic in many domains like artificial intelligence, robotics, smart city development, etc. Some of the researchers have proposed different methodologies and frameworks for Text detection and recognition. The detail review of some of key points in their research is illustrated in this section.

Savita Choudhary, Nikhil Kumar Singh, Sanjay Chichadwani [1] used Maximally Stable Extremal Region (MSER) for text area detection and self-trained Neural Network for text recognition from natural images. The authors performed the image resize, conversion to gray scale, removal of noise and conversion to binary image pre-processing stage. Researchers used MSER to detect the character region. In that case Canny Edge Detection algorithm is used in combination with MSER which detects both text and non-text regions. For character recognition, character classifier model based on Convolution neural network is used. The network takes the input of 28x28 with four network layers as 2D convolution layer, Max Pooling layer, dropout layer and flattening layer.

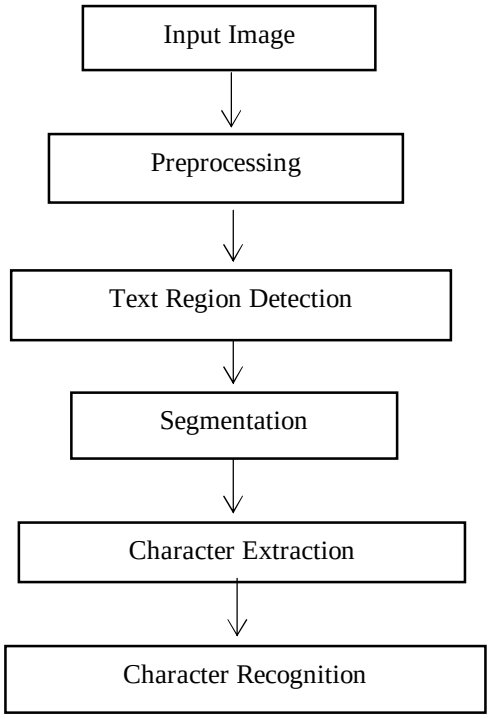

Figure 1: Flow diagram of text detection and recognition process

Annmaria Cherian, Sanju Sebastian [2] proposed work on prospectively distorted text images. For correcting the orientation of distorted text Hough transform was used in their research. Text localization was used to detect the text from image. The connected component analysis is used for segmentation which produced text and non-text regions. Finally, the support vector machine is used for classifying the text region. Text recognition is carried out using Optical Character Recognition (OCR).

Veena Rajan, Shani Raj [3] used fractional poission model. In preprocessing, first researchers converted color image into gray scale then Laplacian filter applied to remove the noise and produced sharp image. After that binarization performed on image. Researchers used MSER for text region detection. Image segmentation performed by morphological operations where square structuring element used. For character extraction researchers used bounding box.

Vyankatesh V. Rampurkar, Sahil K. Shah, Gyankamal J. Chhajed, Sanjay Kumar Biswash [4] used thresholding to convert grayscale image to binary image. Segmentation carried out by using morphological operations. Researchers did character detection by connected component analysis. Text extraction did by adjacent character grouping. Researchers used text line grouping for locating arbitrary orientation.

Shangxuan Tian, Yifeng Pan, Chang Huang, Shijian Lu, Kai Yu, and Chew Lim Tan [5] proposed a unified scene text detection system (Text Flow). Text Flow consist of two parts: Character candidate detection performed by fast cascade boosting algorithm and sliding window scheme. Researchers used Cascade boosting method to detect whole character instead of separate character which is consider as negative sample in training process which become helpful when in some language such as Chinese where character consist of multiple components and sliding window approach used for capturing text shape related information which is of high level such as distinct intensity and gradient distribution, character stroke boundary. For text line extraction researchers used min-cost flow network. It consisted of node and sink. Nodes are nothing but detected character candidate and flow is text line.

Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet [6] proposed work on recognition of multi-digit number which was present on street view imagery. For recognition they used deep convolution neural network that directly worked on pixel level of image. Localization, segmentation, and recognition, all these steps performed by single deep convolutional neural network. Here Researchers used eleven hidden layers to produced best result on high quality images. Researchers also employed the DistBelief implementation to train distributed and large neural network on high quality image.

Junyu Gao, Qi Wang, Yuan Yuan [7] worked on multi-orientation text detection. Researchers adapted novel convolutional regression network (CRN). This network made up of two parts, text locator and region proposal extractor. CRN consisted of HDM (Hierarchical Deconvolution Module) and TGM (Text-line and Geometry segmentation module) which used to segment multi-oriented image and CRM (Classification and Regression module) which adapted to localized final image. This structure trained to detect multi-oriented text.

Obadiah Lam, Feras Dayoub, Ruth Schulz, Peter Corke [8] proposed an approach on text recognition for indoor robotics. Researchers used MSER for text detection, Stroke width transform for character segmentation and Convolutional Neural Network for text recognition. CNN consist of four layers, two convolutional layers and two connected layers. Researchers trained CNN for 25 iterations. Output of the recognition phase was text region. In the sense of robotics, Researchers used Tesseract OCR engine. Images produced by robot was in distorted form. To deal with this problem

Tesseract OCR engine used. Researchers manually cropped image and send it to the Tesseract OCR engine. Researchers also used Google Goggles for web search. This was an image-based search-based app. It was useful for long string of text consist of multiple words.

Ling Hu and Quang Ni [9] worked on object detection algorithm for smart cities. Here, Researchers first converted RBG image into gray scale. After that gradient of the gray scale image calculated. Researchers implemented new filter for correlation coefficient calculation for every pixel of image of car. After applying filter new image with car license plate produced. For this output, first filter applied in horizontal direction, it eliminated noise in horizontal direction, then horizontal axis scan from left to right and cut area from original image. Similarly, applied filter in perpendicular direction, scan perpendicular axis. Their approach is useful for IoT and smart city applications. It reduced data volume generated in smart cities in urban surveillance system.

R. Komalapriya, A. Sangeetha Shiny [10] also worked on object detection algorithm but by using different approach. Researchers used OpenCV for accessing video input in python. In preprocessing they first converted input image into gray scale, to detect multiple number of plate area using edge detection techniques. After that scan high intensity pixels from image from left to right in x-axis and top to bottom on y-axis. Text present in the image is variables, hence we get more region of interest. If area of number plate is identified cropped it and then with the help of OCR image is converted into text.

D. L. Vala, Umesh Baria, Urvi Bhagat, Mohan Khambalkar [11] made project on optical character recognition robot. Project made up of two parts, software and hardware part. Software part processed the image such as conversion of image into gray scale, extract region of interest and OCR processing. In hardware part, raspberry pi-controlled robot circuit. Researchers made a wireless robot controlled by Wi-Fi adapter with the help of IoT and USB camera. Raspberry pi has L293D IC.

Prof. Suneel K Nagavi, Prof. Mahesh S Gothe, Prof. Praveen S. Totiger [12] provided different approach to navigate robot without help of human. Researchers mounted the camera on robot. GPRS helped to communicate between robot and server or PC, cellphone and robot communicated through Bluetooth. When robot crosses any RF card, it stopped and informed mobile to captured the image through Bluetooth. After capturing image, they send image to the server for further process through GPRS. Sever receive the image and by OCR converted into data or text format.

Vinutha M Ha, Sweatha K Na and Sreepriya Kurupa [13] proposed system which continuously monitors to check data was available or not. If data was available, controller informed the mobile to for taking image. Then image send to the server through GPRS, server send the instruction to the mobile then mobile send this instruction to the robot.

Shangxuan Tian, Shijian Lu, and Chongshou Li [14] proposed weakly supervised learning framework. Researchers made this framework to showed the problem related to the character level annotation. Researchers trained the character decorator by word annotation. Here, mostly two steps were iterated, character center mask and character model. Character model was useful for to extract the text in challenging environment. It also produced best result in some condition such as deformed text line and mathematical expression recognition.

