Real-Time Text Reader for English Language

Dr. Jayshree R. Pansare¹, Aditi Gaikwad ²,Vaishnavi Ankam ³, Shikha Sharma⁴, Priyanka Karne ⁵

1,2,3,4,5MES College of Engineering, SPPU, Pune, India

---***---

Abstract—Text classification is interesting task when it comes to classifying text from diverse sources such as images, videos, and handwritten text. Handwritten text may vary as per the varied user. Henceforth, it is tough to find the best technique to categorize such kind of texts due to the absence of standard dataset and evaluation measures. Our system presents a standard method for recognition and classifying the text from all kinds of above-mentioned input sources using the Optical Character Recognition (OCR) and Support Vector Machine (SVM) classifier. Initially it recognizes the text from image and classifies and places the text into predefined classes of parts of speech for English language.

Key Words: — Real-Time Character Recognition, Optical Character Recognition (OCR), Support Vector Machine (SVM), Deep Learning algorithms.

1. Introduction

Text classification is a informative tool especially for diverse language learners like English, Chinese, Kannada, French, and German. Images, videos and handwritten texts are the sources for text recognition. Mining of text from the running slide of video by using detection algorithm, localization method and extraction techniques give the 90.8% accuracy [1]. The text annotation time is reduced by 22% using segmentation for multimedia documents. [2]. For text recognition we need a huge amount of data of city name, state name, ZIP code for recognition of handwritten documents. There are about 3000 classes for city names and 42000 for ZIP codes discussed in [3]. The Multiwriter task and Writer independent task give the accuracy 49.1% for handwritten documents. After addition of SCFG based syntax analysis the accuracy recovers from 49.1% to 54.4% as per [4]. Recognition percentage is reached at 90% for English text recognition based on features combination of structural feature and statistical feature for many letters. The proposed method in [5] for English character recognition may be reduced external noise.The analysis of SVM classifier on diverse language is explained in [6]. The recognition accuracy is achieved 73.33% for Kannada and 96.13% for English lowercase alphabets using SVM classifier. Text recognition for human is not a difficult task but for machines is quite difficult. Following steps are needed for recognition text for machines. The steps are text acquisition, text identification, image to text transformation, character recognition. [7]. The convolutional neural network and genetic algorithm are explained in[8] for feature mining of text. The feed-forward network is used for classification purpose. The Name Entity Recognizer (NER) is process for identifying Location, noun, pronoun in the statement. The Deep Learning Algorithm achieved the 70% accuracy for Name Entity Recognizer [9]. The text extraction from comic images is not same as normal image. These both are different task. For comic image the dialogue balloon is identified first then the text are extracted from that balloon [10]. The paper is organized into following Sections: Section I emphasizes existing technique for Text recognition for Image,Video and Handwritten documents. Section II describes How to detect Text from Comic Image? The overall structure of the proposed method explained in Section III. In Section IV, we present the Comparative Performance of Real-Time Text Reader for English Language based on the performance of specific systems. We conclude with the conclusion in Section V.

2. Detection of Text from Image

Image text recognition is specially separated into four stages viz.: detection, localization, extraction and recognition. The detection stage Identify the text regions. The localization stage finds the borders of strings. The extraction stage filters the background image. Fig. 2.1 portrays how image is transformed into binary image using detection, localization and extraction method. The recognition stage is used to recognized the text.

