Text Detection and Recognition in Multiscript

Mr.R.Raj Bharath, M.E. (Ph.D.)
Associate Professor
Computer Science and Engineering
Manakula Vinagayar Institute of Technology, Puducherry, India

Nagaraj.R , Yuvaraj.M, Aadith.U.K
Computer science and Engineering,
Manakula Vinagyar Institute of Technology,
Puducherry, India.

Abstract — Text is one of the amazing inventions of human. It acts as a bridge between knowledge gainer and knowledge giver. Text can give us the life for what we see. In a field of Computer Vision, text detection and recogoniton plays a vital role in achieving the intelligence for machine. As a human text detection is an easy task. But an achieving that machine is an crucial task. The main objective is to integrate text detection and recognition as one thing. The System consists of 4 modules. They are 1)Text Detection 2)Preprocess the Detected part 3)Text Recogonition 4)Exporting as Documents. The first part is text detection by using Efficiency Accuracy Scene Text Detector and obtaining region of interest. The efficiency accuracy scene text detector contains of two steps only. A first step is to apply Fully Convolution Network to produce text regions. A second step is to apply non-maxima suppression to avoid multiple region of interest. The second part is preprocess the region of interest for increasing the efficiency of recogoniton. This is done by applying threshold to easily separate text from background and apply Gaussian blur for it. The third part is text recogonition by using tesseract OCR. The fourth part is exporting text into documents format by using docx, pypdf package.

I. INTRODUCTION

Technology plays an important role in reducing human work .The content of multimedia format like images and video is increasing day by day due to electronics and network advancement. Artificial Intelligence that tries to spoof like a human. The detecting and recogonizing text in image and videos is one of the key aspects of human . As a human if we can see something new visual, we automatically search for text to identify what the image is saying (or) image represents. As a evident from[1] the text is more focused on image than objects. So text detection and recogoniton is more important to understand images. Also it has lot of real time applications like sign reading, survelliance applications.

But implementing this on machine is a little bit tedious task. Text detection on controlled environment is an easy task. It can be accomplished by using threshold, gradient apporoach. This is due to presence of text in White background. But text detection in natural scenes is a complex task. The challenges are different background, poor lighting, noise in image, resolution as discussed in [1] . These makes the text detection and recogonition challenge. The text detection and recogonition in live video is also a crucial task. The reason for this one video is 24 fps. We have to process 24 images per second. The another factor is distortion in video. In images the objects are static. But in video the objects are moving. So lot of distortion comes in frame. The motion blur also reduce quality that makes the detection difficult.

II. RELATED WORK

Text detection has been an engaging topic to most researchers.A tedious amount of work has been done on image and video.Cai et al [2] uses edge detection,edge strength,edge density and threshold is applied to find text regions.hence it does not perform well for various data sets.Liu et al. [3] uses the Sobel edge details,extracting features from them and classifies pixels into text and non text clusters by applying K-means algorithm. Although it is detecting complex background, but failed to detect low contrast images.It is also computionaly expensive.Yen et al.[4] used robust text detection in natural scene images based on maximaly stable extremal regions as candidates regions and grouping text by single clustering algorithm.The above method considers the image as a high contrast.So it fails when come to video scene images due to low resolution and low contrast of video images.These method struggles in complex background ,distortion..Lee et al.[5] used support vector machines(SVM) to classify each pixel into text and non-text classes.But this methods lead to use large training sets which are computationally expensive. Epshtein et al[6] proposed canny edge images and stroke width transform to detxect text and grouping text by classical connected component algorithm. Tian et al.[7] proposed scene text detection based on weak supervision.The method reduces network dependency by focusing on weak data. The methods solve issues like multi-fonts, orientations and scripts,.But it suffer from good character candidate detection.In addition,it is hard to optimize parameters based on pre-labled samples.

As a conclusion,it requires lot of text and non transfer classifier to achieve it.A Text Attentional Convolutionl Neural Network(Text-CNN)[48] uses rich supervised information that relate text mask region to train the network which increases the efficiency against complex against complex background .Cho et al.[8] introduces a text detection system,which is called Canny Text Detector compares the similarity between image edge and text regions for improved performance. It work for video but orientation is the problem.Rong et al.[9] proposed a two –level algorithm to detect text regions in natural scene images based on the characteristics of character components. Shu Tian et al[10] proposed a unified framework based on video text extraction, text tracking, tracking based on text detection also give better results but struggles on complex background. Pawar et al[11] uses a hybrid methodology which extracts text from natural scene image with complex backgrounds. It consists of four stages. The first one is to extract superimposed text regions based on Area,perimeter,Euler number. Then it is tested for text content by using support vector machines. In the third step, detection of multiple lines in localized text regions is done and line segmentation is performed using horizontal profiles. Pise et al[12] proposed effective features for text image detection, a textregion is detected by using a mostly used features maps, histogram of oriented gradients (HOG). Binarization is applied to segment connected components. Foretext extraction, techniques like normalized height width are taken into consideration to filter out text and non-text components. Payal et al[13] proposed a mid point technique for text detection. But it does not well performed on skew image.

