

# Design and Implementation of Scene Text Detection

Manjushree Kaware<sup>1</sup>, Sapna Khapre<sup>2</sup>, Priyanka Fulare<sup>3</sup>

M.tech. Scholar, Dept. of Computer Science & Engg, GHRIETW, Nagpur, India<sup>1</sup>

Asst. Prof., Dept. of Computer Science & Engg, GHRIETW, Nagpur, India<sup>2</sup>

Asst. Prof., Dept. of Computer Science & Engg, GHRIETW, Nagpur, India<sup>3</sup>

**Abstract:** In recent years, with increasing popularity of portable devices for capturing images, visual processing, text extraction, etc. become key problems which gain the attention of many researchers. Extraction of text information from images or scene involves detection, localization, tracking, segmentation, enhancement and character recognition. But because of variations involved in text such as font style, size, orientation, alignment, reflections and illumination effect, with complex background and low image contrast make text extraction process more challenging. A large number of methods have been proposed to address this problem but still none of them are perfect. This paper presents an effective approach for text detection and recognition in an image. General challenges for performing scene text detection are also discussed.

**Keywords:** text extraction, text detection, localization, segmentation, character recognition.

\*\*\*\*\*

## I. Introduction

Retrieving texts in both indoor as well as outdoor environments provides appropriate clues for a wide variety of vision related tasks. A text in any form or place contains more information related to the place and helps us to understand the objective more easily. The rapid growth in digital technologies and gadgets equipped with megapixel cameras and invention of latest touch screen method in digital devices like PDA, mobile, etc. increase the demand for information retrieval and it leads to many new research challenges [5]. Detection of text and segmentation from natural scene images are useful in many applications. Recognizing text from the detected text lines is a challenging problem due to the variety of colors, fonts, existence of complex backgrounds and the short length of the text strings. Text data available in images contain useful information for indexing, and structuring, automatic annotation of images. Extraction of such type of information involves detection (The text detection stage detects the presence of text in a given input image), localization which is used to determine the location of text in the image & generating bounding boxes around the text, tracking (To reduce the processing time for localization), extraction (In this text components are segmented from the background.), enhancement (To magnify small text at a higher resolution), and recognition in which extracted text image can be transformed into text of the text from a given input image. However, variations of text due to differences in style, size, alignment and orientation as well as complex background and low image contrast make the difficulty in automatic text extraction really challenging [7].

Text can be used to easily and clearly describe the contents of an image. Many varieties of applications are found in current studies that uses extracted text. Recently developed mobile banking application that is provided by the banking institutions

facilitates the customers to carry out the transactions even on passing the image of the cheque to the server. Every such type of application depends on a Textual Information Extraction (TIE) system which can proficiently detect, localize and extract the text information present in the images. Textual Information Extraction system mainly has two phases in which text detection and localization is done in the first phase and text recognition is performed in the second phase in which detected text regions are specified to the OCR by which character recognition is done and textual output is generated [8]. The main challenge in scene text detection is to design a system which is flexible to handle all variability in our daily life including inconsistency in imaging conditions through uneven lighting, aliasing and shadowing, several character fonts and sizes and scene text. Proposed solutions for all text understanding steps must be context independent that means independent of lighting colors, scenes and all different conditions [9].

## II. Related Work

Hyung Il Koo and Duck Hoon [1], proposed a scene text detection algorithm depends on two machine learning classifiers: first classifier is used to create candidate word regions and other is used to sort out nontext regions. Connected components (CCs) in images are extracted by means of the maximally stable external region algorithm then these extracted CCs are grouped into clusters so that candidate regions are generated. Then candidate word regions are normalized and it is determined that whether each region contains text or not. A text/nontext classifier for normalized images is developed because the skew, scale and color of each candidate can be expected from CCs. This classifier is depended upon multilayer perceptrons and with a single free parameter recall and precision rates can be controlled.

