ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 4, April 2016**

# Detecting Multi Oriented Text from Natural Scene Images

Anagha Purohit, Kishori Shekokar

ME Student, Dept. of Computer Engineering, Sigma Institute of Engineering, Bakrol, India

Associate Professor, Dept. of Computer Engineering, Sigma Institute of Engineering, Bakrol, India

**ABSTRACT**: In today's world of globalization the text detection and recognition from natural scene images is one of the most researched topic in the field of computer vision. Due to the wide area of applications the constant need for improvement of text detected from the image is required. The paper gives a fresh approach to improve text detection for Scene images with text appearing in different orientations. The paper usues stroke width transform which is a region based technique. Along with whit it uses ellipse growing method that would help detect curved text regions and for removing false positives we use principle component analysis. Also we use connected component analysis for candidate filtration. The results obtained by the method proposed are promising.

***KEYWORDS***: Multi Oriented Text,SWT, edge detection, ellipse growing, PCA, edge map.

## I. INTRODUCTION

Texts images can either be document images, scene images or born- digital images. Document images are just the image-format of any document. Document images might have text as well as graphics. These images can be scanned or taken through a camera phone so that, these images are transformed into electronic format. The natural scene images hold the text, such as the banners, advertising boards, etc. which the camera captures; therefore, text appears with all the background details [1]. These images are very challenging to detect and recognize, because of the complex and poignant backgrounds. The text in natural scene images might be of different sizes and different orientations. Born-digital images are generated by computer software and are saved as digital images [2]. As compared to the document images and scene images, there are more defects appear in born digital images, like more complex foreground/background, low resolution, compression loss, and severe edge softness. Because of which text extraction becomes difficult from the image background [1].

Image understanding, content-based image filtering, super map, vehicle testing, optical character recognition, text extraction in video sequences, wearable applications et al., are some of the applications in which text detection can be used We'll be talking about the text detection in natural scene images. There are mainly three techniques for detecting text from natural scene images.


Fig. 1. (a)Document Image (b)Scene Image (c)Born Digital Image
Fig. 2.

When we talk about text detection in natural scene images. We need to know what kind of images we are dealing with and what all the ways of detecting the text are. Mainly three text detection methods are described in the literature: Texture based methods, Region based methods and Hybrid methods. As the name suggests the texture based methods

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

deal with the texture properties of the image for identifying the text. These methods require high computing capabilities because it makes use of properties like spatial variance and local intensities and scanning of all locations is to be done comprehensively [8]. Examples of texture based method include- SVM classifier, K-means classifier et al. The region based methods help identify the text based on different parameters viz. the width of each letter candidate, the angle at which each text letter candidate is located and the vector. Till now it is seen that the region based methods give better results as compared to the texture based methods. The region based techniques include methods like Stroke Width Transform (SWT) Maximally Stable Extremal Regions (MSER) and others. The hybrid method is obviously combination of the texture and region based methods. The edge pixels of all possible text regions are extracted, using a detailed edge detection method; the gradient and geometrical properties of region contours are verified to generate candidate text regions, followed by a texture analysis procedure to separate true text regions from non-text regions [11].

We have used region based method for candidate extraction. The method used exhaustively uses the functionalities of region based method along with some hybrid techniques for detecting the circular text. The region based techniques as we know depend largely on the edge map therefore we need to ensure that the edge detection method that we use gives the best possible result.

The applications of the system include tagging and annotations, indexing and searching, computerized aid for visually impaired, and many other fields of computer visioning. Computer visioning intend to acquire, process, analyze and understand image for different purposes.

## II. RELATED WORK

These days most of the research is been done on Stroke Width Transform. Wenyang Dong et al. [18], in their paper they have suggested a method for localized stroke width transform which is proved to be better than SWT. They used adaptive image binarization and implemented SWT in localized regions; this method is robust against edge detection because it reduces contrast between the background of the image and text.

