# Image Text Localization

Ms. Shireen Hussain[1], Dr. Gengaje S. R.[2]

[1] *Master of Engineering student, Department of Electronics Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India.*
[2] *Ph.D. Professor, Department of Electronics Engineering, Walchand Institute of Technology, Solapur, Maharashtra, India.*

## ABSTRACT

*The detection and extraction of text regions in an image is a well known problem in the computer vision research area. Text detection and localization is the process of determining text locations in the image. The proposed approach is based on a method for edge detection, and the localization of text regions using projection profile analysis and geometrical properties. In this paper, an efficient algorithm which can detect, localize and extract text in images with complex backgrounds is presented. A conventional method for text detection and localization is introduced using discrete wavelet transform. The proposed work aims at overcoming the limitation of the existing text detection system. Our proposal is robust with respect to different font sizes, font colors, languages and background complexities. The approach is targeted towards being robust with respect to different kinds of text appearances, including font size, color and language. To achieve this aim, the main focus of the proposed algorithm is centered on the recognition of the specific edge characteristics of characters. Based on the possible different ways text areas are detected and localized. Our method can be classified in two methods which are, edge based and connected component based. This approach uses combination of edge and connected components algorithm. The proposed method will be applied on the various images and detect regions of the text perfectly. The standard images available on the internet and used by the researchers will be used for validation of the proposed system. The entire framework will be developed on MATLAB. The proposed project aims at overcoming the limitation of the existing text detection system.*

**Keyword :** - *Discrete Wavelet Transform, Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), Connected Component (CC).*

## 1. INTRODUCTION

Scenery photos and portrait photos are commonly seen when we are in journey and daily life. These photos taken by people are various and used in different situations. Cameras mounted on various hand-held devices have become very popular. And natural scene images usually taken by digital cameras are focused on more and more in computer vision. Extracting text information from natural scene images has many challenging issues. Natural scenes images contain different objects. A text object is important one among these objects, because it shows the important meanings of image. A text images can be classified into three types namely document image, scene text image, and caption image. Document image is acquired by scanning book covers, printed document etc. Scene text image sometime is referred to graphics text and it finds in natural images that contain advertisements such as street name, road signs etc. Text localization and recognition in real-world (scene) images is an open problem which has been receiving significant attention since it is a critical component in a number of computer vision applications like searching images by their textual content, reading labels on businesses in map applications (e.g. Google Street View) or assisting visually impaired [2]. Text localization and recognition problem has been recently receiving significant attention because text localization method achieved a localization recall of (62%). Text localization methods have been classified into two classes.

### 1.1 Region Based Methods:

These methods work based on color differences between text regions and their background. Region based methods are divided in two sub methods: connected component based method (CC), and edge detection based method. In first method, a text is considered as a set of distinct connected components which have their specific

intensity and color distributions. The second method depends on finding maximum intensity changes between text and background. The main differences between two methods are edge based method is useful to process low contrast text image and with different text size, while connected components methods are simpler to implement. The main drawback of CC method has failed in locating text regions in images which have complex background.

**1.2 Texture Based Methods:**

These methods work by extracting texture features of image firstly, and then a classification process is applied in the second stage to detecting text regions. Discrete transformations such as discrete wavelet transform (DWT) and discrete cosine transform (DCT) are used. These methods suffer from high complexity in nature, but it is robust in processing complex background.

This paper gives a brief review of the recent related works. It presents the proposed method. Also illustrates the experimental results, discussions as well as gives the conclusions.

## 2. PROPOSED SYSTEM

The proposed project aims at overcoming the limitation of the existing text detection system. This project provides a robust technique based on discrete wavelet transform, edge detection, and morphology operation for scene text detection. The proposed method will be applied on a various images such as images of low contrast, complex background images and images of different fonts and size of text. The experimental results show that the proposed method can detect regions of the text perfectly.

First, they apply a color edge detection algorithm in YUV color space and filter out non-text edges using a low threshold. Then, projection profiles are analyzed to localize text regions. Finally, a local thresholding technique is employed in order to keep low-contrast text and simplify the background.