Boris Epshtein , Eyal Ofek, Yonatan Wexler [2], proposed a novel image operator that is used to obtain the value of stroke width for each image pixel, and exhibit its use in natural images for text detection. The proposed image operator is local and data dependent that makes it fast and robust to reduce the need for multi-scale computation. Its simplicity enables the algorithm to detect texts in many fonts and languages. The grouping of letters can be enhanced by considering the directions of the improved strokes and curved text lines can be detected as well.

X. Chen, J. Yang, J. Zhang, A. Waibel [3], joint 1) multiresolution and multiscale edge detection 2) adaptive searching, 3) color analysis, 4) affine rectification in a hierarchical framework used for sign detection with dissimilar priority at each stage to handle the text in different orientations, sizes, color distributions and backgrounds. They used affine rectification to improve text regions deformation caused by an improper camera view angle. They extracted features from an intensity image directly rather using binary information for OCR. They proposed a local intensity normalization method to successfully handle variations in lighting and then a Gabor transform is used to find local features and then for feature selection a linear discriminate analysis (LDA) method is applied. They have utilized this approach in developing a Chinese sign system, which can involuntarily identify and recognize Chinese signs as input from a camera, and able to transform the recognized text into English. This procedure can extensively advance text detection rate and optical character recognition (OCR) accurateness.

K. Subramanian, P. Natarajan, M. Decerbo, D. Castanon [4], approached the text-localization problem using a CC-based approach by first detecting character strokes and then a threshold and stroke-width which are used for character segmentation are estimated by the detected stroke. The sensitivity of the detection algorithm to three key parameters is evaluated against four matrices: stroke precision, character recall, word recall, and computing time. The capability of character detection algorithm is not well on italic fonts or when characters of a word are encrusted together. The performance of the system can be improved by working directly on color space to detect character strokes.

Y. Pan, X. Hou, and C. Liu [5], proposed a hybrid approach to localize scene texts by integrating region information into a robust CC-based method. Parameters of a conditional random field (CRF) model are jointly optimized by supervised Learning and the binary contextual component relationships with the unary component properties are incorporated in the CRF model. The proposed hybrid approach needs further improvements because this approach fails on some texts that are difficult to segment. The speed of the

proposed hybrid approach need to be accelerated further. Text recognition should be included with text localization to complete the need of text information extraction as well.

Yao Li and Huchuan Lu [6], proposed a novel CC-based methodology for text detection in natural scene images. MSERs are first utilized as potential text regions and skeleton is applied to extract stroke width. The robust Connected Component grouping method can group characters into separated words and also eliminate false positives at the same time.

Ravina Mithe, Supriya Indalkar, Nilam Divekar [7], proposed the combination of the functionality of Optical Character Recognition and speech synthesizer. In this paper, the character recognition method is presented by using OCR technology and android phone with higher quality camera. The Optical Character Recognition deals with recognition of optically processed characters. Reliably interpreting text from real-world photos is a challenging problem due to variations in environmental factors even it becomes easier using the best open source OCR engine.

### III. Proposed Methodology

In this section, we have presented an approach for text detection and recognition in an image. Our project aim is to detect text segments in an image by using Stroke Width Transform and to recognize text segments in an image by using Optical Character Recognition.

#### A. STROKE WIDTH TRANSFORM (SWT)

In the image, a stroke is a continuous band of a nearly constant width. The SWT receives an RGB image and returns an image of the same size in which the regions of suspected text are marked. This algorithm has 3 main steps: the stroke width transform, based on their stroke width grouping the pixels into letter candidates, and lastly, grouping letter candidates into regions of text. The SWT Text Detector is designed to locate and mark the regions of an image that are suspected to contain text. The algorithm receives an RGB image, and returns an image of the same size the same as the input image, in which the pixels of each detected text region are marked. Because of the features of the Stroke Width Transform, the resulting system is capable to detect text regardless of its font, scale, direction and language.

