ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 4, October 2012

Review on Text String Detection from Natural Scenes

Rampurkar Vyankatesh Vijaykumar, Gyankamal J.Chhajed, Sahil Kailas Shah

Abstract*— In this paper we have reviewed and analyzed different methods to find strings of characters from natural scene images. We have reviewed different techniques like extraction of character string regions from scenery images based on contours and thickness of characters, efficient binarization and enhancement technique followed by a suitable connected component analysis procedure, text string detection from natural scenes by structure-based partition and grouping, and a robust algorithm for text detection in images. It is assumed that characters have closed contours, and a character string consists of characters which lie on a straight line in most cases. Therefore, by extracting closed contours and searching neighbors of them, character string regions can be extracted; Image binarization successfully processes natural scene images having shadows, non-uniform illumination, low contrast and large signal-dependent noise. Connected component analysis is used to define the final binary images that mainly consist of text regions. One technique chooses the candidate text characters from connected components by gradient feature and color feature. Color-based partition performs better than gradient-based partition, but it takes more time to detect text strings on each color layer. The text line grouping is able to extract text strings with arbitrary orientations. The combination of color-based partition and adjacent character grouping (CA) gives the best performance. Also in this paper, an efficient algorithm which can automatically detect, localize and extract horizontally aligned text in images with complex backgrounds is reviewed which is based on the application of a color reduction technique, a method for edge detection, and the localization of text regions using projection profile analysis and geometrical properties.

***Index Terms*— Closed Contour, Image Binarization, Connected Components Analysis, Adjacent Character Grouping, Image Partition, Text Line Grouping, Text String Detection, Color Reduction Technique, Edge Detection.**

I. INTRODUCTION

A lot of objects on which characters are written exist in our living environment. We humans get much information from these texts. It is expected that robots act in our living environment and support us in the future. If robots can read text on objects such as Packages and signs, robots can get information from them, and they can use it in their activation and support for us. Owing to the progress of OCR, computers have been able to read text in images. However, images have many non-character textures, and they make it difficult to read text by OCR. To cope with that problem, we need to extract character string regions from images. Indexing images or videos requires information about their content. This content is often strongly related to the textual information appearing in them, which can be divided into two groups: Text appearing accidentally in an image that usually does not represent anything important related to the content of the image. Such texts are referred to as *scene text*. Text produced separately from the image is in general a very good key to understand the image which is called *artificial text*. In contrast to scene text, artificial 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. Localization of text and simplification of the background in images is the main objective of automatic text detection approaches. However, text localization in complex images is an intricate process due to the often bad quality of images, different backgrounds or different fonts, colors, sizes of texts appearing in them. In order to be successfully recognizable by an OCR system, an image having text must fulfill certain requirements, like a monochrome text and background where the background-to-text contrast should be high. This paper strives toward methodologies that aid automatic detection, segmentation and recognition of visual text entities in complex natural scene images.

The algorithms of text extraction from images can be broadly classified under three types. They are gradient feature based, color segmentation based, and texture analysis based. The gradient feature based algorithm is based on the idea that pixels which have high gradient are the candidates of characters since edges exist between a character and background. One technique makes use of closed contours to extract string regions.

In this paper, Basilios Gatos proposes a novel methodology for text detection in natural scene images [2]. The proposed methodology is based on an efficient binarization and enhancement technique followed by a suitable connected component analysis procedure. Image binarization successfully processes natural scene images having shadows, non-uniform illumination, low contrast and large signal-dependent noise. Connected component analysis is used to define the final binary images that mainly consist of text regions.

In this paper, Chucai Yi and YingLi Tian propose a new framework to extract text strings with multiple sizes and colors, and arbitrary orientations from scene images with a complex and cluttered background [3]. The proposed framework consists of two main steps: a) image partition to find text character candidates based on gradient feature and

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 4, October 2012

color uniformity. In this step, Chucai Yi and YingLi Tian propose two methods to partition scene images into binary maps of non-overlapped connected components: *gradient-based method* and *color-based method* b) 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, Chucai Yi and YingLi Tian propose two methods of structural analysis of text strings: *adjacent character grouping method* and *text line grouping method*.

Also in this paper, an efficient algorithm which can automatically detect, localize and extract horizontally aligned text in images with complex backgrounds is presented which is based on the application of a color reduction technique, a method for edge detection, and the localization of text regions using projection profile analysis and geometrical properties.