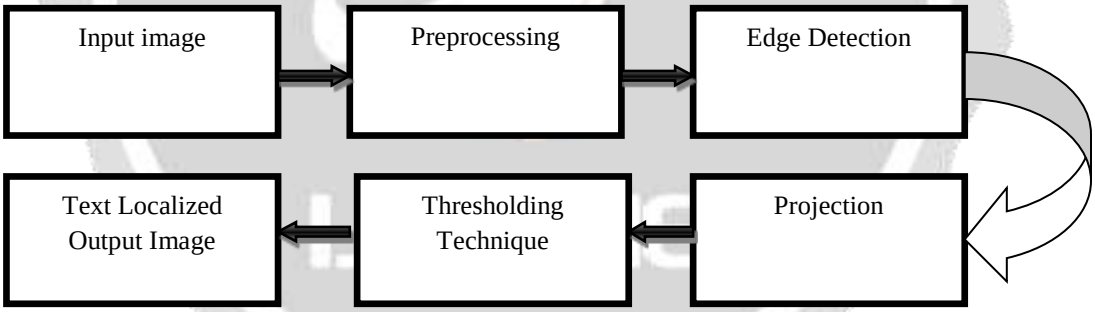



**Fig -1**: Proposed system.

## 3. METHODOLOGY

In this section, a solution for detecting text in an image is provided. As presented in Fig.1 the proposed text detection algorithm passes through the following steps.

1) Preprocessing.
2) Text detection.
3) Text Localization.

The input images used can be in different font, color and language. Input image used for this experiment is as shown in **Fig -2**.


**Fig -2**: Input Image.

**3.1 Preprocessing.**

The input image is RGB, it must be converted into YUV color space by forming weighted sum of the R, G, and B components as in Equation 1.

Y =0:299*R+0:587*G+0:114*B
U =0:14713*R-0:28886*G+0:436*B (1)
V = 0:615*R-0:51499*G-0:10001*B

Our method uses only luminance components (Y) for next processing steps. The Fig.3 shows the YUV image.

(a) (b) (c)

**Fig -3**: YUV image (a).Y image (b).U image (c).V image.

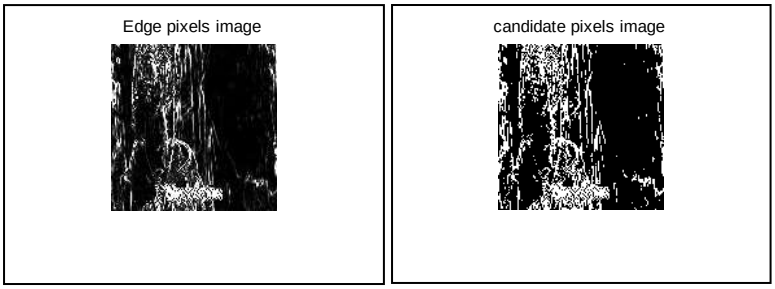
This Y image is further converted into Discrete Wavelet transform. One-dimensional DWT is applied on an image, an image is decomposed into two parts which is coarse and detail elements by using low pass (L) and high-pass (H) filters. Image signal is decomposed into four sub bands when two dimensional DWT is applied on rows at first and then on columns of an image. These sub bands are approximation sub-band (LL), and details sub bands (LH, HL, and HH). Fig.4. shows the Single level decomposition of DWT.

**3.2 Text Localization.**

Text localization will perform edge detection of the image. The text pixels have high variation around its neighbor pixels, therefore a technique based on edge detection will be applied. A wavelet edge pixel E(i,j) at pixel in location i and j can be defined by taking the average of corresponding pixels in three high frequency sub bands (LH, HL and HH) and is shown in **Fig-4.**(a).

$$E(i,j) = \sum_{k=1}^{3}((Dk(i,j)\text{^}2) \ /3$$

Then each pixel in E(i,j) array will be a candidate text pixel if its value is larger than threshold α, which is a parameter its value is selected based on the local statistics such as mean and standard deviation of an image.
**Fig-4.**(b) shows candidate text pixel detection. The value of α=43.89 for image.