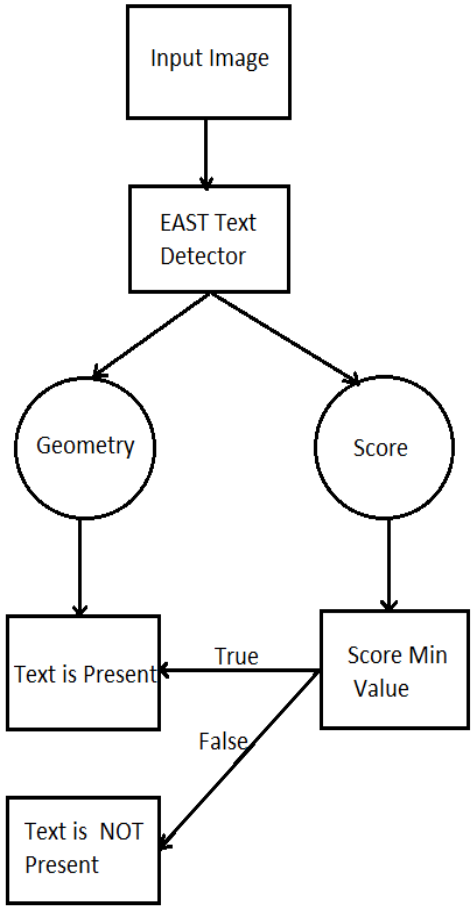

Figure 1
Flow diagram Of EAST text detector

III.PROPOSED SYSTEM

In our system,we use Efficincy Accuracy Scene Text detector for text detection.The efficiency accuracy scene text detector consists of only two steps that are fully convolutional network.A fully Convolutional Neural Network(FCN) is anormal CNN,where the last fully connected layer is substituted by another convolution layer with a large"receptive field".The idea is to capture the global context of the scene .In normal Convolutinal network a image is passed through convolution, pooling layers.After that output is fully connected layers which is 1 X 1.In fully Convolutional Network ,the fully connected layer is unsampled and bring back to input size.So am accurate NMS method is proposed in this letter,which gradually merges the highly overlapped detections in an iterative way. In each iteration, detections overlapped with the highest scored one are grouped with a harder threshold to regress for a new proposal, and then the scores within the group are softly suppressed.

.

Our system consists of 4 modules. The first module is text detection. A Text detection is done by using EAST text detector. It is the architecture that consists of two layers. The first layer is fully convolutional network and second one is non-maxima suppression. The first image is feeded as input. Then the image is resized according to be multiple of 32. Then the image is converted into blob before feeding into network.The blob is consists of two steps. They are mean subraction and scaling. the blob is subtracting each pixel value of three channel from average pixel value of three channel from training dataset. This process is called mean subtraction. After that it is divided by scaling factor.

I. For each pixel [i][j] in image
II. For each channel [i][j] in image
    a. Mean Subtraction is:
        M=pixel -mean value of respective channel
    b. Scaling factor is
        S=M/standard deviation

Then blob is set as input. After that blob is passed through model. We got two output from that model. The first one is geometry, which tells us the bounding box co-ordinates of rectangle. Then score tell us the probability of text in that region. We can loop over the values to compare with minimum confidence. If the probability is higher than minimum

confidence it will be considered as text otherwise not. Finally non-maxima suppression is applied over the co-ordinates to avoid multiple times detecting one object. After that region of interest is cropped. It is converted to graysacle and then inversed threshold. Now text becomes black with white background.Then some amount of gaussian blur to smoothen the image. The images now saved into disk and passed into recogonisation. Now all images are passed into tesseract for recogonization.T he working principle of tesseract are discussed. Blobs are organized into text lines, and therefore the lines and regions are ready for fixed pitch or proportional text. Text lines are broken into words differently according to the type of character space. Fixed pitch text is chopped immediately by character cells.Proportional text is broken into words using definite spaces and fuzzy spaces logics. Recognition then proceeds as a two steps. In the first pass, an effort is formed to acknowledge each word successively . Each word that is passed to an adaptive classifier as training data.Since the adaptive classifier learned something useful to form a contribution near the top of the page, a second pass is run again the page, in which words that not recognized tolerably are recognized again. The tesseract has three arguments. They are language, engine modes and page segmenation modes. We use engime mode as 1 it denotes LSTM engine only.The language is denoted by language code like -eng,tam,tel etc. The page segmentation mode as 8, because we tells the tesseract to assume image as a single word.