In this way here we are going to discuss following algorithms to extract strings from natural scene images which is based on:

1) Closed contour method
2) Image binarization and connected component analysis
3) Structure-based Partition and Grouping
4) A color reduction technique, and a method of edge detection

Fig. 1. Examples of Text in Natural Scene Images [3]

I. CLOSED CONTOUR BASED METHOD [1]

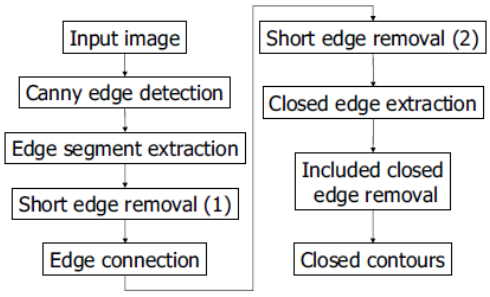

Fig. 2. The Process of Closed Contour Extraction [1]

One of the characteristics of common characters in real images is that most of them are capable of producing closed contour when edge extraction process is applied. So Tomohiro Nishino takes an approach to detect closed contours from images. Moreover, it is assumed that a character string consists of characters which lie on a straight line. From these assumptions, character string should be found from regions where closed contours are arranged with regularly. Fig.2 shows the procedure.

Assuming that characters included in a character string are aligned horizontally, string regions can be extracted by detecting horizontally aligned closed contours. Tomohiro Nishino explains how to detect the horizontally aligned closed contour. First, a circumscribed rectangle of a closed contour is calculated. Next, the rectangle is slid to right by some pixels as much as the width of the rectangle. If the rectangle includes the center of a circumscribed rectangle of another closed contour, these two closed contours are assumed to be aligned horizontally and to be included in the same character string. Closed contours which are isolated are assumed not to be characters. Circumscribed rectangles of each character string are assumed to be string regions. By this process, string regions of horizontally aligned closed contours are extracted. Both closed and unclosed contours which lie left or right of character string regions are extracted, and each thickness is calculated. These contours are added to the character string region if they have the similar thickness to that of characters in that region.

Advantages:

1) String regions can be extracted by detecting horizontally aligned closed contours.
2) However, not all of characters' contours are extracted as closed contours, so these unclosed contours and isolated closed contours are added to character string regions based on the thickness and character string regions are revised.

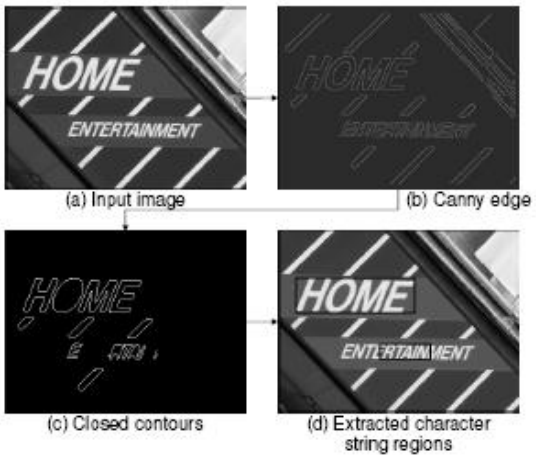

Fig. 3. The Result of Character String Extraction [1]

II. IMAGE BINARIZATION AND CONNECTED COMPONENT ANALYSIS BASED METHOD [2]

The proposed methodology for text detection natural scene images is based on an efficient binarization and enhancement technique followed by a connected component analysis procedure. The flowchart of the Proposed methodology is presented in Figures 4a and 4b. Starting from the scene image, Basilios Gatos produces gray level image and inverted gray level image. Then, calculate the two corresponding binary images using an adaptive binarization and image enhancement technique. In the sequel, the proposed technique involves a decision function that

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 4, October 2012

indicates which image between binary images contains text information. In Fig. 4a the original binary image is selected while in Fig. 5b the inverted binary image is selected. Finally, a procedure that detects connected components of text areas is applied.

Advantages:

1) Image binarization successfully processes natural scene images having shadows, non-uniform illumination, low contrast and large signal-dependent noise. Connected component analysis is used to define the final binary images that mainly consist of text regions.

2) The proposed methodology results in increased success rates for commercial OCR engines.

