Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 9, September 2014, pg.663 – 669

RESEARCH ARTICLE

Scenetext Detection and Recognition in Natural Images with Maximally Stable Extremal Regions and Stroke Width Transform

Sreeja K Soman, Sindhu M.P

M-Tech Student, Sree Narayana Gurukulam College of Engineering and Technology, Kerala

Assoc. Prof, Sree Narayana Gurukulam College of Engineering and Technology, Kerala

sreejaksoman@gmail.com, sindhuvinu@gmail.com

Abstract- Text detection and Recognition from natural scene images are useful in many applications. In this paper, we propose a novel CC-based text detection algorithm, which employs Edge-Enhanced Maximally Stable Extremal Regions (MSER) and stroke Width transform. We extract connected components (CCs) in images by using the edge-enhanced maximally extremal region algorithm then these candidates are filtered using CC based analysis and stroke width filtering. For each non horizontal text perform text rotation to make it as a horizontal text and apply boundary box localization. The detected text can be recognized using an Optical character recognizer. Experimental results on ICDAR 2005 robust reading competition datasets show that our method yields better performance of the proposed method.

Keywords— Text Recognition, maximally stable extremal regions, connected component analysis, stroke width filtering, non-horizontal text

I. INTRODUCTION

Detecting text in natural scene image has high purpose and contains more information related to the place and helps us to understand the objective more easily. Text Detection and Recognition involves detection, localization, tracking, extraction, enhancement and recognition of the text. The scene text information is easily recognized by machines and can be used in a variety of applications such as visually impaired people, translators for tourists, information retrieval systems in indoor and outdoor environments, and automatic robot navigation. Text detection has been considered as a challenging task due to the wide variety of text appearances, such as variations in font and style, geometric and photometric distortions, partial occlusions, and different lighting conditions.

Fig1: Examples of text in natural scene images

Number of Automatic detection and translation of text in images done using different techniques proposed. Generally, Text detection methods can be divided into two categories: Edge-based methods and connected component (CC)-based methods. Region-based methods approved a sliding window scheme, which is basically a brute force scheme which requires a lot of local decisions. Therefore, the region-based methods have attentive on an efficient binary classification (text versus nontext) of a small image area. In other words, they have focused on to determine whether a given patch is a part of a text region. Limitations of Edge-based methods are high computational complexity and the difficulty to select the best features for scene text detection.

On the other hand CC- based methods are simple and efficient text detection approach.CC methods generated separate CC regions. CC-based methods use a bottom-up approach by grouping small components into sequentially larger components until all regions are identified in the image. A geometrical analysis is needed to join the text components using the spatial arrangement of the components so as to filter out non-text components and mark the boundaries of the text regions

II. SYSTEM OVERVIEW

Fig.2 shows the basic block diagram of our method, which contains four steps: candidate generation,text/non textfiltering,text string rotation,text region localization and text region recoginition.The proposed system is also used to detect text in non horizontal orientation.this can be implemented by rotating the text string. Our candidate generation method consists of a MSER-based CC extraction block. Despite their promising properties,MSER have been reported to be delicate to image blur, which can be avoid using combination of MSER and Canny edge detection algorithm

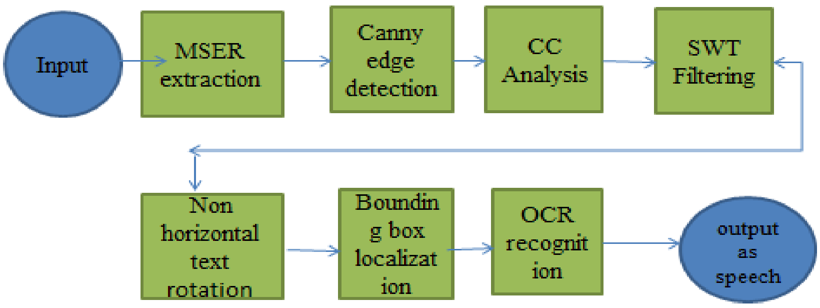


Fig2: Block diagram of our system.

In our method propose two level text/nontext filtering using cc analysis and a novel image operator stroke width transform. The proposed method is work on non-horizontal text detection by rotating the text to make it as a horizontal text. Then the text regions are localized using parallelograms. Detected text regions are recognized using OCR recognizer.

A.*Candidate Generation*

For the generation of candidates, we extract CCs in images. This step includes Detect MSER region, canny edge detection and CC analysis.

