iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** V **Month of publication:** May 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Survey on Localizing Text in Scene Images

Fathima A Muhammadali[#1]

[#]M-Tech Student, Computer Science And Engineering,
MG University, Kottayam, India

Abstract— ***Localizing text method is used for extracting text regions from scene images with complex backgrounds and multiple text appearances. Natural scene images with text information are divided into two categories according to the complexity of the background. 1) Text characters and strings are in high resolution with a relatively simple background, eg: book covers, notice signages, and wrappers. 2) Text into more complex backgrounds with various natural objects, eg: buildings, trees, lawns, roads. In the literaure survey of Localizing text, studied different text detection techniques. These techniques have many advantages and disadvantages. Text localization is evaluated on scene images, born-digital images, broadcast video images, and images of handheld objects captured by blind persons. This method is used for navigation, assistive reading, geocoding, and content-based image retrieval.***

I. INTRODUCTION

Localizing text from scene images are used for navigation, assistive reading, geocoding, and content-based image retrieval, and so on. Natural scene images with text information are divided into two categories according to the complexity of the background. 1) Text characters and strings are in high resolution with a relatively simple background. 2) Text into more complex backgrounds with various natural objects, such as buildings, trees, lawns, roads.

In camera-based scene images, manual text localization is impractical, especially for blind people. Therefore, algorithms of automatic text localization are required to filter out background outliers and localize the regions containing text characters or strings in images for further processes of text segmentation and recognition

Fig. 1 Example Text Localizing from Images

II. DETECTING TEXT IN NATURAL SCENES WITH STROKE WIDTH TRANSFORM

A. Overview

This is a novel image operator that seeks to find the value of stroke width for each image pixel, and demonstrate its use on the task of text detection in natural images. The suggested operator is local and data dependent, which makes it fast and robust enough to eliminate the need for multi-scale computation or scanning windows. Extensive testing shows that the suggested scheme outperforms the latest published algorithms. Its simplicity allows the algorithm to detect texts in many fonts and languages. The main objective of this method is to Find the value of stroke width for each image pixel and its use on the task of text detection in natural images.

The method suggested here differs from previous approaches in that it does not look for a separating feature per pixel, like gradient or color. A pixel gradient is only important if it has a corresponding opposing gradient. This geometric verification greatly reduces the amount of detected pixels, as a stroke forces the co-occurrence of many similarly matched pairs in a small region. Another notable difference from previous work is the absence of scanning window over a multiscale pyramid, required by several other approaches. Unlike previous features used for text detection, the proposed SWT combines dense estimation (computed at every pixel) with non-local scope.

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

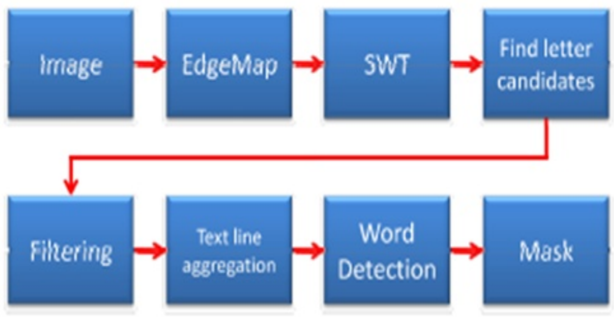


Fig. 1 SWT Algorithm

B. Steps in SWT

1) *Stroke width transform:* The Stroke Width Transform is a local image operator which computes per pixel the width of the most likely stroke containing the pixel. Two neighboring pixels may be grouped together if they have similar stroke width.

2) *Finding the letter candidate:* The output of the SWT is an image where each pixel contains the width of the most likely stroke it belongs to. Which computes per pixel the width of the most likely stroke containing the pixel. This is to group these pixels into letter candidates. Two neighboring pixels may be grouped together if they have similar stroke width.

3) *Grouping letter into Text Lines:* Grouping the letters have similar stroke width, letter width, height and spaces between the letters and words.

C. Advantages and Disadvantages

- One of the fast method.
- Reliable and flexible method for text detection.
- This is not considering the direction of stroke so curved text line not detected.

III. A LAPLACIAN METHOD FOR VIDEO TEXT DETECTION

A. Overview

An efficient text detection method based on the Laplacian operator. The maximum gradient difference value is computed for each pixel in the Laplacian-filtered image. K-means is then used to classify all the pixels into two clusters: text and non-text. For each candidate text region, the corresponding region in the Sobel edge map of the input image undergoes projection profile analysis to determine the boundary of the text blocks. Finally, we employ empirical rules to eliminate false positives based on geometrical properties.

