IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III (Sep – Oct. 2014), PP 50-56
www.iosrjournals.org

Connected Component Clustering Based Text Detection with Structure Based Partition and Grouping

[1]Feby Ashraf , M-Tech student and [2]Nurjahan V A,
Assistant professor
[1]Computer Science and Engineering Department, Ilahia College of Engineering and Technology, MG university, Kerala, India
[2]Computer Science and Engineering Department, Ilahia College of Engineering and Technology,MG university, Kerala, India

Abstract: *Extraction of text from natural scene images is a challenging problem because of its complex backgrounds and large variations of text patterns. In this paper, we presents an innovative scene text detection algorithm with the help of two machine learning classifiers: one for candidate generation and the other which filters out non text ones. And the enhancement technique followed by a text string detection with arbitrary orientations based on structure- based partition and grouping. To extracted the connected components (CCs) in images an algorithm us used ,popularly known as maximally stable extremal region algorithm. These extracted CCs are partitioned into clusters and generate candidate regions. However, the methods use AdaBoost classifier to training the samples and improve the text detection accuracy. The scale, skew, and color of each candidate can be estimated from CCs, and filtered the non text from the normalized images. To find the text string consists of two steps: A) Image partition to detect the text character candidates by using gradient magnitude of character components and B)Text character grouping to detect text strings by using structural analysis of text characters and then merges them into text string for example size of character,distance between neighboring characters.To improve efficiency and accuracy, our algorithms are carried out in multi-scales. The proposed system yield very high identification accuracy and take less time for detection as compared to the existing system.*

Keywords: *Connected component based approach, character grouping, image partition, text string detection.*

I. Introduction

TEXT in images contains most important information and is exploited in many content-based image and video applications, such as content-based web image search, video information retrieval, and mobile based text analysis and recognition, visually impaired people, translators for tourists, information retrieval systems in indoor and outdoor environments, and automatic robot navigation . Human computer interaction (HCI) attains wide precepts during the introduction of mobile devices equipped with high resolution digital cameras are widely available in many research activities. various scene text detection and recognition have received much attention for the last decades. Among them, text detection and recognition in camera based images have been considered as very important problems in computer vision community [1]–[2]. It is because the text data is easily recognized by machines and can be used in a variety of applications. Due to complex background, and variations of font, size, color and orientation, text detection in natural scene images has to be robustly detected before being recognized and retrieved.

With the increasing popularity of practical vision systems and mobile phones, text detection in natural scenes becomes a critical yet challenging task. This is because scene text images suffers primarily from photometric degradations and geometrical distortions so that many algorithms faced the accuracy and/or speed (complexity) issues . To extract scene text information from camera-captured document images (i.e., most part of the captured image contains well organized text with clean background), many algorithms and commercial optical character recognition (OCR) systems have been developed [1].

Scene images are often captured by cameras. Text appearing accidentally in an image that does not represent anything important related to the content of the image. Such texts are known as scene text. In contrast to scene text is not only an important source of information but also a significant entity for indexing and retrieval purposes. Natural scene images contain text information which is often required to be automatically recognized and processed.

It is impossible to recognize text in natural scenes 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 the text strings and their corresponding orientations. Although scene text detection has been studied extensively in the past but the problem remains unsolved. The difficulties mainly come from two aspects: (1) the diversity of the texts and (2) the complexity of the backgrounds.

Document analysis domain is not limited to documents any more – one can have photographs of vehicle number plates, street names, gas/electricity meters, and so on where automatic recognition of scene text is desired. The latter set of challenges constitute the area of scene text detection requiring the researchers to go beyond the traditional techniques for document image analysis to solve them. So we developed a system for the efficient scene text detection connected component clustering based partion and grouping method is proposed. Apart from the scene text image we need to detect image regions containing text strings and their corresponding orientations with in the complex background .