Another method proposed by Anna Zhu et al. [19], use auto image partitioning, two stage grouping and two-layer classification for detecting the text from natural scene images. In this paper they have suggested, three methods divided into different segments, initially the image is segmented into different sub images based on statistical distribution of the pixels. Next comes two stage grouping the first part groups each subimage in the first stage, the second stage groups connected components to text regions. For classification first the similarity score of the regions is checked and next the SVM classifier is applied along with HOG features.

There are some studies related to neural networks for detecting text in scene images. One such work proposed by Lei Sun et al. [20] presents a robust text detection approach based on neural network and colour-enhanced contrasting extremal region (CER) for text & non-text classification. Le Kang et al. [21] use higher order correlation clustering (HOCC) for text line detection in natural images. Maximally Stable Extremal Regions are grouped based on spatial alignment and appearance consistency.

The last paper we'll discuss about has made detection of curved text in natural scene images possible. AnharRisnumawan et al. [7] have suggested a concept of using symmetry characteristics of the text in order to detect horizontal, angular as well as curved text to be detected. A reliable system based on the concepts of Mutual Direction Symmetry, Mutual Magnitude Symmetry and Gradient Vector Symmetry properties to identify text pixel candidates despite of any orientations including curves from natural scene [7]. Canny and Sobel Edge detection methods are used and the symmetry properties are combined together in order to get a better output. In Mutual Direction Symmetry the stroke width direction of pixel is considered, in Mutual Magnitude Symmetry checks the distance between the two pixels selected using MDS and Gradient Vector Symmetry separates the text and non-text pixels. Besides the integration of the symmetry property is done. And growing ellipse method is incorporated to detect the curved texts.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 4, April 2016**


Fig. 2. (a) Original Image (b)Canny Image (c) Sobel Image

## III. PROPOSED METHOD

### *A. Image Preprocessing*

In computer vision the image needs to be acquired, processed, analyzed and understood in order to obtain either statistical or mathematical results. The preprocessing is first step for text detection from real or natural images.

Here, the image will be taken and converted to gray image for further preprocessing. The next step in preprocessing is edge detection. As we know, all the region based techniques need the edge map to identify the text candidates. We need to use the best possible edge detection method for better results of the detected text. Below is the comparison of two edge detection techniques where we can see the significant difference in the edge map for different images.


Fig. 3. (a) Original Image (b) Gray Image (c) Canny Edge Image

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 4, April 2016**

*B. Image Binarization*

With the guidance of edge map an intersection surface is constructed to cut gray image and classify all pixels into two categories to generate the binary image, from which it is much easier to extract character regions [18].


Fig. 4. Binarized Image

*C. Candidate Generation*

The pixels are associated to form connected components using a simple rule that there should be 4 connections with value 1 in the binary image [18].

*D. Stroke Width Computing*

Following are the steps associated with Stroke Width Computing[22]:

Step 1. Initialize each element with ∞.
Step 2. Apply Canny Edge Detection
Step 3. If p lies on stroke boundary
  a. dp is roughly perpendicular to the orientation of stroke
  b. else, check again if the stroke lies on the boundary.
Step 4. Follow the ray: r= p+n*dp until q is found.
  a. Check direction dp at p.
  b. Else, go to Step 6.
Step 5. If dp is roughly to opposite to dp.
  a. dq = -dp ± π/2.
  b. Else, go to Step 6.
Step 6. Discard ray, r.
Step 7. Stroke Width = |p-q|.


Fig. 5. SWT Image

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 4, April 2016**

The figure above shows the Stroke Width Transform results. The results are obtained using cute-80 dataset. As the SWT image can be seen the results look pretty good. The Stroke Width Transform is a region based technique which gives a very good result as compared to other techniques like MSER or texture based methods.

*E. Ellipse Growing*

Assume that the distance between two character candidates is smaller than the character width.Select a random component Ci and grow the ellipse by a unit pixel on major axis denoted by A(Ci) and minor axis denoted by a(Ci), till it touches another text component Cj [7].Also, the minor axis lies exactly at the center of the major axis and is orthogonal to it [7].Create an ellipse component and grow its major and minor axis till it finds the nearest component and mark it as traversed.The terminating condition is defined as - The maximum number of iterations must be less than or equal to bounding box's width and height of the components. In this manner we create different groups of components.