Text detection methods can be classified into three approaches: connected component based, edge based and texture-based. The first approach does not work well for all video images because it assumes that text pixels in the same region have similar colors or grayscale intensities. The second approach requires text to have a reasonably high contrast to the background in order to detect the edges. So these methods often encounter problems with complex backgrounds and produce many false positives. Finally, the third approach considers text as a special texture and thus, uses fast Fourier transform, discrete cosine transform, wavelet decomposition and Gabor filters for feature extraction.

B. Steps in SWT

1) *Text Detection:* Calculate maximum gradient difference, defined as the difference between the maximum and minimum values within a local 1 x N window. The MGD value at pixel (i, j) is computed form image f as follows.

$$MGD(i, j) = max(f(i, j - t)) - min(f(i, j - t))$$

Text regions typically have larger MGD values than non-text regions and use K-means to classify all the pixels into two clusters, text and nontext.

2) *Boundary Refinement:* Compute the binary Sobel edge map SM of the input image. The horizontal projection profile is defined as follows.

$$HP(i)=\sum SM(i, j)$$

If HP is greater than a certain threshold, then it is a text line; otherwise, it is gap between different text lines or words.

C. Advantages and Disadvantages

- Efficient method.
- Get accurate boundary of each text block.
- Arbitrary orientation texts are not detected.

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

IV. Text String Detection from Natural Scenes by Structure Based Partition and Grouping

A. Overview

Text information in natural scene images serves as important clues for many image based applications such as scene understanding, content-based image retrieval, assistive navigation, and automatic geocoding. However, locating text from complex background with multiple colors is a challenging task. This is a new framework to detect text strings with arbitrary orientations in complex natural scene images.

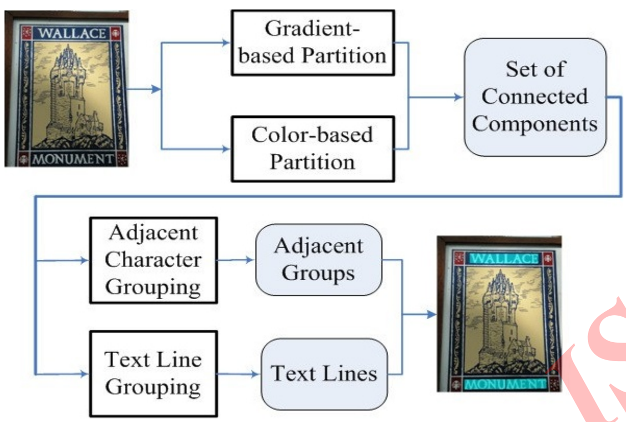


Fig. 3 Flowchart of the proposed framework of text string detection

B. Steps

The proposed framework consists of two steps Image partition and Character candidate grouping.

1) Image Partition: There is two methods to partition images. Gradient based method and colour based method. In gradient based, model the character by distribution of gradient magnitudes and stroke size including width, height and aspect ratio. In color based *to* label a region of connected pixels with similar colors as a connected component,develop color-based partition method.

2) *Character grouping*: There is two methods to grouping images. Adjacent character grouping and text line grouping. The adjacent character grouping method calculates the sibling groups of each character candidate as string segments and then merges the intersecting sibling groups into text string. The text line grouping method performs Hough transform to fit text line among the centroids of text candidates.

C. Advantages and Disadvantages

- Use gradient feature and color feature.
- Extract text strings with arbitrary orientations.
- No text extraction from complex backgrounds.

V. Scene Text Detection via Connected Component Clustering and Nontext Filtering

A. Overview

This is a new scene text detection algorithm based on two machine learning classifiers: one allows us to generate candidate word regions and the other filters out nontext ones. To be precise, we extract connected components (CCs) in images by using the maximally stable extremal region algorithm. These extracted CCs are partitioned into clusters so that we can generate candidate regions. Unlike conventional methods relying on heuristic rules in clustering, we train an AdaBoost classifier that determines the adjacency relationship and cluster CCs by using their pairwise relations. Then we normalize candidate word regions and determine whether each region contains text or not. Since the scale, skew, and color of each candidate can be estimated from CCs, we develop a text/nontext classifier for normalized images.