II. Related Works

Most of the text detection algorithms in the literature can be classified into two categories: texture-based and connected component (CC)-based method[1][2]. Texture-based approaches view text as a special texture that is distinguishable from the document image background. Here the features are extracted over a certain region and a classifier (trained using machine learning techniques or by heuristics) is employed to identify the existence of text. Connected component based methods extract character candidates text from images by connected component analysis followed by grouping character candidates into text; additional checks may be performed to remove false positives. More recently, Maximally Stable Extremal Regions (MSERs) [7] based methods, which can be categorized as connected component based method. Connected component analysis method is used to define the final binary images that mainly consist of text regions. After the CC extraction, CC-based approaches filter out non-text. Finally, CC-based approaches infer text blocks from the remaining CCs. This step is also known as text line formation, or text line grouping.

The region-based methods have focused on many binary classification text versus non text of a small image patch. In other words, they have focused on the following problem: 1) Problem (A): to determine whether a given patch is a part of a text region. Eventhough the Problem-(A) is still challenging. It is not straightforward even for human to determine the image patch when we do not have knowledge of text properties such as scale, skew, and color. Many Experimental results shows that this region based approach is efficient, however, it yields worse performance compared with CC-based approaches. CC-based methods begin with CC extraction and normalize text regions by processing only CC-level information. Therefore, they have focused on the following problems: 2) Problem (B): to extract text-like CCs,3) Problem (C) : to filter out non text CCs,4) Problem (D): to infer text blocks from CCs.Some of the works contributed to development of the present work are

Lee [4] applied a CC-based method to the detection and recognition of text on cargo containers, which can have uneven lighting conditions and characters with different sizes and shapes. Edge information is used for the CC generation. The difference between adjacent pixels is used to determine the boundaries of potential characters after quantizing an input text image. Local threshold values are then selected for each text candidate, based on the pixels on the boundaries. These potential characters are used to generate CCs with the same gray-level. Thereafter, several heuristics rules are used to filter out non-text components based on aspect ratio, contrast histogram, and run-length measurement. Despite the method could be effectively used in other domains, experimental results were only presented for cargo container images.

Kim et al. [5] used cluster-based templates for filtering out non-text components for multi-segment characters to alleviate the difficulty in defining heuristics for filtering out non-text components. A similar approach was also reported by Ohya et al. [3]. Cluster-based templates are used along with geometrical information, such as size, area, and alignment. They are constructed using a K-means clustering algorithm from actual text document images.

More recently, Maximally Stable Extremal Regions (MSERs)[7] have become one of the commonly used region detector because of their high repeatability and partly because they are somewhat complementary to many other commonly used detectors .As described above, MSER-based methods have demonstrated very promising complementary performance in many real projects. However, current MSER-based methods still have some key limitations, i.e., they may suffer from detecting of repeating components and also insufficient text candidates construction algorithms. In this section, we will review the MSER-based methods focusing on these two problems. The main advantage of MSER-based methods over traditional connected component based method is able to detect most characters even when the image is in low quality (low resolution, strong noises, low contrast, etc.

Wolf et al. [8] improved Otsu's method to binarize document text regions from background, followed by a sequence of morphological processing to reduce noise and correct classification errors. To group together text characters and filter out non text components, these algorithms employed similar constraints involved in character, such as the minimum and maximum size, aspect ratio, contrast between character text strokes and background. However they usually fail to remove the background noise resulted from foliage, pane, bar or other background objects that resemble text characters. To reduce background noise, the algorithms which partition images to blocks and then groups .

Weinman et al. [10] used a group of filters to the analysis of 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 into blocks of equal size before grouping is performed . Non-content-based image partition is very likely to break up text characters or text strings into fragments which fail to satisfy the texture constraints Recently, Epshtein et al. [6] designed a content-based partition named as stroke width transform to extract text characters with stable stroke widths. To find the value of stroke width is calculated for each image pixel, and demonstrate its use on the task of scene text detection in natural images.