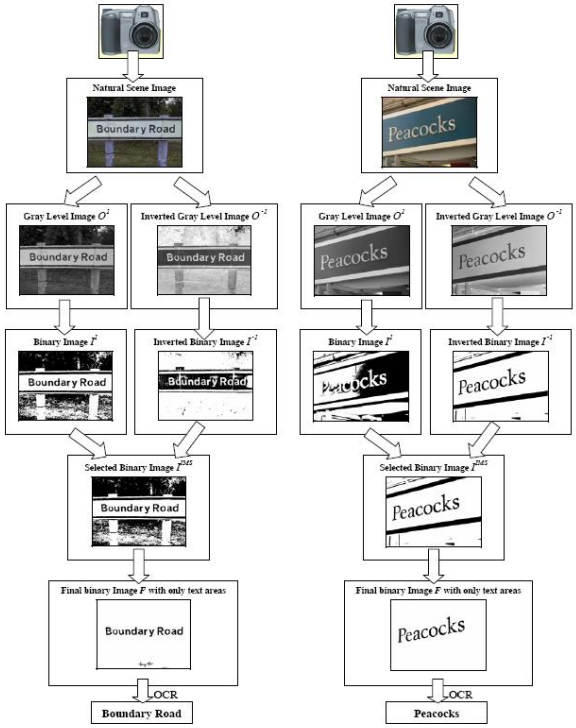

Fig. 4 A) Flowchart of the Proposed Method for Text Detection in Natural Scene Images (Original Binary Image Is Selected).

B) Flowchart of the Proposed Method for Text Detection in Natural Scene Images (Inverted Binary Image Is Selected). [2]

III. STRUCTURE-BASED PARTITION AND GROUPING BASED METHOD [3]

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, Chucai Yi and YingLi Tian propose two methods to partition scene images into binary maps of non overlapped connected components: gradient-based method and color-based method. A post processing is then performed to remove the connected components which are not text characters by size, aspect ratio, and the number of inner holes.

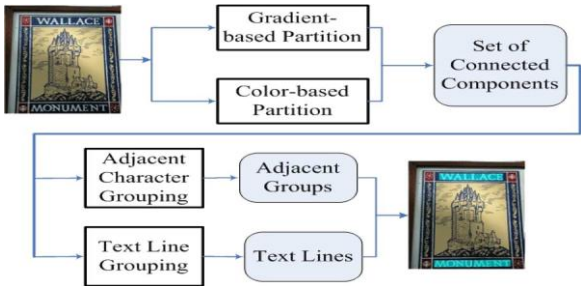

Fig. 5. Flowchart of the Framework [3]

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, Chucai Yi and YingLi Tian propose two methods of structural analysis of text strings: adjacent character grouping method and text line grouping method.

Adjacent Character grouping method:

Text strings in natural scene images usually appear 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 1/T1 and T1.

2) 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 Y-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 Ts.

In their system Chucai Yi and YingLi Tian set T1=T4=2, T2=3, T3=0.5, T5=40

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'.

To create sibling groups corresponding to complete text strings, Chucai Yi and YingLi Tian merge together any two

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 4, October 2012

sibling groups SG(C1) and SG(C2) when their intersection contains no less than two connected components. At this point, each sibling group can be considered as a fragment of a text string. Repeat the merge process until no sibling groups can be merged together. Text string in scene images can be described by corresponding adjacent character groups. To extract a region containing a text string, Chucai Yi and YingLi Tian calculate rectangle covering all of the connected components in the corresponding adjacent character group.

Advantages:

1) The structure features among sibling characters can be used to determine whether the connected components belong to text characters or unexpected noises.

2) Character grouping is performed to combine the candidate text characters into text strings which contain at least three character members in alignment.

Fig. 6. Two Detected Adjacent Character Groups Marked In Red and Green [3]

Text Line grouping method:

In order to locate text strings with arbitrary orientations, Chucai Yi and YingLi Tian 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, Chucai Yi and YingLi Tian use centroid as the descriptor of each connected component.

In order to locate text strings with arbitrary orientations, Chucai Yi and YingLi Tian 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, Chucai Yi and YingLi Tian 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…

M= {m | C ε S & m = centroid (C)}

L = {G | G is subset of M, |G|>=3, they are character centroids & are collinear}

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.

