TEXT SEGMENTATION FROM IMAGES WITH TEXTURED AND COLORED BACKGROUND

Roshanak Farhoodi and Shohreh Kasaei
farhoodi@mehr.sharif.edu, skasaei@sharif.edu

Department of Computer Engineering
Sharif University of Technology, Tehran, Iran

ABSTRACT

In this paper, a new method to segment text regions from color images with textured background is proposed. The method is based on finding the text edges using information content of the subimage coefficients of the discrete wavelet transformed input images. Then, the detected edges are combined to form the exact location of the characters. In the final stage, the regions that are not acceptable as the text regions are removed (based on some general structural rules) to improve the overall performance. The experimental results show that the proposed method is robust against size, font, language, color and direction changes of the text regions.

Keywords: text extraction, text segmentation, wavelet transform, image documents, and OCR.

1. INTRODUCTION

The fast increase of digital media resources has emerged the content-based information retrieval and indexing approaches. Specifically, characters as the key semantic objects have a great role in understanding and farther processing of the automated information retrieval systems. In this regard, the *optical character recognition* (OCR) systems will result in a better performance when their inputs are some "clear" characters on a simple and smooth background.

Detection and extraction of texts from images involves some challenging difficulties. For instance, the characters are often mixed with other objects, the characters may be of any alphabets with any color, the background color may differ only slightly from that of the characters, the font style and size of the characters may vary, and finally the luminance of the images may also vary. The previous efforts in this field of research are mostly focused on solving the problem with restricted conditions imposed on the used characters or the background.

Some previous works on this field can be found in [1-12]. Yu Zhang *et al.* used the intensity variation information encoded in the *discrete cosine transform* (DCT) domain to present a fast text caption method in the JPEG compressed images and the I-frames of the MPEG compressed videos [7]. In another paper, Yu Zhang *et al.* used horizontal spatial variance and color information to locate text in the *compact disk* (CD) and book cover images, as well as in the images of traffic scenes captured by a video camera [6]. Kongqiao Wang and Jari A. Kangas have proposed a method for detecting characters in the scene images taken by a digital camera.[5]. This is one of researches on the color images that utilizes the color histogram for segmenting text regions. K. Jain and B. Yu. introduced an algorithm for text localization that is suitable for newspaper advertisements, web pages, images and videos [12]. That algorithm is base on a connected component analysis and requires either the text or its background to be monochrome. Some other researchers have employed machine-learning algorithms for text identification. Li uses a neural network classifier for text identification [8-9]. The adopted features are the second and the third order central moments in the wavelet domain. Leinhart *et al.* proposed a system for localizing and segmenting text in images and videos based on image gradient features and a complex-valued multilayer feed forward neural network classifier [10].

This paper presents an efficient algorithm to segment text from images with colored and textured background using wavelet transform. Wavelet transform have also played an important roles in classification of texture and abnormalities in medical images. An application of wavelet transform is the formation of classification features using the statistical characteristics of wavelet coefficients.

The proposed algorithm is described with details in Section 2. In the Section 3, the experimental results are illustrated to show the efficiency and robustness of the algorithm. Finally, Section 4 concludes this work.

2. PROPOSED ALGORITHM

The functional diagram of the proposed algorithm is presented in figure 1. As can be seen in the flow chart, a color image is entered to the system as the input data and the segmented text on a clear black background will be the output. In the following, we will describe each stage in detail.

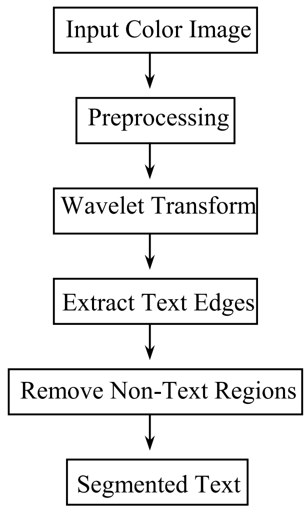


Figure 1. Block diagram of the proposed text segmentation algorithm.

2.1. Preprocessing

The preprocessing stage is performed on the input images commonly captured by digital cameras (as such images are usually contaminated with noise). Firstly, as the pictures are often in the *Red-Green-Blue* (RGB) color space we first convert the 3 color components (*i.e.,* R, G and B) to the intensity component by:

$$Y=0.299R+0.587G+0.114B. \tag{1}$$

This is the actually *Value* component of the *Hue-Saturation-Value* (HSV) color space. Then, the noise of the images is reduced using a weighted median filter that is applied on this component using the mask of:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

After this filtering step, a great part of noise will be removed while the edges in the image are still preserved.

2.2. Discrete Wavelet Transform

Transforms in the field of image processing are based on the idea that often a linear, invertible transform will result in transform coefficients that are more appropriate for the subsequent processes.
Most textured images are well characterized by their contained edges. The *discrete wavelet transform* (DWT) provides a powerful tool for modeling the characteristics of textured images [3]. The DWT is very useful for signal analysis and image processing, especially in multi-resolution representation [1]. It can decompose signal into different components in the frequency domain. Two-dimensional DWT decomposes image into four subimages: one averaged component (LL) and three detailed components (LH, HL, HH) as shown in figure 2. The Haar wavelet is one of the simplest wavelets. Some of the advantages of the Haar wavelet are that it is orthogonal and compactly supported. Due to its very short support, it is the only wavelet that allows perfect localization in the transform domain.

