Structure-Based Partition and Grouping For Text String Detection In Natural Scenes

Miss. Sonal S. Jadhav, Prof. Archana C.Lomte

1.Student, JSPM's Bhivarabai Sawant Institute of Technology and Research ,Wagholi, Pune, India

2.Asst Prof (computer) at Department Of Computer Engineering, , JSPM's Bhivarabai Sawant Institute of Technology and Research ,Wagholi, Pune, India

ABSTRACT

Text in natural image perform excellent task for many image based applications used in computer vision like scene understanding, content-based image retrieval, assistive navigation, and automatic geocoding.With these applications ,now a days text detection and recognition became challenging and trendy task. However, locating text from a complex background with multiple colors is a challenging task. Here, a new framework is explored to detect text strings with arbitrary orientations in complex natural scene images. The proposed framework of text string detection consists of two steps: 1) image partition to find text character candidates based on local gradient features and color uniformity of character components and 2) character candidate grouping to detect text strings based on joint structural features of text characters in each text string such as character size differences, distances between neighboring characters, and character alignment. We propose two algorithms of text string detection: 1) adjacent character grouping method and 2) text line grouping method. With adjacent character grouping sibling groups of each character candidate get calculated and then merges the intersecting sibling groups into text string. The text line grouping method used to perform transformation to fit text line among the centroids of text candidates. The fitted line describes orientation of potential text string. The detected text string is presented by a rectangle region covering all characters whose centroids are cascaded in its text line.

Keyword : - *Adjacent character grouping, character property ,image partition, text line grouping, text string detection.*

---

1. INTRODUCTION

Now a days, variety of text detection technique used for real world applications such as content based image retrieval, automatic geocoding and understanding scene, and text vision. Text detection and reorganization provides a way to directly access and utilize textual information in natural scenes ,which get attracted by computer vision based system and by document analysis community. To extract text information from camera-captured document images many algorithms and commercial optical character recognition systems have been developed .Different from document images, in which text characters are normalized into elegant poses and proper resolutions, natural scene images embed text in arbitrary shapes, sizes, and orientations into complex background. It is impossible to recognize text in natural scene images directly because the off-the-shelf OCR software cannot handle complex background interferences and non-orienting text lines. Thus, we need to detect image regions containing text strings and their corresponding orientations. With knowledge of text string orientations, we can normalize them to horizontal. However, the algorithms described in this paper will focus on text detection. Text detection related work focuses on text region initialization and extension by using distinct features of text characters.

1.1 Related work

To extract candidates of text regions, Kasar *et al.* first assigned a bounding box to the boundary of each candidate character in the edge image and then detected text characters based on the boundary model.Tran *et al.*calculated ridge points in different scales to describe text skeletons at the level of higher resolution and text orientations at the level of low resolution. . To group together text characters and filter out false positives, these algorithms employed similar constraints involved in character, such as the minimum and maximum size, aspect ratio, contrast between character strokes and background, and the number of inner holes. However, they usually fail to remove the background noise resulting from foliage, pane, bar, or other background objects that resemble text characters. To reduce background noise, the algorithms in the other category partition images to blocks and then groups the blocks verified by the features of text characters. Shivakumara *et al.* applied different edge detectors to search for blocks containing the most apparent edges of text characters. Lefevre *et al.* further designed a fusion strategy to combine detectors of color, texture, contour, and temporal invariance, respectively. Weinman *et al.* used a group of filters to analyze texture features in each block and joint texture distributions between adjacent blocks by using conditional random field. One limitation of these algorithms is that they used non content-based image partition to divide the image spatially into blocks of equal size before grouping is performed. Noncontact- based image partition is very likely to break up text characters or text strings into fragments which fail to satisfy the texture constraints. Thus, Phan *et al.* performed line-by-line scans in edge images to combine rows and columns with high density of edge pixels into text regions.

Fig. 1. Examples of text in natural scene images.

1.2 Existing system

The text extraction methods carry out based on image partition which leads to spatially dividing image blocks of equal size before grouping is performed .These leads to break up text characters into fragments which fails to satisfy the texture constraints. In addition, the color uniformity of text characters in natural scene image is taken into account for content-based partition . However the unexpected background noises might share the same colors with text characters, so texture features of characters are still required.
The algorithms in our proposed framework belong to this category of partition and grouping, but our content-based partition is involved in both gradient features and color features. Although many research efforts have been made to detect text regions from natural scene images, more robust and effective methods are expected to handle variations of scale, orientation, and clutter background.