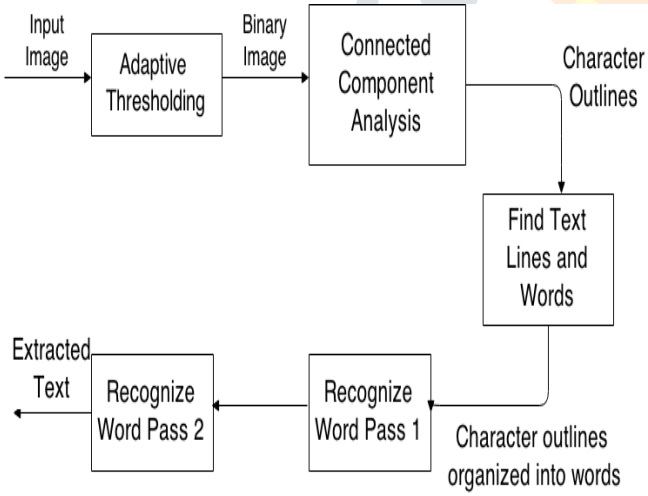

Fig 2:Working of Tesseract

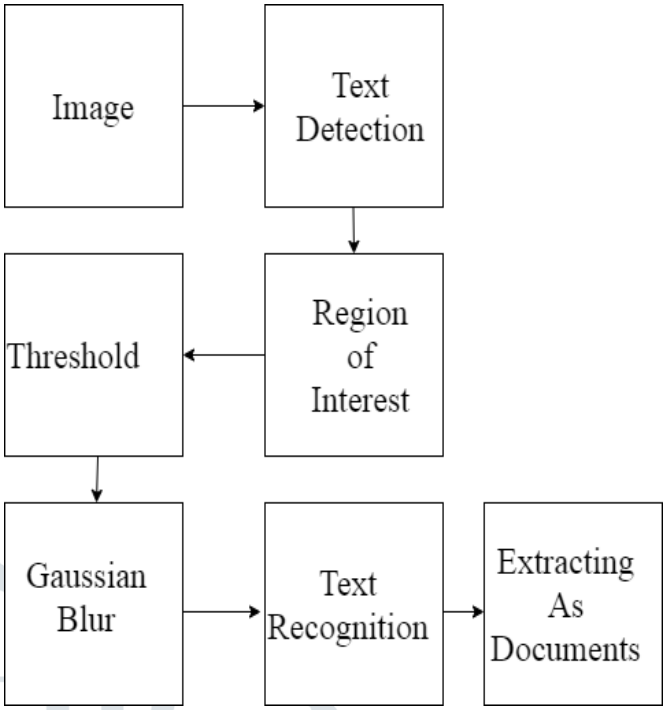

Fig 3 Flow Diagram of Proposed System

After this all modules are integrated and access via graphical user interface. We divide the document into three types. The first one is text detection. This is used for detecting only text. The next one is text recgonization. It is used for both detecting and recognizing text. The third one is Document Extraction. In some cases the image is full of text. (Example:Document).At that case text detection is not efficient. Because there is lot of text present to each other. So the image is straightly passed to tesseract by considering image has full of text.

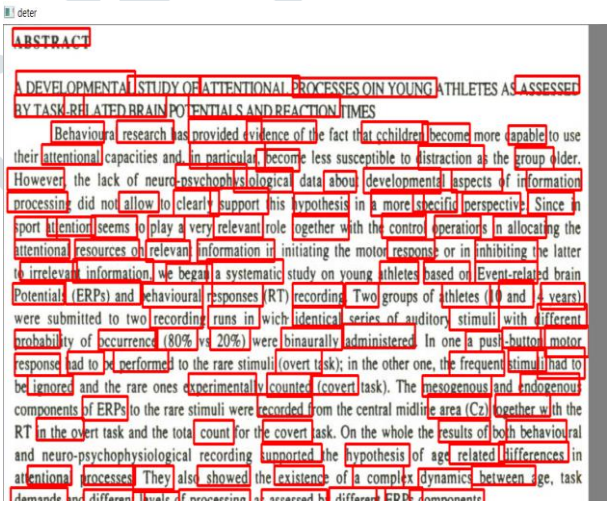

Fig 4:Text Detection with lot of text

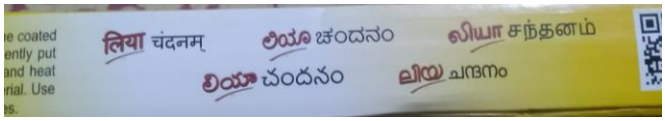

Fig 5: Input Image

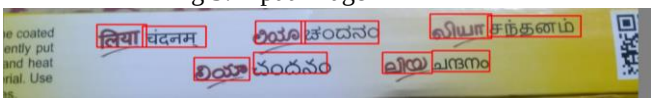

Fig 6: Text Detected Image

Fig 7:Processed Image before recogoniton

Then this images are passed into tesseract for recogoniton. Then text are recogonized and written in documents.