2.1 Text Detection:

The primary stage in text recognition techniques is text detection. This method is classified into two different stages. In the first texture class the entire image is distributed into chunks. For this purpose numerous methodologies can be used, e.g. wavelet transforms, spatial variance, or Gabor filter. In second class the text block and non-text blocks are categorized. It may be done by using the neural network or support vector machine. To classify text and non-text block the background of an image should be clear. For background exclusion of image background-complexity-adaptive thresholding algorithm is used.
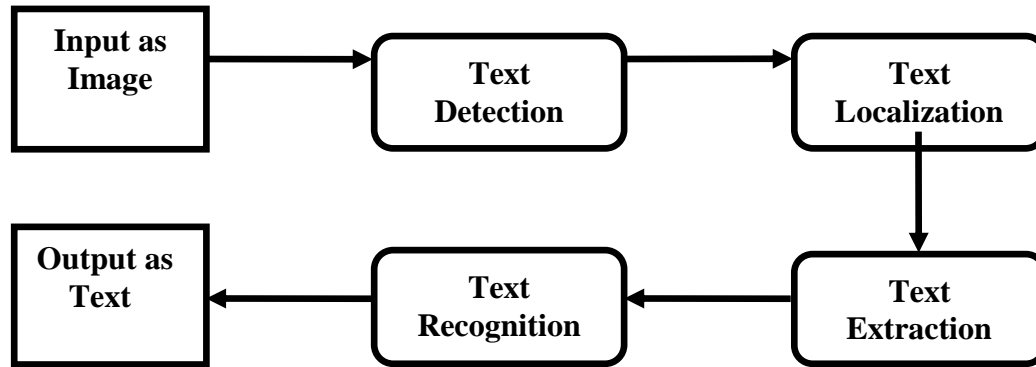
Fig 2.1 Conversion of Input Image into Binary Image

2.2 Text Localization:

Localization is secondary stage in image to text recognition. In localization stage the exact boundaries of text are determines. There are two techniques for text localization main is bottom-up paradigm and another one is top-down paradigm. In bottom-up paradigm clusters a small text regions into entire text strings based on some rules. In top-down approach is based on splitting the image into regions. The first region may be whole image, and then dividing this whole image in horizontal and vertical directions based on color or edge distribution.

2.3 Text Extraction:

The text extraction method filters the background image. Extraction may be done by two groups. One is color-based method and other is stroke-based method. The color-based method is based on assumption that the text pixels color is different than background pixel color. On the other stroke-based method apply some filters only on specific content to filter the background image.

2.4 Text Recognition:

The text recognition is the last stage of image to text conversion process. The binary image is generated by text detection, text localization and text extraction method and this binary image is given as input to text recognition phase for recognizing the text. This whole process can be done by using various techniques such as Optical Character Recognition (OCR), BP Network System etc.

3. How to detect Text from Comic Image ?

Specially, the flow adopted in Character Recognizer from Comic Image has been shown in Fig. 3.1 is fragmented in six consequent parts applied to comic images; Comic image as Input, Pre-processing, Detection of Balloon, Extraction of Text Blob, Recognition of Text, Extraction of Text.

3.1 Input as comic image:

Comic Image denotes the diverse scenes, with dialog by using balloons in an art form. For research, the comic image is considered as input which contains English character.

3.2 Pre-Processing on Comic Images:

The RGB image is converted into binary by applying the threshold values between 0 to 1. If noise containing an image is taken as input, then pre-processing steps are applied to remove the noise for improving the efficiency of text extraction. The median filter is used for noise reduction.

3.3 Detection of Balloon:

Balloon detection is necessary for text extraction from comic images. CCL algorithm is applied on noise-removed image for detecting the connected components in images. First regions boundaries are detected using CCL algorithm. Balloon detection is needed for separating the text blob and non-text blob detection.
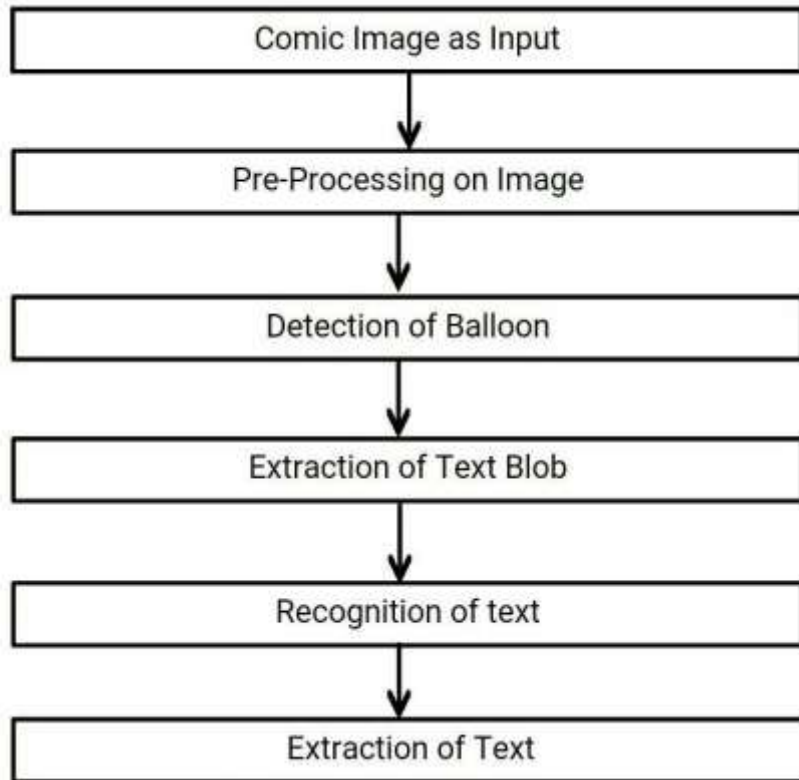
Fig. 3.1 Work Flow of Character Recognizer from Comic Image