III. Problem Domain

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. CC-based approaches have shown better performance than region-based ones, they usually suffer from the computational complexity. It is because their performances depend on the quality of CCs and they adopted sophisticated CC extraction and filtering methods.

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. Compared with the existing methods which focus on independent analysis of single character, the text string structure is more robust to distinguish background interferences from text information. And also used to determine whether the connected components belong to text characters or unexpected noises.

IV. Proposed System

The proposed method generally focus scene text detection from different document images. After the detection the more informative scene text image that is free from complex back ground, variation in font, size and orientation, color distortions, and noise be the result. The proposed method consists of five steps: candidate Generation, candidate normalization, non text filtering image partition and character candidate grouping. The figure shows the block diagram of our proposed method.

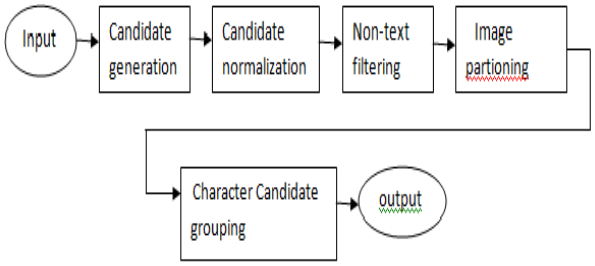

Figure 1: framework of our proposed method

our candidate generation method is based on popular CC based approaches which consists of a MSER-based CCextraction block and an AdaBoost-based CC clustering block. The maximally stable extremal region (MSER) algorithm is invariant to scales changes and affine intensity changes, and other blocks in our method are also designed to be invariant to these changes. In our method, both problems ((D) and (A)) are addressed based on machine learning techniques, so that our method is largely free from heuristics. We have trained a classifier that determines adjacency relationship between CCs for Problem-(D) and we generate candidates by identifying adjacent pairs. In training, we have selected efficient features and trained the classifier with the AdaBoost algorithm [9].

Figure 2: input image

Candidate normalization means not only geometric normalization but also binarization is also involved. We can localize character candidate regions with CC-level information and this localization allows us to build a simple but reliable text/nontext classifier. From that normalized image we built a binary image. Filtering means nothing but a text –nontext classifier is developed and rejects non-texts from normalized images.

The image partition creates a set of connected components from an input image, including both text characters and unwanted noises. Image partition is used to find text character candidates based on gradient features. Here to extract text information from complex background, image partition is first performed to group together pixels that belong to the same text character, obtaining a binary map of candidate character components. The gradient-based partition generates a binary map of candidate character components on black background. By the model of local gradient features of character stroke, we can filter out background outliers while preserving the structure of text character.

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. It also preserve multichannel information that is reduce the uncertainty and minimize redundancy in the output.

V. Solution Methodology

In the proposed method text detection approach consists of the following modules:

a) Candidate Generation

In the proposed method for the generating a candidates, extract CCs in images and partition the extracted CCs into set of clusters, using clustering algorithm is based on an adjacency relation classifier. In this section first explain the CC extraction method. Then, explain the approaches (i) building a training samples,(ii) to train the classifier, and (iii) using that classifier in CC clustering method.

i. CCs Extraction

The MSER algorithm[7] is the most efficient in CC extraction because it shows a good performance with a small computation cost . only the MSER algorithm could provide the stable binary results and also help us find most of the text components.

Figure 3: MSER results About MSER algorithm it is invariant to scales and affine intensity changes and other blocks and also design to invariant to these changes.

ii. Building Training sets

Our classifier is based on pairwise relations between CCs and CC pairs. Therefore, rather than focusing on this difficult problem, we address a relatively simple problem by adopting an idea of region-based approaches. If we have ci ~cj it will yields a character candidate components consisting of non text CCs and this candidate will be rejected at the non text rejection step. Also, we will perform word segmentation as a post processing step. Based on these observations, we build training sets. Specifically, we first obtain sets of CCs by applying the MSER algorithm to a training set released.