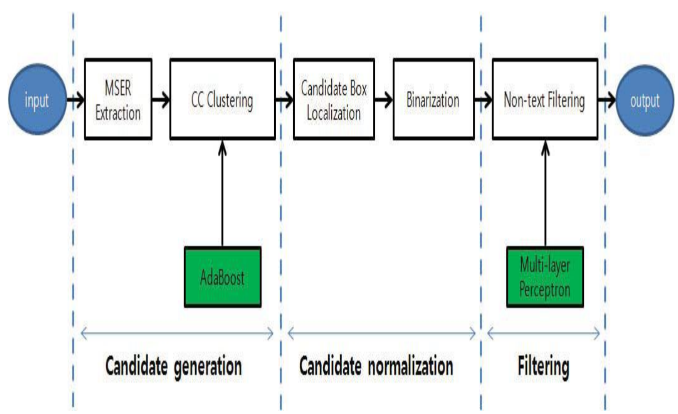


Fig. 4 Block diagram

B. Steps

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

The proposed framework consists of three steps Candidate generation, Candidate normalization, Filtering

1) *Candidate Generation*: For the generation of candidates,extract CCs in images and partition the extracted CCs into clusters, where clustering algorithm is based on an adjacency relation classifier.

2) *Charater Normalization*: In this first normalize the candidate and then binarize.

3) *Filtering:* Develop a text/nontext classifier that rejects nontext blocks among normalized images.

C. Advantages and Disadvantages

- It shows good performance
- Small computation cost

VI. LOCALIZING TEXT IN SCENE IMAGES USING BOUNDARY CLUSTERING STROKE SEGMENTATION AND STRING FRAGMENT CLASSIFICATION

A. Overview

Localizing text in scene images extract text regions from scene images with complex backgrounds and multiple text appearances. This framework consists of three main steps: boundary clustering (BC), stroke segmentation, and string fragment classification. In BC, we propose a new bigram color-uniformity-based method to model both text and attachment surface, and cluster edge pixels based on color pairs and spatial positions into boundary layers. Then, stroke segmentation is performed at each boundary layer by color assignment to extract character candidates. Propose two algorithms to combine the structural analysis of text stroke with color assignment and filter out background interferences. Further, we design a robust string fragment classification based on Gabor-based text features. The features are obtained from feature maps of gradient, stroke distribution, and stroke width. The proposed framework of text localization is evaluated on scene images, born-digital images, broadcast video images, and images

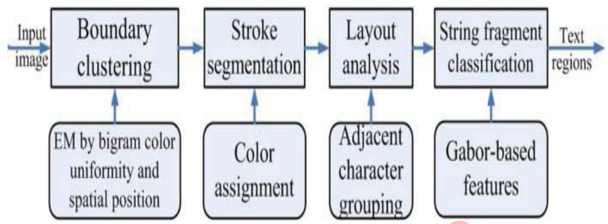


Fig. 5 Flowchart of text localization in natural scene

B. Steps

The proposed framework consists of three steps Boundary clustering, stroke segmentation, string fragment classification.

1) *Boundary clustering*: Text boundaries on the border of text the extracted CCs and its attachment surface are described by characteristic color pairs. To extract text by distinguishing the boundaries of characters and strings from those of background outliers based on color pairs.

2) *Stroke Segmentation*: Text boundaries are probably broken into tiny segments or connected into the boundary of a nontext background object. Character labeling uses stroke as basic unit. Stroke as a connected image region with uniform color and half-closed boundary, which keeps consistent distance in one direction while stays extensible in the perpendicular direction.

3) *String Fragment classification:* Develop a text/nontext classifier that rejects nontext blocks among normalized images.

C. Advantages and Disadvantages

- Complex background text extracted.
- Small computation cost

VII. CONCLUSION

Localize text regions is used among various scene images, born-digital images, broadcast video images, and blind-captured object images. Here different text detection techniqes surveyed. Localizing text used for assistive reading, geocoding, and content-based image retrieval.