Loay Alzubaidi, Ghazanfar Latif [15] proposed work on Saudi license plate detection with the help of raspberry pi. Researchers proposed system consist of software and hardware parts. Firstly, if any vehicle is coming toward gate then sensor present on raspberry pi send message to the camera to captured the image of vehicle license plate. After capturing the image, it processed by using K-nearest neighbor classifier (KNN). KNN is used for character detection from license plate. After that, license plate data compared with data present in database. If it is matched, then system open the gate.

Sheetal S. Pandya and Nilesh B. Kalani [16] proposed review work on Deep Neural Network model. Researchers did research on Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) both are main component of Deep Neural Network. After research, researchers found that CNN worked accurately on text and sentimental classification techniques and RNN worked best for sequence modelling and large-scaled dependency tasks.

Manal Abdullah, Afnan Agal, Marium Alharthi, Marium Alrashidi [17] proposed work on Arabic handwriting recognition by using Neural Network. Researchers did prepreprocess by removing space between words, thinning and bold and baseline estimation and correction. Feature extraction did by using MATLAB software.

For text detection and recognition, researchers applied different techniques on images. They gathered different images from different places and made datasets of them. Some datasets used are shown in table 1 and techniques applied are shown in table 2.

Table 1: Datasets used

Sr No.	Author's Name	Datasets	Number of Images
1.	Shangxuan Tian, Nikhil Kumar Singh, Sanjay Chichadwani	ICDAR2011, ICDAR2013, Multilingual dataset	ICDAR2011: 229 training images and 255 testing images ICDAR2013: 229 training images and 233 testing images Multilingual dataset: 248 training images and 239 testing images

2.	Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet	SVHN dataset	200k street numbers
3.	Annmaria Cherian, Sanju Sebastian	Scene text Perspective	200 images
4.	Junyu Gao, Qi Wang, Yuan Yuan	ICDAR2013, ICDAR2015, MSRA TD500	ICDAR2013: 229 natural training images and 233 testing images ICDAR2015: 1,000 training images and 500 testing images MSRA TD500: 300 training images and 200 testing images
5.	Shangxuan Tian, Shijian Lu, and Chongshou Li	ICDAR 2013, FORU, COCO-Text, SWT	ICDAR 2013: 229 training image and 233 testing images FORU: 1162 images and 14888 annotated characters COCO-Text: 14712 images SWT: 307 images

Table 2: Comparative study on different techniques

Sr no.	Author's Name	Techniques	Advantages	Accuracy
1.	Savita Choudhary, Nikhil Kumar Singh, Sanjay Chichadwani	Text region extraction: MSER Character recognition : CNN	Very less error rate No character is dropped during prediction of image	CNN character classifier: 85-90% On individual characters Overall: 70-75% of text recognition
2.	Annmaria Cherian, Sanju Sebastian	Distortion removing: Haugh transform Character detection: Connected component analysis Filtering non – text region: Model work on basis of SVM Recognition: OCR	correct the orientation of image	72% recognition accuracy on perspective words
3.	Veena Rajan, Shani Raj	Noise removed: Laplacian filter Text region detection: MSER Image segmentation: Dilation and Erosion Text extraction: Bounding box	Enhanced fine details in image	improved the performances of text detection and recognition
4.	Vyankatesh V. Rampurkar, Sahil K. Shah, Gyankamal J. Chhajed,	Text detection: Connected component analysis Text Extraction: Adjacent	Recognized image in complex background	Get precision value up to 0.70.