3.4 Extraction of Text Blob:

After applying the CCL algorithm we get the script blob and non-script blob. Blob extraction is needed for categorizing the script blob from non-script blobs. By using features of blob size identification is done. If the part is 10 %(8 %) of the unique image then it is categorized as a script blob and another is a non-script blob. By using the equation (a) the part of the blobs are planned.

$$X.TB[i] = TB[j].Width * TB[j].Height \quad (a)$$

3.5 Recognition of Text:

The OCR is executed to know the text from the balloon after text blob extraction. The OCR process involves various phases like classification, feature extraction, and segmentation. Then lastly text characters are kept in a text file.

4. Proposed System

Specifically, the Fig. shows the method adopted in Text Recognition and classification. It is divided in six subsequent steps useful to captured image and conversion namely, Image scanning, Pre-Processing, Segmentation, Parts Of Speech(POS), Name Entity Recognition.

4.1 Image Scanning:

The captured image is considered as input. The image must be in .JPEG, .JPG and .PNG format. Image is captured with the help of either scanner or digital camera or mobile.
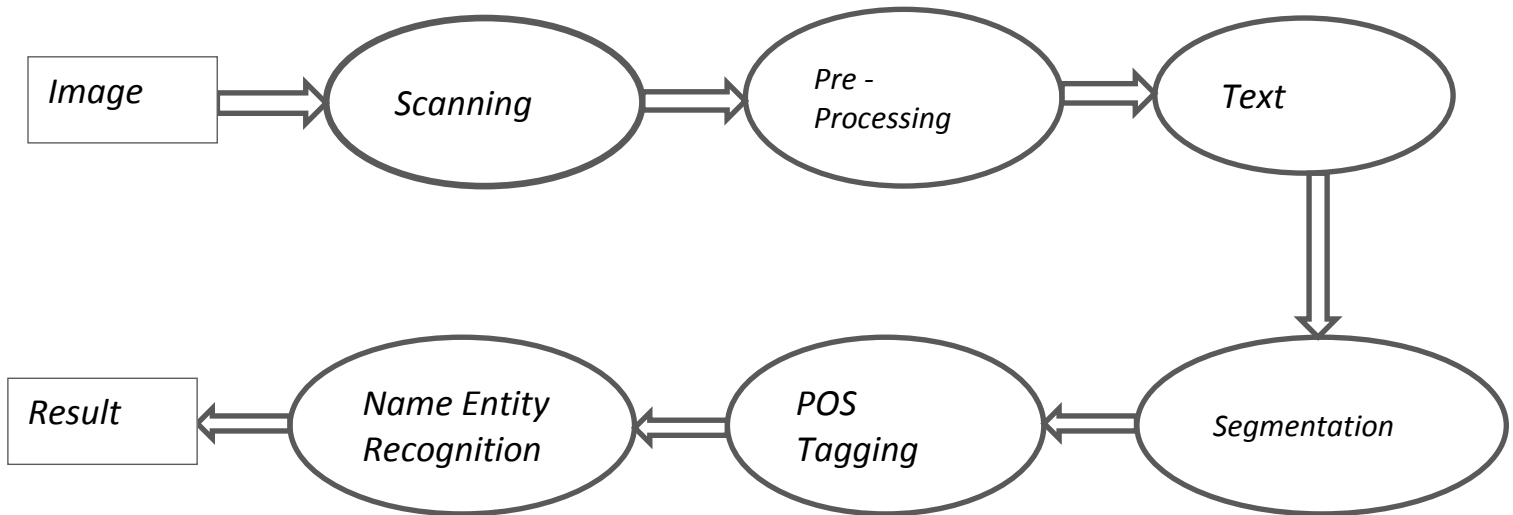
Image → Scanning → Pre - Processing → Text → Segmentation → POS Tagging → Name Entity Recognition → Result