REFERENCES

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

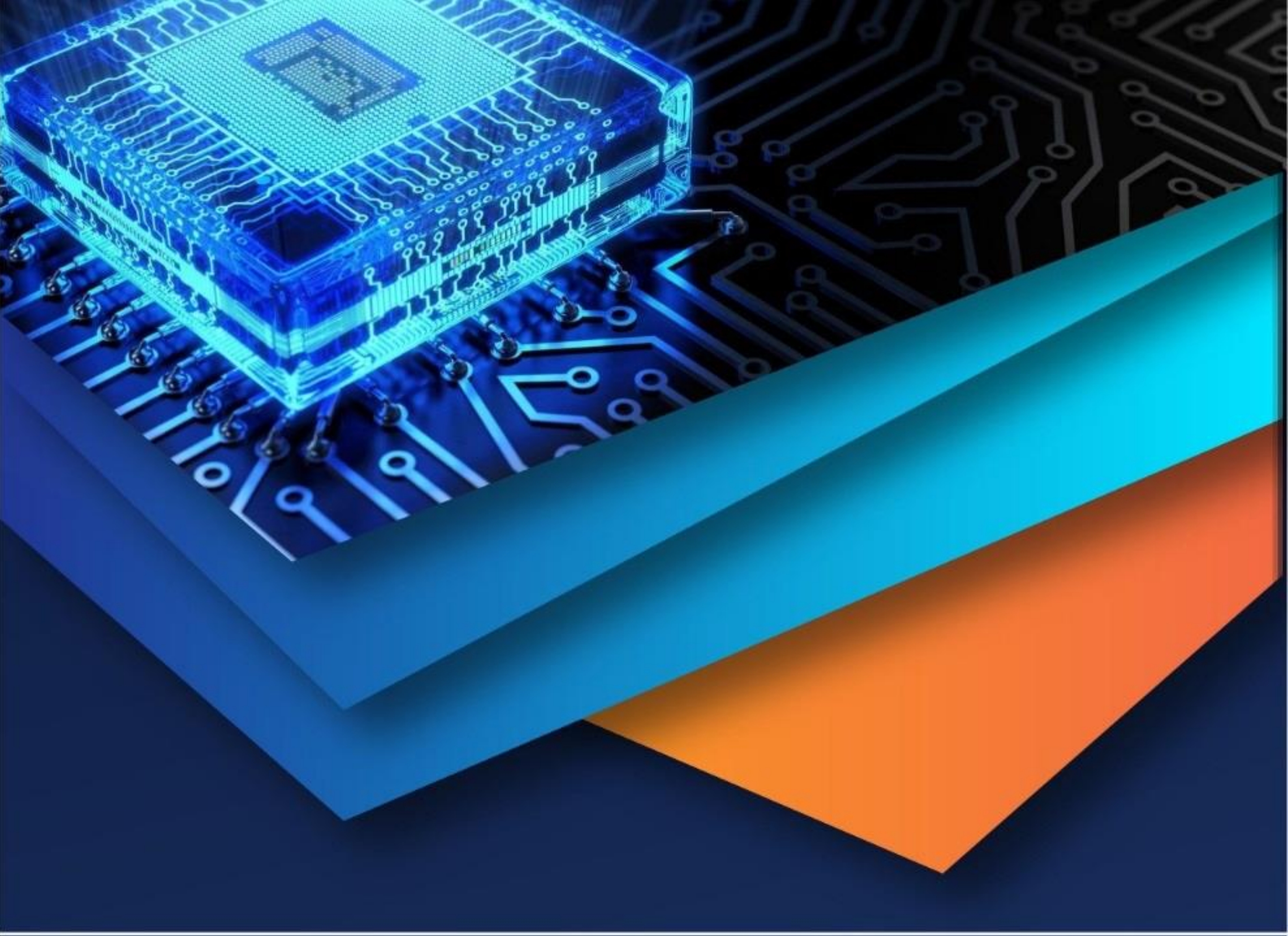
[1] B. Epshtein, E. Ofek, and Y. Wexler, "*Detecting text in nature scenes with Stroke width transform"*, in Proc. IEEE Conf. Comput. Vision Pattern Recogn, Jun. 2010, pp. 2963-2970.

[2] T. Phan, P. Shivakumara, and C. L. Tan, "*A Laplacian method for video text detection"*, in in Proc. 10th Int. Conf. Document Anal. Recogn.,2009, pp. 66-70

[3] C. Yi and Y. Tian, "*Text string detection from natural scenes by structure based partition and grouping"*, in IEEE Trans. Image Process, . 20, no. 9, pp. 2594-2604, Sep. 2011.

[4] 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.

[5] Chucai Yi and Yingli Tian Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, and String Fragment Classification IEEE Trans. Image Process. VOL. 21, NO. 9, SEP 2012



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY