Text Extraction from Images with Complex Background

Rakshith Ranjan
Department of ISE,
New Horizon College of Engineering,
Bangalore, India

Preethi Venugopal
Department of ISE,
New Horizon College of Engineering,
Bangalore, India

Prithvi S
Department of ISE,
New Horizon College of Engineering,
Bangalore, India

Priyanka N S
Department of ISE,
New Horizon College of Engineering,
Bangalore, India

Abstract—**Extraction of text with complex background is a thriving field in Image Processing since there is a tremendous growth in the usage of images and videos in databases. Extraction of text is a two stage process that encompasses Text Detection and Text Extraction. In this paper, we explore novel methods for text detection and extraction. During the detection stage, edge maps are obtained for the images and K-means algorithm is applied to the average edge map. Morphological operations like opening and dilation are employed. Lastly, false positive filtering is executed to eliminate false positives. During the extraction stage, an adaptive seed fill method and an adaptive thresholding method is used to obtain two binary images that are fused to obtain a final binary image. This image is the result of extraction. Text Extraction is a difficult task due to the presence of a complex background that poses challenges such as sharply varying contours and background pixels that have the same intensities as text pixels. In this paper, we have proposed an algorithm that results in successful extraction of text from images with complex background and outperforms the existing methods in terms of efficiency.**

I. INTRODUCTION

The Text extraction from images with complex background remains rife with obstacles. With a rapid increase in the usage of images and videos for information retrieval, it is a pressing task to develop an efficient algorithm that successfully extracts text from images. Although many methods have been described for text extraction, very few methods exist that address the issue of presence of complex background. In this paper, we propose a novel text extraction method that initially detects the presence of text and extracts the detected text. The detection phase is composed of the following steps.

1.   A method to attain edge maps in all the four directions of text. An average edge map is then obtained.

2.   A K-means algorithm is applied to the average edge map.

3.   Morphological operations such as opening and dilation are applied.

4.   A strategy to eliminate false positives is devised.

The text extraction phase is composed of the following steps.

1.   An adaptive seed fill method that results in a binary image namely P.

2.   An adaptive threshold method that also results in a binary image namely Q.

3.   A method to fuse the two binary images P and Q to obtain a single binary image that represents the result of extraction.

The rest of the paper is organized as follows. Section 2 describes related work. Section 3 describes the text detection phase. Section 4 describes the text extraction phase. Section 5 discusses experimental results. The paper concludes with Section 6 that draws our inferences.

II. RELATED WORK

Previous Methods for text extraction from images can be divided into four categories: global thresholding method [1], local thresholding method [2, 3], model-based method [4] and seed-fill method [5].Global thresholding method takes account for only gray level information in the extraction of text. A pixel in the original image is classified as a text or background pixel simply according to its gray level in the global view. The binary image B=|F|T, where T is a constant. The Ostu method [1] is a well-known global thresholding method. The threshold selection of Ostu method is based on discriminate analysis. The global thresholding method can extract text from image with simple backgrounds. However, it cannot handle the images with complex backgrounds.

Local thresholding method [2, 3] estimates   a separate threshold for each pixel according to its local neighbors. The binary image B=|F|T, where Tisa threshold image T={t(x,y)}.The drawback of these methods is that when the backgrounds include sharply varying contours, the bright pixels around the contours distinguishing bright backgrounds from darker ones are classified as text pixels.

X.Yeetal. [4] proposed a stroke-model-based method that applies the double-edge model of text stroke as a unified solution to the problem of text extraction. This method may fail to extract some text pixels at corners and intersections of strokes. And, when backgrounds contain bright thin connected

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCAISE-2015 Conference Proceedings

components, this method will classify these thin connected components as text strokes.

Lyu, M.R et al proposed a seed fill method to remove background pixels that are classified as text pixels in the binary images. In this method, dam points are introduced to prevent the filling from flooding into text pixels. However, some background pixels inside the characters may be misclassified as text pixels.

As seen from the above analysis, it is difficult for the described methods to extract text from complex backgrounds, so we propose a novel method to address the issue.

III. TEXT DETECTION

The role of text detection is to find those regions in the image that contain only text. These regions are highlighted with bounding boxes and provided as input to the extraction algorithm. There are two conditions that should be met in order to achieve accurate text detection.

1. Attaining high detection rate, low false alarm rate, low misdetection rate and low inaccurate boundary rate for text of various font-size, font color, orientations, scripts and in complex background.
2. Consistent performance especially for low contrast images, besides high contrast images regardless of background complexity.