Chucai Yi and YingLi Tian design an efficient algorithm to extract regions containing text strings. At first, they remove the centroids from the set M if areas of their corresponding connected components are smaller than the predefined threshold Ts. Then, three points m_i, m_j, m_k are randomly selected from the set to form two line segments. They calculate the length difference, and incline angle difference between line segments mimj and mjmk as shown

$\Delta d = D(m_i, m_j) / D(m_j, m_k)$

$\Delta\Theta = |\Theta_{ij} - \Theta_{jk}|$, if $|\Theta_{ij} - \Theta_{jk}| <= \Pi/2$

$= |\Theta_{ij} - \Pi - \Theta_{jk}|$, if $|\Theta_{ij} - \Theta_{jk}| > \Pi/2$

The three centroids are approximately collinear if $1/T6 <= \Delta d <= T6$ and $\Delta\Theta <= T7$

Here T6=2, T7=Π/12

Thus, they compose a preliminary fitted line lu={mi, mj, mk} , u =index of fitted line. Other collinear centroids along lu can be added into the end positions to form a complete text string increasingly. For now, each text string is described by a fitted line. The location and size of the region containing a text string is defined by the connected components whose centroids are cascaded in the corresponding fitted line.

Advantages:

1) In order to locate text strings with arbitrary orientations, Chucai Yi and YingLi Tian develop text line grouping method.

2) To group together the connected components which correspond to text characters in the same string which is probably non horizontal, they use centroid as the descriptor of each connected component.

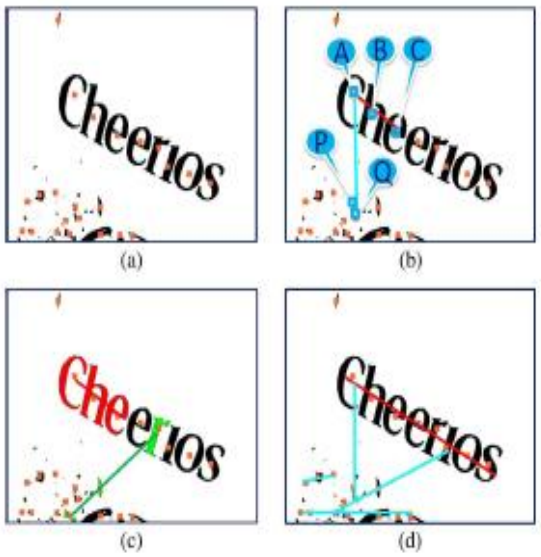

Fig. 7. Resulting Fitted Lines From Centroids Cascading. Red Line Corresponds To Text Region While Cyan Lines Are False Positives to Be Removed [3]

IV. A COLOR REDUCTION TECHNIQUE, AND AN EDGE DETECTION BASED [4]

An efficient algorithm which can automatically detect, localize and extract horizontally aligned text in images (and

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 4, October 2012

digital videos) with complex backgrounds is presented. The proposed approach is based on the application of a color reduction technique, a method for edge detection, and the localization of text regions using projection profile analyses and geometrical properties.

The different steps of our approach are as follows.

Step 1: Image Preprocessing-If the image data is not represented in YUV color space, it is converted to this color space by means of an appropriate transformation. After that, luminance value thresholding is applied to spread luminance values throughout the image and increase the contrast between the possibly interesting regions and the rest of the image.

Step 2: Edge Detection-This step focuses the attention to areas where text may occur. Julinda Gllavata employs a simple method for converting the gray-level image into an edge image. As a result, all character pixels as well as some non-character pixels which also show high local color contrast are registered in the edge image. In this image, the value of each pixel of the original image is replaced by the largest difference between itself and its neighbors (in horizontal, vertical and diagonal direction). Despite its simplicity, this procedure is highly effective. Finally, the contrast between edges will be increased by means of a convolution with an appropriate mask.

Step 3: Detection of Text Regions- The horizontal projection profile of the edge image is analyzed in order to locate potential text areas. Since text regions show high contrast values, it is expected that they produce high peaks in horizontal projection.

Step 4: Enhancement and Segmentation of Text Regions-First, geometric properties of the text characters like the possible height, width, and width to height ratio are used to discard those regions whose geometric features do not fall into the predefined ranges of values. All remaining text candidates undergo another treatment in order to generate the so-called text image where detected text appears on a simplified background. The binary edge image is generated from the edge image, erasing all pixels outside the predefined text boxes and then binarizing it. This is followed by the process of gap flling. If one white pixel on the binary edge image is surrounded by two black pixels in horizontal, vertical or diagonal direction, then it is also filled with black. The gap image is used as a reference image to refine the localization of the detected text candidates. Text segmentation is the next step to take place. Then, the segmentation process concludes with a procedure which enhances text to background contrast on the text image.