iii. Adaboost Learning & CC Clustering

Using the collected samples, we train an AdaBoost classifier that tells us whether it is adjacent or not. For these operations we shall define some local properties of CCs. We have used 6-dimensional feature vectors consisting of five geometrical features and one color- based feature. All of geometric features are designed to be

invariant to the scale of an input image and the color feature is given by the color distance between two CCs in RGB space. All of these features are informative and consider each feature as a weak classifier.From these weak classifiers, build a strong classifier with the AdaBoost learning algorithm. The AdaBoost is easy to implement and known to show good performance in many applications.
Based on the adjacency relations, C is partitioned into a set of clusters
W = {wk}
After clustering discarded the clusters having only one cc.

b) Candidate Normalization

After CC clustering, have a set of clusters are there. In this section, we will normalize corresponding regions for the reliable text/nontext classification.

I. Geometric normalization

Here first localize its corresponding region. Eventhough text boxes can experience perspective distortions, then approximate the shapes of text boxes with parallelograms whose left and right sides are parallel to y-axis. This approximation alleviates difficulties in estimating text boxes having a high degree of freedom (DOF).the normalization method only is to find a skew and four boundary supporting points. To estimate the scale and skew of a given word candidate wk , we build two sets:
$T_k = \{t(c_i) | c_i \in w_k\}$
$B_k = \{b(c_i) | c_i \in w_k\}$
where t (ci) and b(ci) are the top-center point and the bottom center point of a bounding box of ci , respectively. Then, perform geometric normalization by applying an affine transform mapping that transforms that corresponds region to a rectangle. During the transformation, use a constant target height and preserve the aspect ratio of the box.

II. Binarization

Given geometrically normalized images, build a binary images. In many cases, MSER results can be considered as binarization results. However, perform the binarization separately by estimating document text and background colors. It is because (i) the MSER results may miss some character components and/or yield noisy regions (mainly due to the blur) and (ii) it have to store the point information of all CCs for the MSER-based binarization

c) Non text filtering

A text/nontext classifier that rejects nontext blocks among normalized images. The main challenge of the approach is the variable aspect ratio. One possible approach to solve this problem is to split the normalized images into patches covering one of the letters and develop a character/non-character classifier. However, character segmentation is not an easy problem so split a normalized block into overlapping squares and develop a classifier that assigns a textness value to each square block. Finally, the decision results for all square blocks are integrated so that the original block is classified.

I. Feature Extraction from a Square Block

Here each square blocks can be divided into 4 horizontal and vertical ones and extract the features.For a horizontal block Hi (i = 1, 2, 3, 4), we consider
1) the number of white pixels,
2) the number of vertical white-black transitions,
3) the number of vertical black-white transitions
as features, and features for a vertical block is similarly defined.

II. Multilayer Perceptron Learning

For the training, need normalized images for this goal, applied the method (i.e., candidate generation and normalization algorithms) to the training images. Then ,manually classified them into text and nontext. And also discarded some images showing poor binarization results. However, the text/nontext images are divided into squares and have trained a multi-layer perceptron for the classification of square patches . To help the learning, input features are normalized.

D) Image partition

To extract text information from complex background, image partition is first performed to group together pixels that belong to the same text character, obtaining a binary map of candidate character

components. Based on local gradient features of text characters, we design a gradient-based partition respectively.

Although text character and string vary in font,size,color and orientation. figure shows Connected components with closed width boundaries and uniform intensities each pixel is mapped the stroke width in which it is located, and then the consistency of the stroke width is used to extract a candidate character component. In our proposed 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, as shown in Fig. Each pixel couple that is connected by a path. Then the distribution of gradient magnitudes at pixels of the connecting path is computed to extract candidate character component.