LL	HL
LH	HH

Figure 2. One level of 2-D DWT decomposition.

The wavelet transform can be interpreted as a multiscale edge detector that represents the singularity content of an image at multiple scales and different orientations. Wavelets overlying a singularity yield large wavelet coefficients; wavelets overlying a smooth region yield small coefficients.

In this step of our algorithm, the illumination components are transformed to the wavelet domain using Haar wavelet. This stage results in the four LL, HL, LH and HH subimage coefficients. Here, the coefficients of the horizontal, vertical, and diagonal subimages of the first level are used.

2.3. Extracting Text Edges

The "dense edges" are the distinct characteristics of the text blocks which can be used to detect possible text regions. Here, by finding the edges in the mentioned subimages and fusing the edges contained in each subimage, the candidate text regions can be found. As a result, we firstly need to employ an edge detector. Here, for computational efficiency, the *Sobel* edge detector is used. The Sobel edge detector is efficient to extract the strong edges that are needed in this application (unlike the Canny that extracts both strong and weak edges). We apply the Sobel edge detector on each subimage. An illustrative example is shown in figure 3.

As the next step, these candidate text edges are used to form an edge map. Here, we use a weighted "OR" operator. This operator decides whether a pixel belongs to an edge in the image or not regarding the coefficients in the vertical, horizontal and diagonal subimages. The edges associated with each subimage are weighted by [1, 1, 0.5], respectively.

Then we perform a morphological dilation operation on the processed edge map. The structuring element has the size of 1x6. This operation results in filling the gaps inside the obtained characters' regions.

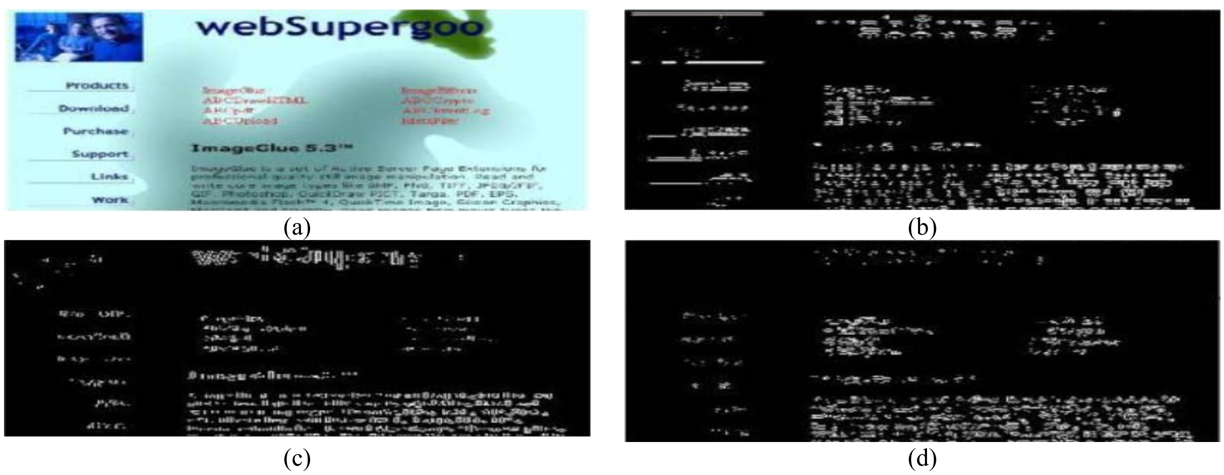


Figure 3. Original image and its DWT coefficients after detecting candidate text edges.
(a) Original image, (b) horizontal subimage, (c) vertical subimage, (d) diagonal subimage.

2.4. Removing Non-Text Regions

In this stage, to further enhance the results, we remove the non-text regions using some structural rules. To do so, we first summarize the common attributes of horizontal texts as:

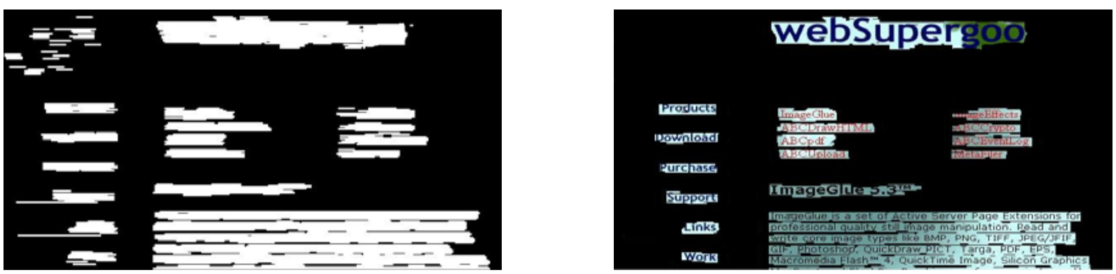
a) texts always contain edges,
b) texts are some bars whose widths are larger than their heights,
c) texts are bounded in size, and
d) texts have a special texture property.