Text detection is a four step process [6] which has been briefly described below.

Fig 1(a) Input Image

A. Edge Maps method for Text Detection

The input image is resized to the dimension of 256*256 to facilitate uniformity. The resized image is converted to a grayscale image. We then implement an edge maps based method for detection. Edge is a distinctive characteristic which can be utilized to discover potential text areas. Text is mainly composed of strokes in horizontal, vertical, up-right and up-left directions. For the gray-scale image, a Sobel operator is employed to attain four directional edge maps which represent the edge density and edge strength in the above mentioned directions. Next, the average edge map is formulated by performing the average function on the previously obtained edge maps. Fig 1 (b) represents the average edge map for the input image.

Fig 1(b) Average Edge Map

B. K-means Algorithm

The second step in the Detection phase involves the application of the K-means algorithm on the average edge map procured in the first step. K-means algorithm is employed in order to classify the edge map into two clusters: background and text candidates.

C. Opening and Dilation

The third step in the Detection phase encompasses the execution of morphological operations such as opening and dilation. The reasoning behind this step is to procure connected components and to discard too small objects respectively. Fig 1 (c) corresponds to the result after the application of opening and dilation.

Fig 1 (c) Opening and Dilation

D. False Positive Filtering

Text Detection using the edge based methods often leads to many false positives. In order to eliminate the false detected text blocks [7] we compute a pre-defined features of the text like height, width, area (height * width), aspect ratio (width / height) and density which is given by the sum of all the pixels in the detected text block divided by the area (Edge area/ area).Based on the above features we eliminate the false positive text regions if it satisfies the below condition:

1) AR < T1 || density < T2 || height < 6 || width <= 5 || (height * width) < 24 || height > 70.

Thresholds T1 and T2 are determined empirically as: T1 = 0.5 and T2 = 0.1. Fig 1(d) represents the effect of false positive filtering.

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCAISE-2015 Conference Proceedings

Fig 1(d) False Positive Filtering

IV. TEXT EXTRACTION

The role of text extraction is to convert the grayscale image obtained as output from the Detection phase to binary images that are OCR-ready. Text Extraction is a three step process [8] and is elaborated as follows.

A. Locally adaptive seed fill method

There exists continuity between text and background in the image that is obtained as a result from Text Detection. Hence, filling from each pixel on the boundary of the image is an efficient way to find the background pixels. Once the background pixels are identified, they can be eliminated from the binary image. In order to prevent the seed filling from flooding into the text pixels and eliminate maximum amount of background pixels, an adaptive seed filling method is adopted.

Let $H = \{h(x, y)\}$ denotes a binary image. For a pixel (x, y), $h(x, y) = 1$ indicates the pixel is a background pixel, and $h(x, y) = 0$ indicates we are not sure whether it is a background pixel or not. Each pixel (x, y) in the binary image H is initialized to 0, i.e. $h(x, y) = 0$. Then, we take each pixel (x, y) on the boundary of the text image as a seed to find the background pixels and set the corresponding $h(x, y)$ to value 1. For each pixel of the original image $F = \{f(x, y)\}$ with gray level $f(x, y) \in [0, 255]$, the local threshold $g(x, y)$ and local variance $c(x, y)$ are calculated, respectively.

$$g(x, y) = (F_{max}(x, y) + F_{min}(x,y))/2$$
$$c(x, y) = F_{max}(x, y) - F_{min}(x,y)$$

In which $F_{max}(x, y)$ and $F_{min}(x, y)$ are the highest and lowest values in a local neighborhood centered at pixel (x, y). A non-background pixel q will be considered to be a background pixel, if it has the following properties:
1) q is one of the 4-neighborhoods of a background pixel p $= (x, y)$: $q \in \{(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)\}$;
2) $|f(q) - f(p)| < d_0 \cdot c(q)$ or $((f(q) > g(q)$ or $c(q) < d_2)$ and $|f(q) - f(p)| < d_1)$.
Here, $0 < d0 < 1$ is a constant; $0 < d1, d2 < 255$ are two constant thresholds. In our application, we heuristically chose the constant d0 as 0.05, and chose the thresholds d1 and d2 as 0.3 and 0.2 times the Ostu threshold for the local variance image $C = \{c(x, y)\}$. The process of seed-fill is repeated until there is no pixel has the above properties. The result of this step is a Binary image H that is shown in Fig 1(e).