	Sanjay Kumar Biswash	character grouping		
5	Shangxuan Tian, Yifeng Pan, Chang Huang, Shijian Lu, Kai Yu, and Chew Lim Tan	Character candidate detection: Fast cascade boosting algorithm and sliding window scheme Text line extraction: Min-cost flow network	Produced high recall and provides reliable character confidence	Text detection results on dataset using Text Flow` ICDAR2011: F-score: 80.89 ICDAR2013: F-score:80.25 Multilingual dataset: F-score:81.4
7.	Junyu Gao, Qi Wang, Yuan Yuan	novel convolutional regression network	Detected multi-oriented text in natural images	85 % F-measure on ICDAR2013 dataset 75% F-measure on MSRA-TD500.
8.	Shangxuan Tian, Shijian Lu, and Chongshou Li	weakly supervised scene text detection method (WeText)	Overcome data annotation constraints	Text Line detection results on ICDAR 2013 dataset using COCO-Text Weakly TL : F-score:86.9%
9.	Loay Alzubaidi, Ghazanfar Latif	Raspberry pi	Automatically detect vehicle license plate, reduces labor and repetitive work	90.6%

3. THE PROPOSED METHOD

It is an android mobile application which convert image containing text into machine readable form with the help of react native technology. Basically, react native is a mobile application open source framework which is created by Facebook. For performing the operation on images, we need to import some packages such as react-native-tesseract-ocr.
In this application, we will have to choose image from mobile gallery. Here, react native image picker is used to choose the image from gallery. After choosing image from gallery, it gets loaded on home screen and within minute result which is nothing but extracted text is get display on screen. Original image is shown in figure 2 and after processing through application, application's home screen shown in figure 3.

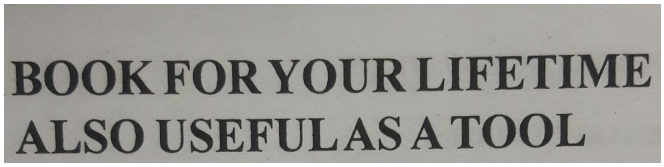

Figure 2: Original image

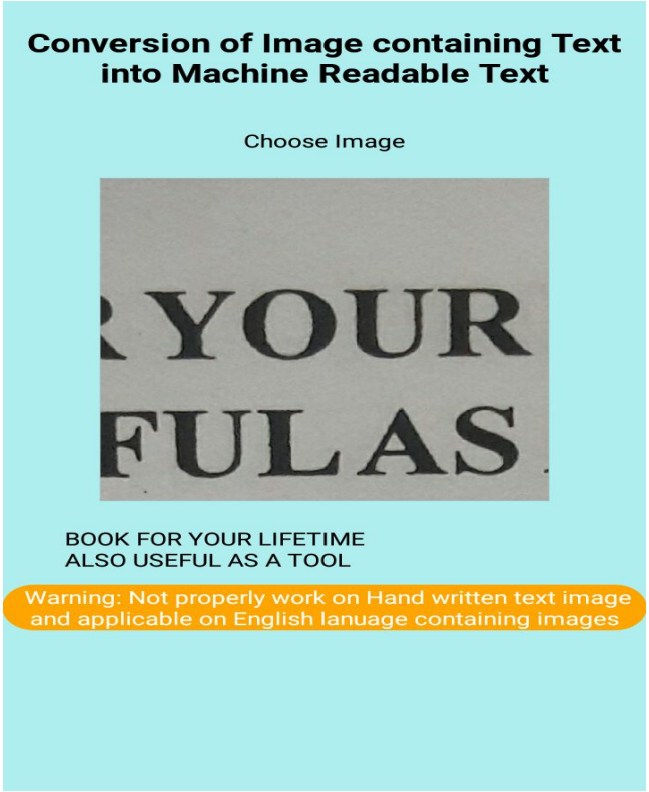

Figure 3: Processed image

4. RESULT

We performed operation only on that images which contain text only in English language. It works fine on it, for that we have to captured image from mobile camera and later we choose this image for further processing.

Future improvement could be achieved when application will work on different language as well as work on hand written text image. There is scope of such improvement because tesseract-ocr is work on different language such as French, Italian, Japanese etc.

5. CONCLUSION

In this paper, we made mobile application using react-native technology that aim to convert image containing text into machine readable form. It works absolutely accurate on image containing text in English language.