2. OVERVIEW OF OUR FRAMEWORK

Here , a new framework is used to extract text strings with multiple sizes and colors, and arbitrary orientations from scene images with a complex and cluttered background. Fig. 2 depicts the flowchart of our framework. The proposed framework consists of two main steps, given here.
Step 1) Image partition to find text character candidates based on gradient feature and color uniformity.
In this step, two methods to partition scene images into binary maps of non overlapped connected components used : The first one is *gradient-based method* and second is *color-based method* .After these processes 1post processing is performed to remove the connected components which are not text characters.
Step 2) Character candidate grouping to detect text strings based on joint structural features of text characters in each text string such as character sizes, distances between two neighboring characters, and character alignment. In this step, two methods of structural analysis of text strings used :the first one is *adjacent character grouping method* and second one is *text line grouping method*.

The proposed framework is able to effectively detect text strings in arbitrary locations, sizes, orientations, colors and slight variations of illumination or shape of attachment surface.

• Most existing work of text detection from natural scene images focuses on detecting text in horizontal orientation or independent analysis of single character .We propose a new framework to robustly detect text strings with variations of orientation and scale from complex natural scene images with clutter background by integrating different types of features of text strings.

• We collect an oriented scene text dataset (OSTD) with text strings in arbitrary orientations, which is more challenging than the existing datasets for text detection.

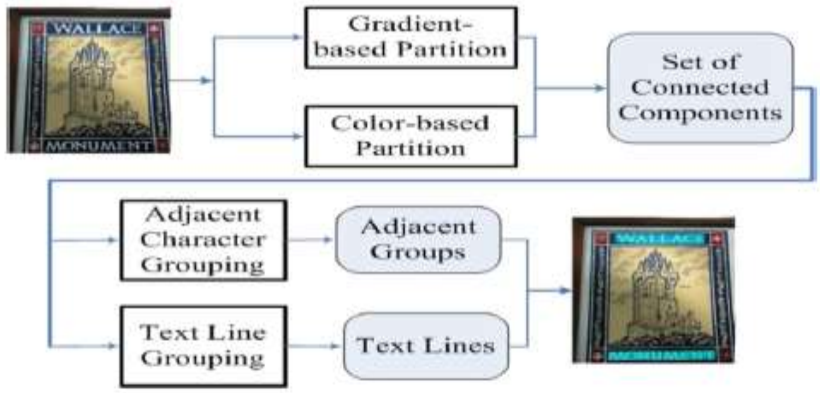

Fig. 2. Flowchart of the proposed framework of text string detection.

2.1 IMAGE PARTITION

A. Gradient-Based Partition by Connecting Paths of Pixel Couples

Although text characters and strings vary in font, size, color, and orientation, they are composed of strokes which are rectangle connected components with closed-width boundaries .In our method, each pixel is mapped to the connecting path of a pixel couple, defined by two edge pixels and on an edge map with approximately equal gradient magnitudes and opposite directions. Each pixel couple is connected by a path. Then the distribution of gradient magnitudes at pixels of the connecting path is computed to extract candidate character component.
A character boundary consists of a number of pixel couples. We model the character by distribution of gradient magnitudes and stroke size including width, height, and aspect ratio. The partitioned components are calculated from connecting path of pixel couple across the pixels with small gradient magnitudes.

For each pixel couple whose connecting path falls on this length range, we establish an exponential distribution of gradient magnitudes of the pixels on its connecting path, denoted by

$$g(G_{\mathrm{mag};\lambda}) = \lambda \exp(-\lambda G_{\mathrm{mag}})$$

At last, we perform morphological close and open as post processing to refine the extracted connected components, The refined connected components are taken as candidate character components.

B. Color-Based Partition by Color Reduction

In most scene images, text strings are usually composed of characters with similar colors. Thus, we can locate text information by extracting pixels with similar colors. To label a region of connected pixels with similar colors as a connected component, we develop color-based partition method. we perform color reduction by using color histogram and weighted K-means clustering through the following steps. First, a Canny edge detector is employed to obtain edge image. Second, we calculate color histograms of the original input image. To capture the dominant colors and avoid drastic color variations around edge pixels, only non edge pixels are sampled for color histogram calculation to obtain a set of sampled pixels . Third, after mapping all of the pixels from spatial domain to RGB color space, weighted K-means clustering is performed to group together the pixels with similar colors. By using the initial mean point which is randomly selected from the sampled pixels and an initial radius or clusters in RGB color space is established covering any pixel q whose color is close to pi

$$\mathrm{Cover}(q|p_i) = \begin{cases} 1, & IF\ K1 \text{ is satisfied} \\ 0, & \text{otherwise} \end{cases}$$
$$\mathrm{Cluster}(p_i) = \{q|\mathrm{Cover}(q|p_i) = 1\}.$$

The number of color layers depends on the number of dominant hues in original image and the initial radius . The larger the cluster radius is, the more pixels will be covered by each color cluster, so the total number of color clusters is reduced, which results in less color layers.