#### B. OPTICAL CHARACTER RECOGNITION (OCR):

In Optical Character Recognition (OCR), the text lines, words and symbols in an image must be segmented properly before recognition. Correctness or incorrectness of text line segmentation directly affect accuracy of word/character segmentation and consequently affect the accuracy of

word/character recognition. Line and word segmentation is one of the important step of OCR systems. Segmentation of text line from images is a major and critical component of an Optical Character Recognition (OCR) system. The image is first preprocessed and then it is passed through the process of binarization, line segmentation, word segmentation, and character recognition. In general, text segmentation incorporates line segmentation, word segmentation from an image. It is the process through which the text component within an image is isolated from the background. For proper reconstruction of the editable text lines from the recognized characters, the line of text is first segmented, then from the segmented line the words are segmented and then from that the characters are segmented. Texts in scene images may possibly contain much information and thus separation of text strings is an important issue. In most business card images, graphic backgrounds are commonly found. So as to identify the text information from the card, the background contents and the text must be separated. OCR [10] technique enables us to effectively extract the text from an image and change it into an editable text document. In map interpretation, form processing, bank cheque processing, engineering drawing interpretation and postal address sorting, extraction of text is important. Therefore, our key purpose is to extract text from scene images.

The proposed work is divided into following steps as shown in figure1.

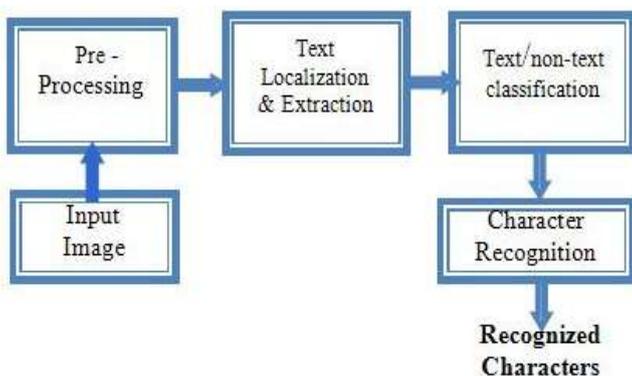


Fig 1: Methodology for text extraction from image

- (a) Preprocessing
- (b) Text localization
- (c) Text and non-text classification
- (d) Character recognition/extraction

#### a) Preprocessing

In this step, firstly we take an input image in the RGB color image. After that this RGB input image is converted to gray-scale image. To decrease the processing overload, we convert RGB image into grey scale image. After that filtering is applied to this image to remove any noises present in the

image. Then we use an edge detection algorithm to extract the edges from the final image.

#### b) Text localization

In this step, the input edge image received from the last step is binarized. Then text candidate regions are determined through thresholding. Then the entire connected component is extracted using structuring element. And then all the edges are connected by using closing operation.

#### c) Text and non-text classification

Now there are both text and non-text components presents in the localized text. So we have to separate these text and non-text components by making the bounding box for all the objects and connected components statistical metrics are used to remove the non-text components. Thresholding is used to remove very small components and, height & width of bounding box to remove the very big components.

#### d) Character Recognition / Extraction

Character recognition is performed by using Optical Character Recognition. An OCR system takes an image as input and generates a character set in editable form as an output.

### IV. Experimental Setup

The proposed approach focuses on images having English as text medium because we could measure the performance of the upgrading in the readability of such images by employing an OCR. Readability of the segmented foreground text is checked in terms of character and word recognition rates.

First of all we take RGB color image as an input after that this image is converted into binary image and then line segmentation and word segmentation is performed on that image.



Fig 2: Image binarization, text line segmentation and word segmentation

After getting word segmented image, by employing OCR we get recognized text.

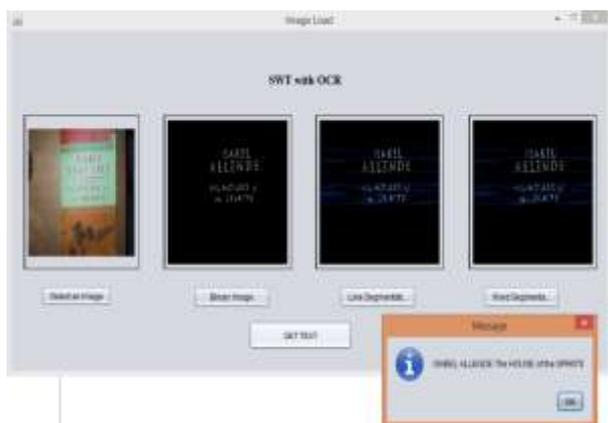


Fig 3: Text recognition from an image

#### IV. Conclusion

In this paper, we have proposed an enhanced text detection and recognition technique from images and this proposed method is tested with different types of images, both images with scene text and caption text. All interrelated methods specified in references are studied and the drawbacks are reduced and thus getting an enhanced version of the previous works such that in this work, we get reduced noise levels and major importance has been given in eliminating false positives and making it efficient for large size text.

#### References

- [1] Hyung Il Koo and Duck Hoon Kim, "Scene Text Detection via Connected Component Clustering and Nontext Filtering", IEEE Transactions On Image Processing, Vol. 22, No. 6, June 2013.
- [2] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2010, pp. 2963-2970.
- [3] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes" IEEE Transactions on Image Processing VOL.13, NO. 1, January 2004.
- [4] K. Subramanian, P. Natarajan, M. Decerbo, D. Castanon, "Character-Stroke Detection for Text-Localization and Extraction", International Conference on Document Analysis and Recognition (ICDAR), 2005.
- [5] Y. Pan, X. Hou, and C. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. on Image Processing, vol. 20, no. 3, pp. 800, Mar. 2011.
- [6] Yao Li and Huchuan Lu, "Scene Text Detection via Stroke Width" International Conference on Pattern Recognition (ICPR 2012), November 2012.
- [7] Ravina Mithe, Supriya Indalkar, Nilam Divekar, "Optical Character Recognition", International Journal of Recent Technology and Engineering (IJRTE), Volume-2, Issue-1, March 2013.
- [8] K. Jung, "Text information extraction in images and video: A survey," Pattern Recognit., vol. 37, no. 5, pp. 977-997, May 2004.

- [9] M. Swamy Das, B. Hima Bindhu , A. Govardhan, "Evaluation of Text Detection and Localization Methods in Natural Images," International Journal of Emerging Technology and Advanced Engineering (IJETA), Volume 2, Issue 6, June 2012.
- [10] Ms. Saumya sucharita Sahoo, "Review of Methods of Scene Text Detection and its Challenges," International Journal of Electronics and Communication Engineering (IJECET) ,Volume 5, Issue 1, January (2014), pp. 74-81
- [11] Mohammad Shrif Uddin,Madeena Sultana, Tanzila Rahman and Umme Sayma Busra, "Extraction of texts from a scene image using morphology based approach",IEEE Transactions on image processing, 978-1-4673-1154/12(2012)
- [12] Rishpa Sachdeva, Pooja Nagpal, " Text Localization and Extraction in Images Using Mathematical Morphology and OCR Techniques", International Journal of Scientific Engineering and Research (IJSER), Volume 1 Issue 1, September 2013
- [13] Saurav Kumar, Andrew Perrault, "Text Detection on Nokia N900 Using Stroke Width Transform" ,December 14,2010.
- [14] Gili Werner, "Text Detection in Natural Scenes with Stroke Width Transform", February 2013.
- [15] Harpreet Singh, Deepinder Singh, "Text Confining and Extraction in Image Using Mathematical Morphology" International Journal of Science and Research (IJSR), Volume 3 Issue 6, June 2014.
- [16] Satadal Saha, Subhadip Basu, Mita Nasipuri and Dipak Kr. Basu, " A Hough Transform based Technique for Text Segmentation", Journal Of Computing, Volume 2, Issue 2, February 2010.
- [17] G. Louloudis, B. Gatos, I. Pratikakis, C. Halatsis, "Line And Word Segmentation of Handwritten Documents"
- [18] Vikas J Dongre , Vijay H Mankar, "Devnagari Document Segmentation Using Histogram Approach", International Journal of Computer Science, Engineering and Information Technology (IJCSUIT), Vol.1, No.3, August 2011.
- [19] P. Nagabhushan, S. Nirmala, "Text Extraction in Complex Color Document Images for Enhanced Readability" Intelligent Information Management, 2010, 2, 120-133