On the gradient map, G map(p) and dp are used respectively to represent the gradient magnitude and direction at pixel p. We take an edge pixel p from edge map as starting point and probe its partner along a path in gradient direction. If another edge pixel q is reached where gradient magnitudes satisfy

G map(p) -G map(q)<20 and directions satisfy dq-(dp-(dp/dp)*Π)< Π/6 , we obtain a pixel couple and its connecting path from p to q. This algorithm is applied to calculate connecting paths of all pixel couples.we establish an exponential distribution of gradient magnitudes of the pixels on its connecting path, denoted by g(G mag;λ)= λ exp(-λGmag)

E) Connected components grouping

The image partition creates a set of connected components from an input image, including both text characters candidates and unwanted noises. The text information appears as one or more text strings in most natural scene images, we perform heuristic grouping and structural analysis of text strings to distinguish connected components representing text characters from those representing noises. Assuming that a text string has at least three characters in alignment, we develop a methods to locate regions containing text strings. A connected component C is described by four metrics: height (), width (.), centroid (.), area (.). In addition, we use D (.) to represent the distance between the centroids of two neighbouring characters.

I. Adjacent character grouping

Text strings in natural scene images usually appear in alignment, namely, 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 T2 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 thanT3 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 because the characters in the same string have similar colors.

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. At each sibling group can be considered as a fragment of a text string. To create sibling groups corresponding to complete text strings, we merge together any two sibling groups . Repeat the merge process until no sibling groups can be merged together.

III. Conclusion and Future Work

We have presented in the paper an improved text string detection method which can effectively detect text from the document background. Due to the unpredictable text appearances and complex backgrounds, text detection in natural scene images is still an unsolved problem to locate text regions embedded in those images, In this paper, we have presented a novel scene text detection algorithm with the help of two machine learning classifiers: one for candidate generation and the other which filters out non text ones. And the enhancement technique followed by a text string detection with arbitrary orientations based on structure- based partition and grouping. we have presented an approach to detect, localize, and extract texts appearing in grayscale Images as well as locate text strings with arbitrary orientations.

Our future work will focus on developing learning based methods for text extraction from complex backgrounds and text normalization for OCR recognition.We also attempt to improve the efficiency and transplant the algorithms into a navigation system prepared for the wayfinding of visually impaired people.

References

[1]. K. Jung, "Text information extraction in images and video: A survey," Pattern Recognit., vol. 37, no. 5, pp. 977–997, May 2004.

[2]. J. Zhang and R. Kasturi, "Extraction of text objects in video documents: Recent progress," in Proc. 8th IAPR Int. Workshop Document Anal. Syst., Sep. 2008, pp. 5–17.

[3]. J. Ohya, A. Shio, and S. Akamatsu, Recognizing Characters in Scene Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (2) (1994) 214-224.

[4]. C.M. Lee, and A. Kankanhalli, Automatic Extraction of Characters in Complex Images, International Journal of Pattern Recognition Artificial Intelligence, 9 (1) (1995) 67-82.

[5]. E. Y. Kim, K. Jung, K. Y. Jeong, and H. J. Kim, Automatic Text Region Extraction Using Cluster-based Templates, Proc. of International Conference on Advances in Pattern Recognition and Digital Techniques, 2000, pp. 418-421.

[6]. B. Epshtein, E. Ofek, and Y. Wexler, "Detecting Text in Nature Scenes with Stroke Width Transform," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[7]. J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in Proc. Brit. Mach. Vis. Conf., 2002, pp. 384–393.

[8]. C. Wolf, J. M. Jolion, and F. Chassaing, "Text localization, enhancement and binarization in multimedia documents," In Proceedings of the International Conference on Pattern Recognition, Vol. 4, pp. 1037–1040, 2002.

[9]. J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," Ann. Stat., vol. 28, no. 2, pp. 337–407,1998.

[10]. J. Weinman, A. Hanson and A. McCallum, "Sign Detection in Natural Images with Conditional Random Fields," In IEEE International Workshop on Machine Learning for Signal Processing, 2004. approaches," Comput. Vis. Image Understand., vol. 91, pp. 6–21, 2003.