Fig. 6 (a) Original Image (b) Sample text components of word "STAR" (c) Ellipse for the whole word [7].

The above figure shows ellipse growing method. The word is covered with an ellipse with major axis covering the word length and the minor axis covering its height. The further analysis is done using Principal component Analysis where the ellipse major and minor axis will be compared with PCA major and minor axis and based on results false positives will be eliminated.

*F. False Positive Elimination*

For finding the Principal Component we first find the direction of maximum variance in the samples and align it with the first axis. Continue this process with orthogonal directions of decreasing variance, aligning each with the next axis. For each group of components, compare the angle of the major axis given by the ellipse and the angle computed by the Principal Component Analysis (PCA). For a component, if the angles computed for both its ellipse's major axis and PCA are the same, then confirm that it is a text component or else discard it as a non-text component.Divide the group of components into two equal sub-groups if the ellipse covers a word with greater than or equal to four components as shown in fig 6.Check whether the minor axis, pixel distribution, direction, and standard deviation of the proximity matrix of each sub-group are equal.

*G. Text Letter Candidates:*

Finally, the detected text is obtained by identifying the text letter candidates. The figure shows the final result after false positive elimination. The results may vary for different images because of the light and quality of image.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 4, April 2016**

Fig. 7. Filtered Text

## IV. PERFORMANCE PARAMETERS AND EXPERIMENTAL RESULTS

The terms Precision measure (p), Recall measure(r) and F-measure (f) are the measures to compare accuracy of text segmentation.

These performance parameters are described below:

Precision (p) = TP / (TP +FP) (1)

Recall (r) = TP / (TP +FN) (2)

F-measure (f) = 2* (p * r) / (p + r) (3)

TP = Correctly detected text pixels

FP = False positives

FN = False negatives

Also, false positive means pictorial pixels are detected as text where in real case they are non-text. False negative means the text pixels are not detected or detected as pictorial pixels.

Following are the performance parameter results based on different angles used with Stroke Width Transform.

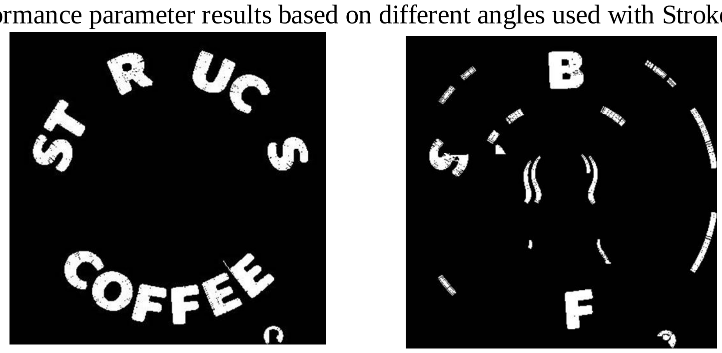

Fig. 8. (a)SWT with Canny (b)SWT with Sobel

Table. 1. Comparison of SWT with Canny and Sobel based on performance parameters

| Methods | Precision | Recall | F-measure |
|---|---|---|---|
| **SWT with Canny** | 0.69 | 0.66 | 0.67 |
| **SWT with Sobel** | 0.23 | 0.35 | 0.28 |

Fig. 8 shows how the edge map is so important for the edge detection using region based method. Thus, reflects the performance parameters table that if the edge map is clear the precision and recall are better. Again the precision parameters are based on the confusion matrix which would be heuristically created. Therefore, the measurements may change from person to person based on the perception. The only way to know whether our detected text is accurate is through character recognition.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

## V. CONCLUSION AND FUTURE WORK

We've seen different methods for detecting text from natural scene images. Most of the methods till now have been successful in detecting only the horizontal text. Some of them have succeeded in finding the text located at different angles but not much of work has been done on the curved text detection. This might be interesting to find text which is placed in a curved manner. The general idea of taking advantage of the Stroke Width Transform results along with ellipse growing might efficiently detect the text from natural scene images. Also using Principle component analysis we can remove the false positives. Future work might help detecting differently oriented texts from scene images.