Fig 1(e) Adaptive Seed Fill Method

B. Locally adaptive threshold method

Locally adaptive thresholding method calculates a threshold for each pixel in the image in accordance to its local neighbors. For each pixel of the input binary image, a local threshold and a local variance is determined with the formulae provided in the locally adaptive seed fill method. Each pixel in the input binary image is classified as a text pixel (indicated by value 1) or a background pixel (indicated by value 0) based on the conditions that follow:
1) $p(x, y) = 1$ if $f(x, y) > g(x, y)$ and $c(x, y) > c^*$
2) $p(x, y) = 0$ otherwise.
In the above condition, c^* refers to the Ostu threshold for the local variance image $C = \{c(x, y)\}$. The binary image $P = |C|_c \cap |F|_G$ where $G = \{g(x, y)\}$ is the local threshold image. The image thus obtained is shown in Fig 1(f).

Fig 1(f) Locally Adaptive Threshold Method

C. Fusing Binary Images

The two binary images that are obtained as a result of locally adaptive seed fill method and locally adaptive thresholding method are fused together to obtain the final binary image that represents the result of extraction. The underlying implementation behind the fusion involves the logical ANDing of the two binary images to attain the final result of extraction.


Fig. 1(g) Text Extraction Result.

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCAISE-2015 Conference Proceedings

V. CONCLUSION

In this paper, we have attempted to develop an efficient algorithm that could accomplish Text Extraction from Images with complex background. We have designed the following methods.

1. An efficient detection algorithm was developed to detect the presence of text in Images and localize it. The algorithm composed of Edge Maps method, K-means algorithm, and operations including Opening and Dilation and False Positive Filtering.

2. An efficient extraction algorithm was developed that composed of two methods, namely locally adaptive seed fill method and locally adaptive threshold method. The concluding step in the extraction method is to fuse the two binary images that are procured as outputs from the locally adaptive seed fill method and locally adaptive threshold method.

3. Satisfactory results were obtained from the algorithm that we have developed with minimum false positive rate, false boundary rate and efficiency of up to 92%

V.   EXPERIMENTAL RESULTS
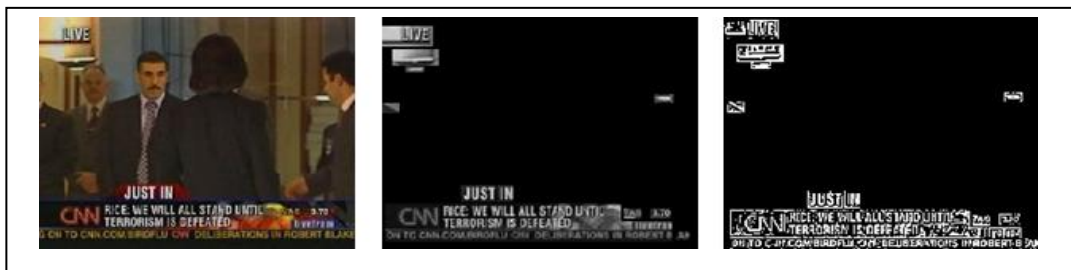
Input   Detection Result   Output

Input   Detection Result   Output

Input   Detection Result   Output

REFERENCES

[1] N. Ostu, "A threshold selection method from gray-scale histogram," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, 1978, pp. 62-66.

[2] J. Bernsen, "Dynamic thresholding of gray-level images," in *Proc. Int. Conf. Pattern Recognition*, vol. 2, 1986, pp. 1251-1255.

[3] X. Ye, M. Cheriet, C. Y. Suen, and K. Liu, "Extraction of bank check items by mathematical morphorlogy," *Int. J. Doc. Anal. Recognit.*, vol. 2, pp. 53-66, 1999.

[4] X. Ye, M. Cheriet, and C. Y. Suen, "Stroke-model-based character extraction from gray-level document images," *IEEE Trans. Image Processing*, vol. 10, Aug. 2001, pp. 1152-1161.

[5] Lyu M. R., Jiqiang Song, Min Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, 2005, pp243- 255

[6] "Accurate video text detection through classification of low and high contrast images" by Palaiahnakote Shivakumara , Weihua Huang, Trung Quy Phan, Chew Lim Tan, School of Computing, National University of Singapore, Singapore, 2010

[7] "A Laplacian Method for Video Text Detection" by Trung Quy Phan, Palaiahnakote Shivakumara and Chew Lim Tan *School of Computing, National University of Singapore*

[8] "Method Combination to Extract Text from Images and Videos with Complex Backgrounds" by Wuyi Yang, Shuwu Zhang, Zhi Zeng, Haibo Zheng at Digital Content Technology Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China, 2008