1. *MSER Extraction*

In computer vision, maximally stable extremal regions (MSER) are used as a method of blob detection in images. Informally, a blob is a region of a digital image in which some properties are constant or vary within a prescribed range of values; all the points in a blob can be considered in some sense to be similar to each other. The original algorithm is proceeds by first sorting the pixels by intensity. After sorting, pixels are marked in the image, and the list of growing and merging connected components and their areas is maintained using the union-find algorithm. In practice these steps are very fast. During this process, the area of each connected component as a function of intensity is stored producing a data structure.

A merge of two components is viewed as termination of existence of the smaller component and an insertion of all pixels of the smaller component into the larger one. In the extremal regions, the 'maximally stable' ones are those corresponding to thresholds where the relative area change as a function of relative change of threshold is at a local minimum, i.e. the MSER are the parts of the image where local binarization is stable over a large range of holds have consistent color, we begin by finding regions of similar intensities in the intensities in the image using the MSER region detector.

Fig 3:MSER regions in an image

The MSER algorithm can be divided into four major parts:

1) Pre-processing. Pixels are sorted in intensity order, and the number of pixels for each intensity is determined.
2) Clustering: A representation of all regions at each intensity level is created.
3) MSER detection. The size, |Q|, of all regions are tracked and the growth rate, q are monitored for local minimums
4) Display result. All pixels belonging to a detected MSER are identified and presented as an output.

2. *Canny Edge Detection*

Edge detection is the name for a set of mathematical methods which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. Since written text is typically placed on clear background, it tends to produce high response to edge detection

Canny edge detection is a four step process.

1. A Gaussian blur is applied to clear any speckles and free the image of noise.
2. A gradient operator is applied for obtaining the gradients' intensity and direction.
3. Non-maximum suppression determines if the pixel is a better candidate for an edge than its neighbors.
4. Hysteresis thresholding finds where edges begin and end

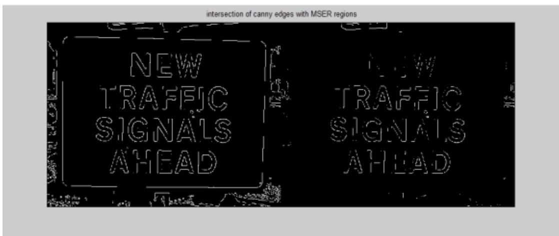


Fig4: canny edges with MSER region

An intersection of MSER regions with the edges is going to produce regions that are even more likely to belong. The original MSER regions in MSER Mask still contain pixels that are not part of the text. We can use the edge mask together with edge gradients to eliminate those regions. Grow the edges outward by using image gradients around edge locations. This mask can now be used to remove pixels that are within the MSER regions but are likely not part of text.

3. CC Analysis

Once region boundaries have been detected, it is often useful to extract regions which are not separated by a boundary. Any set of pixels which is not separated by a boundary is call connected. Each maximal region of connected pixels is called a connected component. Some of the remaining connected components can now be removed by using their region properties. The thresholds used below may vary for different fonts, image sizes, or languages

The basic steps in finding the connected components are:

1. Search for the next unlabeled pixel, p.
2. Use a flood-fill algorithm to label all the pixels in the connected component containing p.
3. Repeat steps 1 and 2 until all the pixels are labeled.

Fig5: connected component analysis output

B. STROKE WIDTH FILTERING

Stroke width is defined as the length of a straight line from a text edge pixel to another along its gradient direction. The basic motivation of our stroke width extraction algorithm is that stroke width almost remains the same in a single character; however, there is significant change in stroke width in non-text regions as a result of their irregularity.

Fig6: Stroke Width Filtering

The initial step of stroke width extraction is to get skeletons of MSERs remained. On every foreground pixel on the skeleton, distance transform is applied to compute the Euclidean distance from this pixel to the nearest boundary of the corresponding MSER. Then we obtain a skeleton-distance map. Characters in most languages have a similar stroke width or thickness throughout. It is therefore useful to remove regions where the stroke width exhibits too much variation.

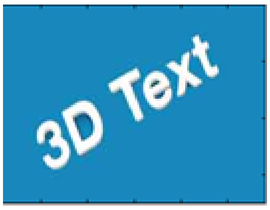
Algorithm 1 Finding stroke width

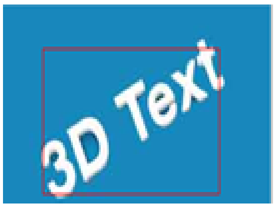
```
Input: binary image BW
Output: stroke width image SW
 D: = Distance Transform (BW);
 D: = round (D);
For p = each foreground pixel in D do
          PVal: = D (p);
            Lookup (p):= p's 8 neighbors whose    value< PVal;
End for
MaxStroke: = max (D);
For Stroke = MaxStroke to 1 do
    StrokeIndex: = find (D==Stroke);
    NeighborIndex = Lookup (StrokeIndex);
    While NeighborIndex not empty do
          D(NeighborIndex):=Stroke;      NeighborIndex:= Lookup(NeighborIndex);
End while
End for
Return SW: = D;
```