2.2 Connected Component Grouping

A. Adjacent Character Grouping

Text strings in natural scene images normally appears in alignment each text character in a text string must possess character siblings at adjacent positions. The structure features among sibling characters can be used to determine whether the connected components belong to text characters or unexpected noises. Here, five constraints are defined to decide whether two connected components are siblings of each other.

1) Considering the capital and lowercase characters, the height ratio falls between and 1/T1 and T1

2) Two adjacent characters should not be too far from each other despite the variations of width, so the distance between two connected components should not be greater than T 2 times the width of the wider one.

3) For text strings aligned approximately horizontally, the difference between -coordinates of the connected component centroids should not be greater than T3 times the height of the higher one.

4) Two adjacent characters usually appear in the same font size, thus their area ratio should be greater than 1/T4 and less than T4.

5) If the connected components are obtained from gradient based partition the color difference between them should be lower than a predefined threshold T5 because the characters in the same string have similar colors.

In our system, we set T1=T4=2 and T2=3,T3=0.5 AND T5=50. According to the five constraints, a left/right sibling Fl/Fr set is defined for each connected component C as the

set of sibling components located on the left/right of C. To extract regions containing text strings based on adjacent character grouping, we first remove the small connect components (area< Ts)from the set of connected components S. In our system, we set Ts=20. Then, a left-sibling set and a right-sibling set for each connected component C are initialized to record its sibling connected components on the left and right, respectively. For two connected components C' and C" , they can be grouped together as sibling components if the above five constraints are satisfied. When C' and C" are grouped together, their sibling sets will be updated according to their relative locations, that is, when C is located on the left of C" ,C" will be added into the right-sibling set of C', which is simultaneously added into the left-sibling set of C". The reverse operation will be applied when C is located on the right of C" . When a connected component corresponds to a text character, the five constraints ensure that its sibling set contains sibling characters rather than the foliage, pane or irregular grain.

B. Text Line Grouping

In order to locate text strings with arbitrary orientations, we develop text line grouping method. To group together the connected components which correspond to text characters in the same string which is probably non horizontal, we use centroid as the descriptor of each connected component. Given a set of connected component centroids, groups of collinear character centroids are computed, as shown in

$$M = \{m \mid C \in S \text{ and } m = \text{centroid}(C)\}$$
$$L = \{G \mid G \subseteq M, |G| \geq 3, \forall m_i, m_j, m_k \in G, \text{ they are character centroids and they are colinear.}\}$$

where M denotes the set of centroids of all of the connected components obtained from image partition, and L denotes the set of text lines which are composed of text character centroids in alignment.

A solution is to search for satisfied centroid groups in the power set of M, but the complexity of this algorithm will be $O(2^{|M|})$, where M represents the number of centroids in the set M . We design an efficient Algorithm to extract regions containing text strings.

3. Algorithm

Locating Text Strings by adjacent character groups

$S := S - \{C | C \in S, area(C) < T_s\}$;
for every connected component $C \in S$
    Initialize the sibling sets F_L and F_R;
endfor

for two connected components C and C' with sibling sets $F_L \cup F_R$ and $F_L' \cup F_R'$ respectively
    if $1/T_1 < height(C)/\text{height}(C') < T_1$
        & $D(centroid(C).x, centroid(C').x) \leq T_2 * \max\{width(C), width(C')\}$
        & $D(centroid(C).y, centroid(C').y) \leq T_3 * \max\{height(C), height(C')\}$
        & $1/T_4 < Area(C)/Area(C') < T_4$
        & difference of mean RGB color value is less than T_5
            if $centroid(C).x \leq centroid(C').x$,
                $F_R := F_R \cup \{C'\}$; $F_L' := F_L' \cup \{C\}$;
            else
                $F_L := F_L \cup \{C'\}$; $F_R' := F_R' \cup \{C\}$;
            endif
    endif
endfor

for every connected component C
    if $F_L \neq \emptyset$ & $F_R \neq \emptyset$ & $||F_R| - |F_L|| \leq 3$
        $SG(C) := F_L \cup F_R \cup \{C\}$;
    endif
endfor

Repeat the following until no merge is performed and the rest sibling groups will be upgraded to adjacent character groups
    for two sibling groups $SG(C_1)$ *and* $SG(C_2)$
        if $|SG(C_1) \cap SG(C_2)| \geq 2$
            $SG(C_1) := SG(C_1) \cup SG(C_2)$; $SG(C_2) := \emptyset$;
        endif
    endfor

Filter out false positives by the three filters decided by the area, distance and stroke width respectively.

Calculate extracted regions based on the adjacent character groups.