REFERENCES

1. Savita Choudhary Nikhil Kumar Singh, Sanjay Chichadwani, "Text detection and recognition from scene images using maximally stable extremal region and character classification based on neural network ", Second International Conference on Advances in Electronics, Computer and Communications, pp. 335-341, 2018. https://doi.org/10.1109/ICAECC.2018.8479419
2. Annmaria Cherian and Sanju Sebastian, "Automatic Localization and Recognition of Perspectively Distorted Text in Natural Scene Images", Conference on text detection and recognition techniques, pp. 822-828, 2016.
3. Veena Rajan and Shani Raj, "Text Detection and Character Extraction in Nature Scene Images Using Fractional Poission Model" IEEE International Conference on Computing Methodologies and Communication, pp. 1136-1141, 2017. https://doi.org/10.1109/ICCMC.2017.8282651
4. Vyankatesh V. Rampurkar, Sahil K. Shah, Gyankamal J. Chhajed, Sanjay Kumar Biswash "An Approach towards Text Detection from Complex Images Using Morphological Techniques", Second International Conference on Inventive Systems and Control, pp. 969-973, 2018. https://doi.org/10.1109/ICISC.2018.8398945
5. Shangxuan Tian, Yifeng Pan, Chang Huang, Shijian Lu, Kai Yu, and Chew Lim Tan" Text Flow: A Unified Text Detection System in Natural Scene Images", IEEE International Conference on Computer Vision (ICCV), pp. 4651-4659, 2015.
6. Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet "Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks", Conference on text detection and recognition techniques, April 2017.
7. Junyu Gao and Qi Wang, Yuan Yuan, "Convolutional Regression Network for Multi-Oriented Text Detection", IEEE transaction on Image Processing, Vol. 7, 2019.
8. Obadiah Lam, Feras Dayoub, Ruth Schulz, Peter Corke, "Text recognition approaches for indoor robotics: a comparison", Australasian Conference on Robotics and Automation, 2014.
9. Ling Hu and Qiang Ni, "IoT-Driven Automated Object Detection Algorithm for Urban Surveillance Systems in Smart Cities", IEEE Internet of Things Journal, 2016.
10. R. Komalapriya, A. Sangeetha Shiny, "An IOT Based Vehicle License Plate Recognition Using Object Detection Algorithm in Smart Cities", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 7, 2019. https://doi.org/10.22214/ijraset.2019.3362
11. D. L. Vala, Umesh Baria, Urvi Bhagat, Mohan Khambalkar, "Optical Character Recognition Robot", International Conference on Research and Innovations in Science, Engineering & Technology, Vol. 1, pp. 220–225, 2017
12. Prof. Suneel K Nagavi, Prof. Mahesh S Gothe, , Prof. Praveen S. Totiger, "Optical Character Recognition based Auto Navigation of Robot by Reading Signboard", International Journal of Engineering Research & Technology (IJERT), 2015. https://doi.org/10.17577/IJERTV4IS070130
13. Vinutha M Ha, Sweatha K Na, Sreepriya Kurupa, "Optical Character Recognition Based Auto Navigation of Robot", International Journal of Current Engineering and Technology, Vol.3, 2013.
14. Shangxuan Tian, Shijian Lu, Chongshou Li, "WeText: Scene Text Detection under Weak Supervision", IEEE International Conference on Computer Vision (ICCV), 2017.
15. Loay Alzubaidi, Ghazanfar Latif, "Real time License Saudi Plate Recognition Using Raspberry Pi", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, pp. 42-47, 2019. https://doi.org/10.30534/ijatcse/2019/0981.12019
16. Sheetal S. Pandya and Nilesh B. Kalani," Review on Text Sequence Processing with use of different Deep Neural Network Model", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, pp. 2224-2230, 2019. https://doi.org/10.30534/ijatcse/2019/56852019
17. Manal Abdullah, Afnan Agal, Marium Alharthi, Marium Alrashidi, "Arabic Handwriting Recognition Model based on Neural Network Approach", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, pp. 253-258, 2019. https://doi.org/10.30534/ijatcse/2019/4581.120419