## REFERENCES

Jian Zhang, Renhong Cheng, Kai Wang, Hong Zhao, "Research on the text detection and extration from complex images", Fourth International Conference on Emerging Intelligent Data and Web Technologies. Vol. 10, 2013, Page no. 708-713.

Uma B. Karanje, Rahul Dagade, "Survey on Text Detection, Segmentation and Recognition from a Natural Scene Images", International Journal of Computer Applications (0975 – 8887), Volume 108 – No. 13, December 2014, Page no. 39-43.

K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. PAMI, vol. 25, no. 12, pp. 1631–1639, 2003.

X. Chen and A. Yuille, "Detecting and reading text in natural scenes," in Proc. of CVPR, 2004.

L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in Proc. of ACCV, 2010.

H. Chen, S. Tsai, G. Schroth, D. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," in Proc. of ICIP, 2011.

AnharRisnumawan, PalaiahankoteShivakumar, Chee Seng Chan, Chew Lim Tan, "A Robust Arbitary Text Detection System for Natural Scene Images", Expert Systems with Applications, ELSEVIER, July 2014.

J. Gllavata, R. Ewerth, and B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients," in Proc. of ICPR, 2004.

Y. Pan, X. Hou, and C. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. Image Processing, vol. 20, no. 3, pp. 800–813, 2011.

Cong Yao, Xin Zhang, Xiang Bai, Wenyu Liu, Yi Ma, and ZhuowenTu, "Detecting texts of Arbitary orientations in Natural Images", IEEE Transactions on Image Processing, May 10, 2012

Y. Zhong, K. Karu, and A. K. Jain, "Locating text in complex color images," Pattern Recognition, vol. 28, no. 10, pp. 1523–1535, 1995.

V. Wu, R. Manmatha, and E. M. Riseman, "Finding text in images," in Proc. of 2nd ACM Int. Conf. Digital Libraries, 1997.

H. P. Li, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," IEEE Trans. Image Processing, vol. 9, no. 1, pp. 147–156, 2000.

Y. Zhong, H. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," IEEE Trans. PAMI, vol. 22, no. 4, pp. 385–392, 2000.

J. Weinman, A. Hanson, and A. McCallum, "Sign detection in natural images with conditional random fields," in Proc. Of WMLSP, 2004.

K. Wang and S. Belongie, "Word spotting in the wild," in Proc. ECCV, 2010.

Wenyan Dong, ZhouhuiLian, Yingmin Tang, and Jianguo Xiao, "Text Detection in Natural Images using Localized Stroke Width Transform", Springer International Publishing, Switzerland, 2015.

Anna Zhu, Guoyou Wang, Yangbo Dong, "Detecting Natural Scene text via auto image partitioning, two-stage grouping and two-layer classification", Pattern Recognition Letters, ELSEVIER, June 2015.

Lei Sun, QiangHuo, Wei Jia, Kai Chen, "A Robust Approach in the Text Detection from Natural Scene Images", ELSEVIER, April 2015.

Le Kang, Yi, Li, David Doermann, "Orientation Robust Text Line Detection in Natural Images", IEEE Xplore, CVPR 2014.

Zheng Zhang, Wei Shen, Cong Yao, XaingBai, "Symmetry-Based Text Line Detection in Natural Scenes", IEEE Xplore, CVPR2011.

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

## BIOGRAPHY

**Anagha Purohit**is a ME student at the Computer Engineering Department, Sigma Institute of Engineering. Her area of interest is Image Processing and she is in the final year of her Master's programme.

**KishoriShekokar** is an Associate Professor at the Computer Engineering Department, Sigma Institute of Engineering. She completed her Masters of Engineering in Amravati, Maharashtra. She is keen in teaching and guiding students for their BE projects and ME dissertation.