---

corresponding connected components are smaller than the predefined threshold Ts . Then, three points ,mi,mj,mk are randomly selected from the set M to form two line segments.

3.1 Performance Evaluation

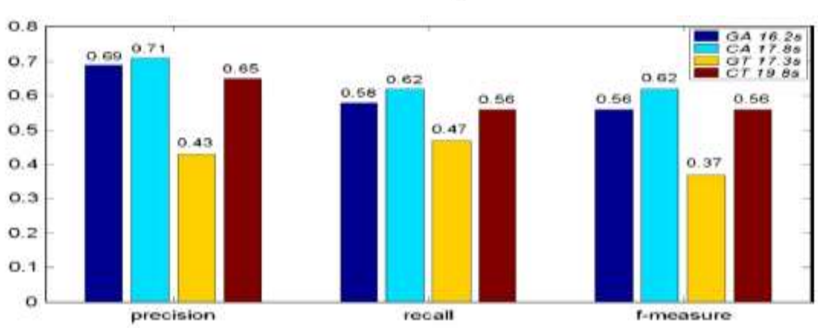

Fig.3. Performance evaluation of datasets.

The performance evaluation of four combinations of partition and grouping on robust reading dataset is shown in figure. The text detection in each image indicates that new method have good performance over other systems.

3.2 Results

By comparison with the algorithms presented in the text locating competition in ICDAR, the precision of our algorithm achieves the first rank while the recall and -measure is comparable with the algorithms with the high performance, as shown in Table. As a ruled-based algorithm, no trained classifiers can be applied directly, so it takes more time to perform the text string detection than the learning-based algorithms when not taking into account the time spent on training.

	Precision	Recall	*f*-measure
Ours	**0.71**	**0.62**	**0.62**
H. Becker	0.62	0.67	0.62
A. Chen	0.60	0.60	0.58
Ashida	0.55	0.46	0.50
HWDavid	0.44	0.46	0.45
Q. Zhu	0.33	0.40	0.33
Wolf	0.30	0.44	0.35
J. Kim	0.22	0.28	0.22
Todoran	0.19	0.18	0.18
N. Ezaki	0.18	0.36	0.22

Table 1.comparision between our algorithm and text detection algorithms

4. CONCLUSIONS

As text may appear in any variation and due to complex background text detection remains as a tedious work. To locate text in natural image a new way used based on image partition and connected components grouping. First, with gradient feature and color feature candidate text character obtained from connected components then, character grouping is performed to combine the candidate text characters into text strings which contain at least three character members in alignment. The text line grouping is able to extract text strings with arbitrary orientations. The combination of color-based partition and adjacent character grouping gives the best performance, which outperforms the algorithms presented in ICDAR.

5. ACKNOWLEDGEMENT

The authors would like to thank both the researchers and publishers for providing their resources and making them easily available and teachers for their guidance. Finally, we would like to extend a heartfelt gratitude to friends and family members

6. REFERENCES

[1] Robust unsupervised segmentation of degraded document images with topic models. by T. J. Burns and J. J. Corso in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009

[2] Detecting text in nature scenes with stroke width transform, by B. Epshtein, E. Ofek, and Y. Wexler in *Proc IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010

[3] Contextual restoration of severely degraded document images,”. by J. Banerjee, A. M. Namboodiri, and C. V. Jawahar, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009

[4].“The OCRopus open source OCR system,” by T. M. Breuel in in *Proc.IS&T/SPIE 20th Annu. Symp.*, 2008,

BIOGRAPHIES

	**Sonal S.Jadhav,** M.E Pursuing (computer) from Department of Computer Engineering, Jayawant Shikshan Prasarak Mandal's (JSPM'S), Bhivarabai Sawant Institute Of Technology And Research (BSIOTR),Pune,India.Savitribai Phule Pune University, Pune, Maharashtra,India- 411007. Received B.E (computer) Degree from Dr.d.Y.Patil Collage of Engineering College Of Engineering,Kolhapur,India-416234. Shivaji University, Kolhapur, Kolhapur, Maharashtra, India.. Area of interest is,Cloud Computing,Network Security,Image processing.
	**Prof. Archana Lomte ,** Ph.DReceived(computer)Received M.Tech(computer) PG from Savitribai Phule Pune University,Pune,Maharashtra,India-411007. Current working as PG .Coordinator and Asst Prof (computer) at Department Of Computer Engineering, Jayawant Shikshan Prasarak Mandal's (JSPM'S), Bhivarabai Sawant Institute Of Technology And Research(BSIOTR), Pune,India.Savitribai Phule Pune University, Pune, Maharashtra, India-411007. Area Of Interest is Cloud Computing,Network Security,Image Processing