Fig 4.1 System Architecture of Real-Time Character Recognizer

4.2 Pre-processing:

To improve quality of image pre-processing is applied on the scanned image. There are variety of methods are available for pre-processing like edge detection, noise reduction, Binarization, filing of image and dilating.

a. Noise Reduction:

Noise reduction has multiple procedures which are used to eliminate noise from an image and with this elimination it is easy to compress. Frequent algorithms are available which try to blur the unwanted part of image. For noise reduction different operations like morphological and filtering operations are used.

b. Binarization:

Binarization is nothing but conversion of gray scale image to binary image. For binary image two colors like black and white colors are used. Thresholding method is used. In thresholding the pixels which are same in gray scale are included in similar group.

c. Edge Detection:

Edge detection is a technique used for detection the boundaries of object in images. Discontinuities in image are detected by this method. Canny technique, Sobel method, Prewitt method, Roberts's system and fuzzy logic all this methods are used to perform edge detection.

d. Filing and Dilation:

In filing process holes present in images are filled. In dilation the pixels are added to the boundaries in the image, on different side the practice of removing pixels on item's margins of image is called erosion.

4.3 Text:

After performing all the steps of pre-processing image is converted into simple text. This text contains may be paragraph or sentence or single word which is provided to the segmentation step.

4.4 Segmentation:

In the segmentation whole text is converted into single words which are called as tokens. After conversion if there is a paragraph or a sentence then segmentation process splits this paragraph or sentence into tokens and the process is called as tokenization.

4.5 Parts Of Speech Tagging(POS Tagging):

After segmentation this tokens provides the meaning according to grammar such as noun, verb, adjective, infinitive, etc. which are used to learn English this process is named as Parts Of Speech Tagging(POS Tagging). For example "Chennai is the capital of Tamil Nadu.". Sentence is classified as Chennai and Tamil Nadu are as proper nouns, is as a verb, the is determiner and capital is noun.

4.6 Name Entity Recognition(NER):

Name Entity Recognition(NER) is effective and important part of machine learning. NER is nothing but classify the proper noun as entities like name, place, department, organization, etc. for example, "Savia works in Minority Department, Gurugram, Haryana." This sentence is classify as Savia as name, Gurugram and Haryana as location and Minority Department as organization.
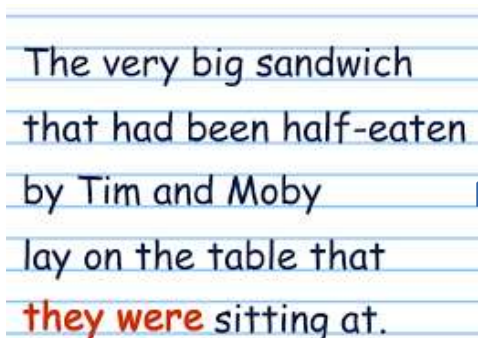
5. Results :-

The proposed Real -Time Text Reader works well on multifonts and various different types of images. We tested some of images in our system. The syatem is divided in two phases, firstly it scans the image then recognize the text by using OCR process in Phase 1.Afterwards the Overall detected text is categorizerd into Parts of Speech as noun,pronoun,verb, etc. by using Name entity Recognizer and SVM method in Phase 2. So the proper working of our system is shown in following figures and proven that this Real- Time Text Reader System is very effective.

Our system is also working for Blob images so the processing of Blob images is explained as follows:-Specially, the result adopted in Character Recognizer from Comic Image has been shown in Figure 1.6 The System is fragmented in six consequent parts applied to comic images; Comic image as Input, Pre-processing, Detection of Balloon, Extraction of Text Blob, Recognition of Text, Extraction of Text.

Input as comic image:

Comic Image denotes the diverse scenes, with dialog by using balloons in an art form. For research, the comic image is considered as input which contains English character.

Pre-Processing on Comic Images:

The RGB image is converted into binary by applying the threshold values between 0 to 1. If noise containing an image is taken as input, then pre-processing steps are applied to remove the noise for improving the efficiency of text extraction. The median filter is used for noise reduction.

Detection of Balloon:

Balloon detection is necessary for text extraction from comic images. CCL algorithm is applied on noise-removed image for detecting the connected components in images. First regions boundaries are detected using CCL algorithm. Balloon detection is needed for separating the text blob and non-text blob detection.

Extraction of Text Blob:

After applying the CCL algorithm we get the script blob and non-script blob. Blob extraction is needed for categorizing the script blob from non-script blobs. By using features of blob size identification is done. If the part is 10 %(8 %) of the unique image then it is categorized as a script blob and another is a non-script blob. By using the equation (a) the part of the blobs are planned.

$$X.TB[i] = TB[j].Width*TB[j].Height \quad (a)$$

Recognition of Text:
The OCR is executed to know the text from the balloon after text blob extraction. The OCR process involves various phases like classification, feature extraction, and segmentation. Then lastly text characters are kept in a text file.
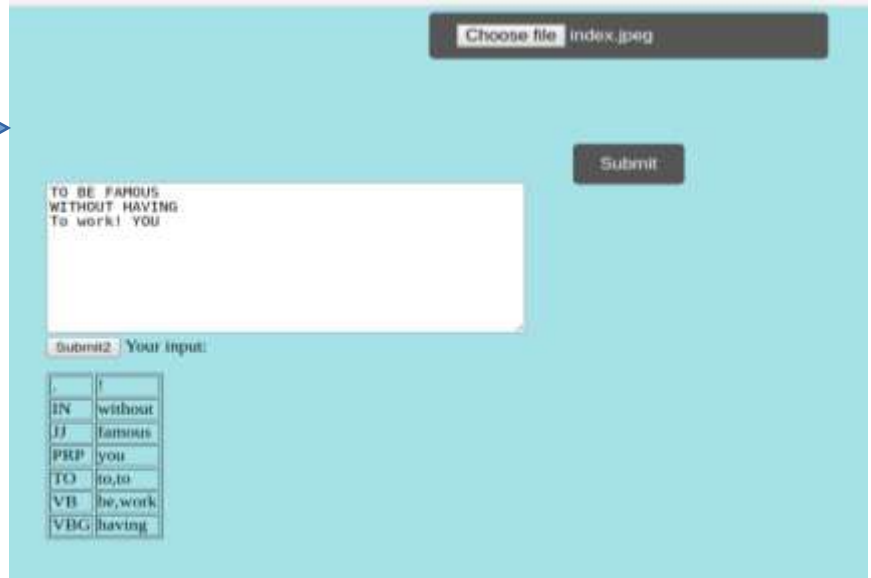
I. Comparative Performance of Real-Time Text Reader for English Language

Table 5.1 Comparative Study for Text Reader

Sr. No.	System	Methodology	Input	Accuracy
1	OCR system	Text discovery and localization by OCR method.	Text in current slide of video	67.3%
2	Kannada SVM classifier	SVM classifier character recognition	Kannada consonants	73.33%
3	English SVM classifier	SVM classifier character recognition	English lowercase alphabets	96.13%
4	Kannada and English SVM classifier	SVM classifier character recognition	Mixed English and Kannada characters	83.02%
5	DNN for parts of speech	Deep neural network Cosine similarity for Name entity recognizer	English sentences	50% - 60%
6	BP Network System	Recognition of character using BP network	English text	100% Recognition Rate
7	Recognition System	Writer Independent task for text recognition	Handwritten text	49.1%
8	Text Recognition System	SCFG-based syntax analysis for text recognition	Handwritten text	51.09%

6. Conclusion

We have studied different types of techniques which are used to recognize and classify texts in detail. This paper describes in detail processing of image. A new version for classifying parts of speech of words has been presented. The system revealed in this paper is used only for the English language. In future same method can be supported for other different Languages like Bangla, Chinese, Kannada etc. We can also create applications like text summary, billing system, etc with the help of text which is obtained after pre-processing.

References

[1] T. Q. Phan et.al. , "A Gradient Vector Flow-Based Method for Video Character Segmentation", International Conference on Document Analysis and Recognition, pp:- 1520-5363. doi:-10.1109/ICDAR.2011.207.

[2] Dr.M.Sundaresan, et.al. , "Text Extraction from Digital English Comic Image Using Two Blobs Extraction Method", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME), pp.449-452, 2012.

[3] H. Mehta, et.al. , "Optical Character Recognition (OCR) Sys-tem for Roman Script and English Language using Artificial Neural Network (ANN) Classifier" , IEEE International Conference on Research advances in Integrated Navigation Systems (RAINS- 2016), R. L. Jalappa Institute of Technology Doddaballapur, Bangalore, India, 2016.

[4] J. Bai, et.al. , "Image Character Recognition Using Deep Convolutional Neural Network Learned From Different Languages", IEEE, pp.2560-2564, 2014.

[5] M. R. Lyu, et.al. , "A Comprehensive Method For Multilingual Video Text Detection, Localization, and Ex-traction", IEEE Transaction On Circuits And Systems For Video Technology, vol.15, no.2, pp.243-255, 2005. doi:10.1109/TCSVT.2004.841653

[6] T. Q. Phan, et.al. , "Semiautomatic Ground Truth Gen-eration for Text Detection and Recognition in Video Images", IEEE Transaction On Circuits And Systems For Video Technology, vol.24, no.8, pp.1277-1287,2014. doi:10.1109/TCSVT.2014.2305515

[7] J. J. Hull, "A Database for Handwritten Text Recognition Research", IEEE Transactions on Pattern Analysis and Ma-chine Intelligence, Vol.16, no.5. pp:-550-554,2017.doi:0162-8828/94

[8] M. Zimmermann, et.al. , "Offline Grammar-Based Recognition of Handwritten Sentences", IEEE Transaction On Pattern Analysis And Machine Intelligence, Vol. 28, No. 5. pp. 818 ? 821, 2006.

[9] Y. Yang, et.al. , "English Character Recognition Based on Feature Combination", 2011 International Conference on Advances in Engineering, pp.159-164,2011. doi:10.1016/j.proeng.2011.11.2619

[10] G. Mukarambi, et.al. , "A Zone-Based Character Recognition Engine For Kannada And English Scripts", pp.3292 ? 3299, 2012.

[11] S. L. Wasankar, et.al. , "Machine Learning with Text Recognition", doi:10.1109/ICCIC.2010.5705811

[12] U. Yadav, et.al. , "A Deep Learning Based Character Recognition System From Multimedia Document", International Conference on Innovations on Power and Advanced Computing Technologies.pp.1-7,2017

[13] S. P. Singh, et.al. , "Deep Neural Based Name Entity Recognizer and Classifier for English Language", Proceeding of Second International Conference on Circuits, Controls and Communications, pp. 241 ? 246, 2017.

[14] A. D. Amensisa, et.al. , "A Survey on Text Document Categorization Using Enhanced Sentence Vector Space Model and Bi-Gram Text Representation Model based on Novel Fusion Techniques", IEEE Xplore Proceedings of the Second Inter-national Conference on Inventive Systems and Control(ICISC 2018), pp.218-225,2018.

[15] J. R. Pansare and M. Ingle, "Comprehensive Performance Study of Existing Techniques in Hand Gesture Recognition System for Sign Languages," Int J. Computer Sci. Inf. Tech-nol., vol. 7, no. 3, pp. 1343?1347, 201.