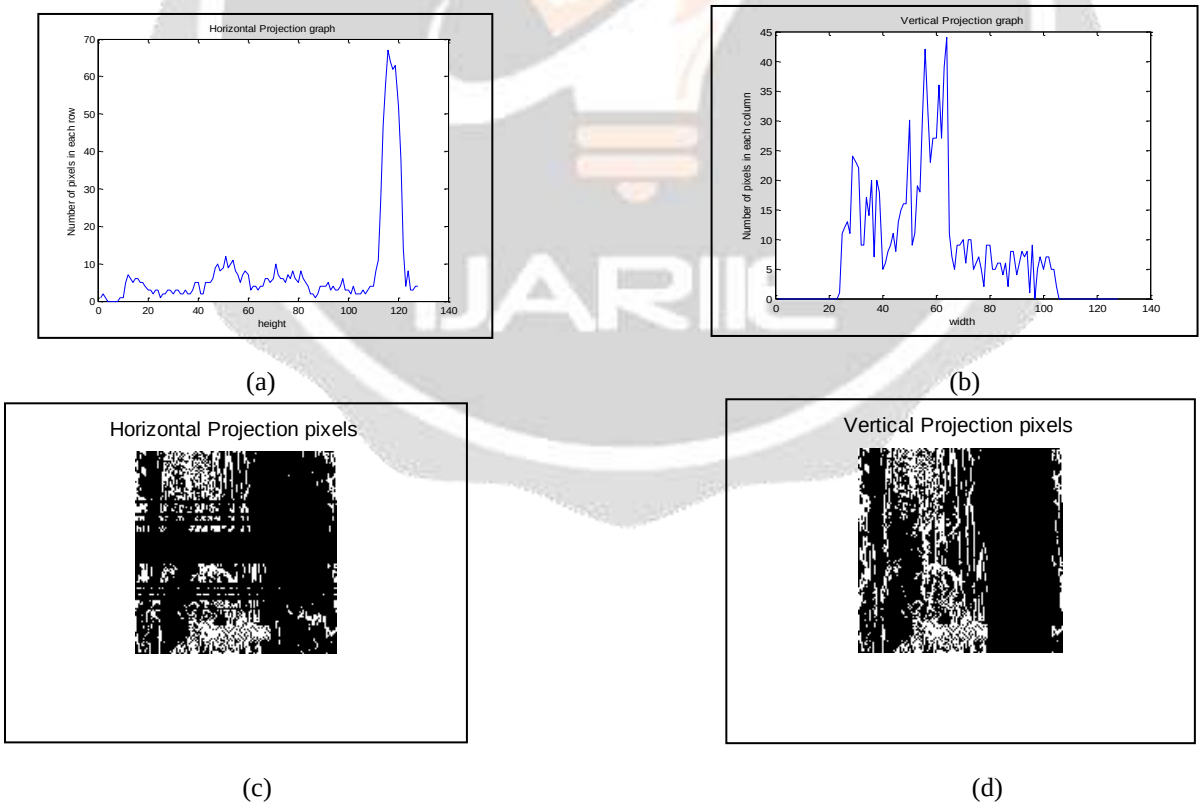



(a) (b)

**Fig -4**: (a) Edge pixels detection of image. (b) Candidate pixels of image.

### 3.2 Text Detection.

Text detection is the process where projection and thresholding technique is performed. Vertical and horizontal projections are one of the amplitude segmentation methods. Vertical and horizontal projections of a binary candidate text pixels image, are performed for horizontal projection Phor of row as the sum of pixel values in that row and all the columns in the image. While vertical projection Pver of column is computed as the sum of pixel values in that column and all the rows in the image. To segment text regions, two thresholds are used horizontal and vertical thresholds, which denoted as Th and Tv respectively. They are defined as average of mean and minimum of respective projection. If Phor is greater than horizontal threshold Th then Phor of that row can be considered as a part of candidate text region; otherwise this row is suppressed. Similarly, if Pver is greater than vertical threshold Tv, then Pver of that column can be considered as a part of a candidate text region. **Fig -5** shows projection of candidate text image.