Rules (a) and (b) have already been used to localize the text regions. Now we employ the other rules to remove the spurious detected regions.
So, as the final step, we remove the false locations due to the general rules of the text blocks defined as follows:

- The text width is greater than the text height about a factor of 2; this rule will result in removing most of the square pictures contained in the web pages that might be found incorrectly as the text regions in previous steps. This rule results in removing singular letters and numbers. In case that in a particular application these single characters are important, then this rule may be simply replaces by the following.
- Founded square text regions are accepted as single characters if their height (width) has a determined size. This reference size is obviously input-dependent.
- For horizontally aligned text regions, the text block height is between two threshold values. This rule is used to remove the vertical lines that might have been detected in the vertical subimage of the wavelet transform. The high threshold is about half of the input image height (though it can be altered for different types of input images), and the low threshold is experimentally set to 8 pixels.

3. EXPERIMENTAL RESULTS

Our database contains 100 color sample images of size between 290x56 and 560x600 pixels. The input images are chosen from a digital collection of web pages, scanned book and journal covers, scene images from digital cameras, advertisements, and video frames. The overall number of characters in the database images is difficult to determine; as there are many small indistinctive or single characters in some images. The results of the proposed algorithm when run on some typical images are shown in figures 4 and 5.



(a) (b)

Figure 4. (a) Dilated edge map, (b) extracted text regions.

(a) (b)

Figure 5. A sample image to illustrate the robustness of the proposed algorithm against font size.
(a) Original image, (b) extracted text regions.

The overall processing cost of the algorithm depends on the size of the image and the size of the obtained text regions according to the image size.

The performance was evaluated by comparing the manually extracted ground truth regions with the automatically extracted regions. The correct detection rate was defined as the ratio of the number of pixels in the real text regions to the number of pixels contained in the detected regions. It was **98.9%** for our database. Also, a comparison between the performance of our algorithm and some other algorithms reported in the literature is shown in Table1.

Table1. Average performance of different algorithms.

AVERAGE CORRECT RATE	
Traditional edge detection [1]	98.76
9-7 taps DWT [1]	98.78
Traditional Haar DWT [1]	98.57
Segment-matrix Haar DWT [1]	98.57
Our algorithm	**98.91**

4. CONCLUSIONS

A robust method for text segmentation is presented to detect and localize characters in color images with textured background. Most of the previous methods fail when the characters are too small or when the characters are not aligned well. They also result in some missing characters when characters have very poor contrast with their background area. The proposed method has overcome these shortcomings. It is also fast and robust against language, font size, and color of the texts.

ACKNOLEDGEMENT

This work was in part supported by a grant from ITRC.

5. REFERENCES

[1] Chung-Wei Liang and Po-yueh Chen, "DWT-Based Text Localization", International Journal of Applied Science and Engineering, 2004.2, 1:105-116.

[2] Jai Li And Robert M. Gray, "Text and Picture Segmentation by the Distribution Analysis of Wavelet Coefficients", International Conference on Image Processing, Chicago, Oct. 1998

[3] Hyeokho Choi and Richard G. Baraniuk, , "Multiscale Image Segmentation using Wavelet-Domain Hidden Markov Models", IEEE transaction on image processing, VOL. 10, No. 9, September 2001.

[4] Qixiang Ye, Wen Gao, Weiqiang Wang, Wei Zeng, "A Robust Text Detection Algorithm in Images and Video Frames", ICICS-PCM 2003, 15-18, December 2003, Singapore.

[5] Kongqiao Wang and Jari A. Kangas, "Character Location in Scene Images from Digital Camera", The journal of the pattern recognition society, Vol. 36, 2287-2299, March 2003.

[6] Yu Zhong, K.Karu and A.K. Jain, "Locating Text in Complex Color Images", IEEE Transactions on Pattern recognition, Vol.28, 1523-1536, October 1995.

[7] Yu Zhang, et al., "Automatical Caption Localization in Compressed Video", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22, 385-392, 2000.

[8] H. Li, D. Doermann, and O. Kia, "Automatic Text Detection and Tracking in Digital Video." Maryland Univ. LAMP Tech. Report 028, 1998.

[9] H. Li, O. Kia and D. Doermann, "Text Enhancement in Digital Videos," In Proc. SPIE99-Document Recognition and Retrieval, 1999.

[10] Rainer Leinhart and Axel Wernicke, "Localizing and Segmenting Text in Images and Videos", IEEE Transactions on circuits and systems for video technology, Vol.12, No. 4, April 2002.

[11] K. Jain and B. Yu, "Automatic text location in images and video frames," Pattern Recognition, Vol. 31, No. 12, 2055-2076, 1998.