C. TEXT MASK DETECTION AND ROTATION

Mask is created for filtered output Image. Find the Connected Component of masked image. Connected component has the state orientation. If Orientation make any angle which is greater than 5 degree. Then following are performed

1. Read demo image.
2. Compute and display the histogram.
3. Get rid of holes inside.
4. Erode to shrink the mask and get rid of letters.
5. Mask the image
6. Get gradient
7. Shrink mask again to get rid of outer outline
8. Get histogram
9. Threshold to get a binary image
10. Get rid of holes inside.
11. Label the image
12. Make measurements of orientation
13. Find the largest blob
14. Pluck out largest one
15. Make measurements of orientation
16. If the angle is less than 5 then rotate image
17. Do text detection





Input image localizing text

D. BOUNDING BOX LOCALIZATION

To compute a bounding box of the text region, we will first merge the individual characters into a single connected component. This can be accomplished using morphological closing followed by opening to clean up any outliers

E. OPTICAL CHARACTER RECOGNITION

The segmentation of text from a muddled scene can greatly improve OCR results. Since our algorithm already produced a well segmented text region, we can use the binary text mask to improve the accuracy of the recognition results.

III.PERFORMANCE EVALUVATION

The performance degree used for text detection, which is easier to express than for localization and extraction, is the detection rate, defined as the ratio between the number of detected text frames and all the given frames containing text. Measuring the performance of text extraction is enormously difficult and until now there has been no comparison of the different extraction methods. Instead, the performance is merely inferred from the OCR results, as the text extraction performance is closely related to the OCR output. Traditionally in evaluation of object detection algorithms, for a single detection file and its corresponding ground truth file, two values, recall and precision, can be calculated. They are defined as follows:

$$Recall = \frac{correct\ Detected}{(Correct\ Detected + Missed\ Text\ Lines)}$$

$$\text{Precision}=\frac{Correct\ Detected}{(Correct\ Detected + False\ Positives)}$$

Table I EXPERIMENTAL RESULT OF THE ALGORITHM.

#images	150
# textlines	370
#correct detected	305
#false positives	40
Recall (%)	82.43%
Precision (%)	88.40%

IV. CONCLUSION

In this proposed system a new text detection system that detects and recognizes texts of arbitrary directions in natural images. In this paper CC-based text detection algorithm is proposed to overcome the difficulties of grouping the characters and remove false positives. The proposed system compares favorably with the state-of-the-art algorithms when handling horizontal texts and achieves significantly enhanced performance on texts of arbitrary 0rientations in complex natural scenes.

REFERENCES

[1] B.Epshtein,E.Ofek,and Y. Wexler, Detecting text in nature scenes with Stroke width transform, in in Proc. IEEE Conf. Computer Vision Pattern Recognition, Jun. 2010,pp. 2963_2970

[2] M. Basavanna , P. Shivakumara, S. K. Srivatsa, G. Hemantha Kumar,_Multi-Oriented Text Detection In Scene Images,_ in International Journal of Pattern Recognition and Artificial Intelligence , Vol. 26, No. 7 Dec 2012.

[3] Yao Li and Huchuan Lu “Scene Text Detection via Stroke Width” in 21st International Conference on Pattern Recognition (ICPR 2012) November 11-15, 2012. Tsukuba, Japan.

[4] Hyung Il Koo, Member, IEEE, and Duck Hoon Kim, Member, IEEE “Scene Text Detection via Connected Component Clustering and Nontext Filtering” in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 6, JUNE 2013

[5] D. Nistér and H. Stewénius, “Linear time maximally stable extremal regions,” in Proc. Eur. Conf. Comput. Vis., 2008, pp. 183–196.

[6] P.-E. Forssén, “Maximally stable colour regions for recognition and matching,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,Jun. 2007

[7] L Neumann and J. Matas. “A method for text localization and recognition in real-world images.” In ACCV, pages 770–783, 2010..

[8] Satadal Saha, Subhadip Basu, Mita Nasipuri and Dipak Kr. Basu, A Hough Transform form based Technique for Text Segmentation in Journal Of Computing, vol 2, Issue 2, February 2010.

[9] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong,and R. Young. Icdar 2003 robust reading competitions. In ICDAR, pages 682–687, 2003.