(a) (b)

(c) (d)

**Fig -5**: Projection of candidate text image(a) graphical representation of horizontal projection (b) graphical representation of horizontal projection (c & d) horizontal and vertical projection image.

The next step perfomed is connected components. In this step, horizontal and vertical images are combined to be one image, which is shown in **Fig -6**(a). And then non-text regions which are false detected are eliminated using geometrical features of regions. Area geometric feature of each region is computed and compared with threshold (thr1) which computed according to Equation (2).

$$thr1=MaxArea/10 \quad (2)$$

where Maxarea represents a largest region area. If the area of each region is less than threshold, then it is considered as non-text region and its pixels should be discarded. If (R) of region larger than (thr2) then discarded region elements and **Fig -6**: (b) shows the result for this connected components step.

(a) (b)

**Fig -6**: Connected components step (a) combined projection image (b) result image after discard non-text regions.

As shown in Fig.6-(b), the resultant image may contain non-text pixels. So a morphology operation are performed where dilation followed by opening operation are applied to further remove nontext pixels. **Fig -7** shows the detection of text regions after applying morphology operations.

(a) (b)

**Fig -7**: Result (a) After Morphological Operation. (b).Text Localized Output Image

## 4. CONCLUSIONS

The proposed system is validated using set of images available using suitable performance metrics. The proposed method will be applied on the various images and detect regions of the text perfectly. The entire framework will be developed on MATLAB platform. A conventional method for text detection and localization is introduced using discrete wavelet transform. The proposed method is more robust in dealing with the contrast, complex background images and images of different fonts and size of text. Using wavelet transform gives good tools for detecting candidate text regions. The proposed work provides a system for detecting the text regions in scene images which is an open problem.

## 6. REFERENCES

[1]. Shivananda V. Seeri, J.D. Pujari, P.S. Hiremath, Multilingual text localization in natural scene images using wavelet based edge features and fuzzy classification, Int. J. Emerg. Trends Technol. Comput. Sci. 4 (1) (January-February 2015) 210e218.

[2]. L. Neumann, J. Matas, "Scene text localization and recognition with oriented stroke detection," in: IEEE International Conference on Computer Vision (ICCV), 2013, pp. 97e104.

[3]. Julinda Gllavata1, Ralph Ewerth1 and Bernd Freisleben, '' A Robust Algorithm for Text Detection in Images," SFB/FK 615, University of Siegen, D-57068 Siegen, Germany, Dept. of Math. & Computer Science, University of Marburg, D-35032 Marburg, Germany.

[4]. G. Zhou, Y. Liu, Q. Meng, Y. Zhang, "Detecting multi lingual text in natural scene", in: 2011 1st International Symposiumon Access Spaces (ISAS), IEEE, 2011, pp. 116–120.

[5]. Chip Hong Chang and Mingyan Zhang, "Fuzzy-Art based adaptive digital watermarking scheme", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15 no. 1, 2005, pp. 65-81.

[6]. B. Epshtein, E. Ofek, Y. Wexler, "Detecting text in natural scenes with stroke width transform," in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2963e2970.

[7]. C. Yi, Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Process 20 (9) (2011) 2594e2605.

[8]. S.V. Seeri, S. Giraddi, B.M. Prashant, "A novel approach for Kannada text extraction," in: Proceedings of the International Conference IEEE on Pattern Recognition, Informatics and Medical Engineering, 2012, pp. 444e448.

[9]. H. Koo, D.H. Kim, "Scene text detection via connected component clustering and non text filtering," IEEE Trans. Image Process 22 (6) (2013) 2296e2305.

[10]. H. Raj, R. Ghosh, "Devanagari text extraction from natural scene images," in: International Conference on Advances in Computing. Communications and Informatics (ICACCI), IEEE, 2014, pp. 513e517.

[11]. Keechul Jung, Kwang In Kim, Anil K. Jain, "Text information extraction in images and video: a survey, Pattern Recognit." 37 (5) (2004) 977e997.