REFERENCES

[1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] M. Cai, J.Song and M.R Lyu,"A new approach for Video Text Detection", In proceedings of International Conference on Image Processing ,pp.117-120,2002.

[3] C.Liu, C.Wang and R.Dai,"Text Detection in Images Based on Unsupervised Classification of Edge-based features", In proceedings of International Conference on Document Analysis and Recogonition(ICDAR),pp.610-614,2005.

[4] X.C Yin, X.Yin, K.Huang, and H. W. Rao,"Robust text detection in natural images",IEEE transactions on Pattern And MachineIntelligence(TPMAI),vol36,pp.970-983,2014.

[5] C.W. Lee, K.Jung, and H.J. Kim,"Automatic Text Detection and Removal in Video Sequences,"Pattern Recogoniton Letters,vol.24,pp.2607-2623,2003.

[6] B.Epshtein, E.ofek, and Y.Wexler,"Detecting text in natural scenes with stroke width transform",In proceedings of Computer Vision and Pattern Recognition(CVPR),pp.2963-2970,2010.

[7] X.Zhao, K.H. Lin, Y.Fu, Y.Hu, Y.Liu and T.S Huang,"Text from corners: a novel approach to detect text and caption in videos",IEEE transcations on Image Processing, pp.790-799,2011.

[8] H.Cho, M.Sung, and B.Jun,"Canny Text Detector: Fast and Robust Scene Text Localization Algorithm", In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,pp 3566-3573,2016.

[9] L.Rong, W.Suyu, Z.X.Shi,"A Two level algorithm for text detection in natural images",in proceedings of Document Analysis Systems,pp.329-333,2014.

[10] Shu Tian, Xu-Cheng Yin," A Unified Framework for Tracking Based Text Detection and Recognition from Web Videos on IEEE transaction analysis and patterns

[11] Karmakar, P. , Nayak, B. and Bhoi, N. "Line and Word Segmentation of a Printed Text Document", International Journal of Computer Science and Information Technologies, vol. 5, No. 1, pp.157-160, 2014.

[12] Kaur, N. and Himani. "A Review of Different Skew Detection Techniques", International Journal of Emerging Trends in Engineering and Development, vol.2, No.4, pp. 108-115, 2014.

[13] Tang, Y., Wu, X. and Bu, W. "Text Line Segmentation Based on Matched Filtering and Top-down Grouping for Handwritten Documents", Proc. of the 11 th IAPR International Workshop on Document Analysis Systems, Chennai, India, pp. 365-369,2014.

[14] Garg, R. and Kumar, N. "An algorithm for Text Line Segmentation in Handwritten Skewed and Overlapped Devanagari Script", International Journal of Emerging Trends in Engineering and Development, vol. 4, No.5, pp. 114-118, 2014.

[15] Sneh and Kumar, M. "Segmentation of Connected Components and Overlapping Lines in Handwritten Documents", International Journal of Emerging Trends in Engineering and Development, vol. 4, No.5, pp. 114-118, 2014.

[16] Jain, S. and Singh, H. "A Novel Approach for Word Segmentation in Correlation based OCR System", International Journal of Computer Applications, vol. 99, No.18, pp. 12-20, 2014

[17] Mehdi, M. and Riaz, A. "Optimized Word Segmentation for the Word Based Cursive Handwriting Recognition", Institute of Electrical and Electronics Engineers, pp. 299-304, 2013.

[18] Jindal, S. and Lehal, G. "Line Segmentation of Handwritten Gurmukhi Manuscripts", Proc. of the 3rd International on Advance Computing Conference, Institute of Electrical and Electronics Engineers, , Mumbai, pp. 1797-1801, 2012.

[19] Kumar, A. and Jindal, S. "Segmentation of handwritten Gurmukhi text into lines", Proc. of the International Conference on Recent Advances and Future Trends in Information Technology, pp. 13-17, 2012.

[20] Kumar, A. , Jindal, S. and Singla, G. "Line Segmentation Using Contour Tracing", Journal of Global Research in Computer Science, vol.3, No.1, pp.50-54,2012.