Advantages:

1) The input to our system can be a gray scale or a color image.

2) The current version can only detect texts with a horizontal Alignment.

Fig. 8. An Image from the "Commercial Test Set" [4]

Fig. 9. The Text Detection Result For the Image from the "Commercials Test Set" [4]

V. CONLCUSION

In this paper we have reviewed and analyzed different methods to find strings of characters from natural scene images. We have reviewed different techniques like extraction of character string regions from scenery images based on contours and thickness of characters, efficient binarization and enhancement technique followed by a suitable connected component analysis procedure, text string detection from natural scenes by structure-based partition and grouping, and a robust algorithm for text detection in images. Also, we have presented an approach to detect, localize, and extract texts appearing in grayscale or color Images as well as locate text strings with arbitrary orientations.

REFERENCES

[1] Extraction of Character String Regions from Scenery Images Based on Contours and Thickness of Characters in MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN.

[2] Towards Text Recognition in Natural Scene Images by Basilios Gatos, Ioannis Pratikakis and Stavros Perantonis.

[3] Text String Detection from Natural Scenes by Structure-based Partition and Grouping by Chucai Yi and YingLi Tian, Senior Member, IEEE

[4] A Robust Algorithm for Text Detection in Images by Julinda Gllavata, Ralph Ewerth and Bernd Freisleben

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

[6] X. Chen, J. Yang, J. Zhang, and A. Waibel, “Automatic detection and recognition of signs from natural scenes,” IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 4, October 2012

AUTHOR BIOGRAPHY

Vyankatesh V. Rampurkar has received his Bachelor of Engineering in 2008 from VPCOE, BARAMATI. He is now pursuing ME degree, in Computer Engineering at VPCOE, BARAMATI. Vyankatesh has participated and received certificate of participation in workshop on "Database Management Systems" and "Writing Effective Conference Papers" organized by MHRD and IIT, Bombay. He is life member of ISTE.
Lecturer,B.E.(Computer),M.E.(Computer)Appear

Research Interest are Image processing applications for finding hidden data features, Data Mining and applications. harshaldada@gmail.com

Gyankamal obtained Engineering degree (B.E.) in Computer Science and Engineering in the year 1991-95 from S.G.G.S.I.E.T, Nanded and Postgraduate degree (M.Tech.) in Computer Engineering from College of Engineering, Pune (COEP) in the year 2005-2007 . She is approved Undergraduate and Postgraduate teacher of Pune university and having about 16 yrs. of experience.

Gyankamal authored a book and has 10 publications at the national, international level for Conferences and Journal. She judged paper presentation event at National Conference , delivered expert talk and reviewed paper for IEEE international conference . She is life member of the ISTE & International Association IACSIT.

Research Interests:

Gyankamal is working in the research area of Steganography and Watermarking , specially for Black and White images. Efficient and more data embedding capacity with minimum distortion is her key goal of the research.

Selected Publications:

1. "Review on Binary Image Steganography and Watermarking" .Paper published in international Journal IJCSE Vol.3 11, November 2011.
2. "Steganography in Black and White Picture Images", 978-0-7695-3119-9/08 DOI 10.1109/CISP.2008.626 pg. no. 141-144 IEEE Computer Society. at International Conference CISP 2008 , China.
3. " Efficient Embedding in B&W picture images " with 978-1-4244-5265-1/10 page no 525-528. Paper published at 2nd IEEE International conference ICIME 2010 held in Chengdu , China , 16-18 April 2010.

Assistant Professor and HOD
B.E.(C.S.E.), M.Tech.(C.E.)
gjchhajed@vpcoe.org

Sahil K. Shah has obtained his Bachelor of Engineering in Information Technology in 2010 from Vidya Pratishthan' College of Engineering, Baramati; Pune University. Currently he is pursuing M.E. in Computer Engineering from Vidya Pratishthan's College of Engineering, Baramati; Pune University. Sahil is having 1.5 years of teaching Experience at undergraduate level. Research Interests: His research interests span Information Retrieval, Web Data Mining and Image Processing. He is working on various Topics related to above fields including Semantic similarity and multi document summarization of text documents.

Assistant Professor
B.E.(IT), M.E.(computer engg. Appearing)
sahilshahwnr@gmail.com