Available online at www.sciencedirect.com

ScienceDirect

Karbala International Journal of Modern Science 2 (2016) 138–144
http://www.journals.elsevier.com/karbala-international-journal-of-modern-science/

Wavelet transform based technique for text image localization

Suhad A. Ali [a,*], Ashwaq T. Hashim [b]

[a] Babylon University, Science College for Women, Computer Science Dep., Babylon, Iraq
[b] Technology University, Control and Systems Engineering Dep., Baghdad, Iraq

Received 10 January 2016; revised 8 March 2016; accepted 21 March 2016
Available online 3 May 2016

Abstract

In this paper, a robust technique based on discrete wavelet transform, edge detection, and morphology operation for scene text detection is proposed. There are several stages in the proposed method. In the first stage, a single wavelet decomposition LH, HL and HH subbands are applied for detecting edges in original scene text image. The projection technique is applied in the second stage to preliminary detect text and non-text pixels. In third stage, 4-connected components are applied, and then area geometric feature is used as threshold to remove non-text region. At last stage, morphological operations are applied to connect isolated text components and to remove non-text regions. The proposed method is applied on a various images such as images of low contrast, complex background images and images of different fonts and size of text. The experimental results show that the proposed method can detect regions of the text perfectly.

Keywords: DWT; Image projection; Morphological operator; Connected components

1. Introduction

Natural scenes images contain different objects. A text object is important one among these objects, because it shows the important meanings (semantic) of image. A text images can be classified into three types namely ***document image***, ***scene text image***, and ***caption image***. Document image is acquired by scanning book covers, printed document etc. Scene text image sometime is referred to graphics text and it finds in natural images that contain advertisements such as street name, road signs etc. Caption image contains text which inserted in this image. Caption text is always referred as artificial text.

Many applications need text localization and segmentation from natural scene images [1]; ranging from automatic detection of traffic signs that help in transportation system [2], and helping visually impaired people [3], to multimedia indexing and retrieval [4]. Text localization and recognition problem has been recently receiving significant attention because text localization method achieved a localization recall of (62%) [4]. Text localization methods have been classified into two classes [5]:

1) **Region based methods**: These methods work based on color differences between text regions and their background. Region based methods are divided in

* Corresponding author.
E-mail addresses: suhad_ali2003@yahoo.com (S.A. Ali), ashwaqtalib@yahoo.com (A.T. Hashim).
Peer review under responsibility of University of Kerbala.

http://dx.doi.org/10.1016/j.kijoms.2016.03.004

two sub methods: connected component based method (CC), and edge detection based method [6]. In first method, a text is considered as a set of distinct connected components which have their specific intensity and color distributions. The second method depends on finding maximum intensity changes between text and background [7]. The main differences between two methods are edge based method is useful to process low contrast text image and with different text size, while connected components methods are simpler to implement. The main drawback of CC method is failed in locating text regions in images which have complex background.

2) **Texture based methods**: These methods work by extracting texture features of image firstly, and then a classification process is applied in the second stage to detecting text regions [8]. Discrete transformations such as discrete wavelet transform (DWT) and discrete cosine transform (DCT) are used. These methods suffer from high complexity in nature, but it is robust in processing complex background [9].

The rest of the paper is organized as follows; Section 2 gives briefly a review of the recent related works. Section 3 presents the proposed method. Section 4 illustrates the experimental results and discussions and Section 5 gives the conclusions and suggestion for future works.

2. Related works

This section surveys many methods which are related to our proposed method:

- In 2010, A. Angadi at al. proposed an algorithm for texture features based on discrete cosine transform (DCT). The method is applied on 100 natural scene images, it is inefficient when image background is more complex like trees, vehicles [10].
- In 2010 Epshtein et al. proposed a method to detect texts in many languages with different fonts based on stroke width transform (SWT) [11].
- In 2011 C. Yi and Y. Tian proposed a method for extracting text strings with arbitrary orientations. It is based on text image partitioning and connected components [12].
- In 2012, Seeri et al. proposed an algorithm for Kannada text images using combination of techniques such as median filter, sobel edge detector, connected component labeling, and order static filter. It fails to extract very small characters [13].
- In 2013, H. Koo and D. H, Kim proposed an algorithm for text region detection based on two classifiers, the first one is used for generating candidate word regions and the second for filtering out non-text regions [14].
- In 2014, Raj et al. proposed an algorithm for natural scene text image detection using connected components (CC). It fails for small slanted/curved text [15].

3. The proposed method

As presented in Fig. 1 the proposed text detection algorithm passes through many steps.

3.1. Preprocessing

In this step, if the input image is RGB, it must be converted into YUV color space by forming a weighted sum of the R, G, and B components as in Equation (1).

$$\begin{aligned} Y &= 0.299*R + 0.587*G + 0.114*B \\ U &= -0.14713*R - 0.28886*G + 0.436*B \\ V &= 0.615*R - 0.51499*G - 0.10001*B \end{aligned} \quad (1)$$

Our method uses only luminance components (Y) for next processing steps.

3.2. Discrete wavelet transform

Discrete Wavelet transform is referred as a multi-resolution decomposition approach. In the frequency domain; it decomposes a signal depending on a family of

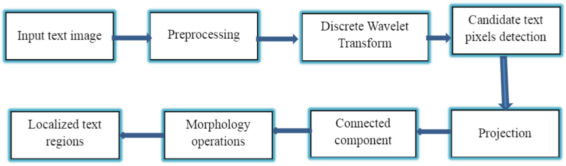


Fig. 1. The proposed method steps.

basis functions obtained through translation and dilation of a mother wavelet. When one-dimensional DWT is applied on an image, an image is decomposed into two parts which is coarse and detail elements by using low-pass (L) and high-pass (H) filters. Image signal is decomposed into four subbands when two dimensional DWT is applied on rows at first and then on columns of an image. These subbands are approximation sub-band (LL), and details subbands (LH, HL, and HH) [16]. For applying multilevel DWT; the approximation subband is used for next level of decomposition. Fig. 2 shows single level decomposition of DWT.

3.3. Candidate text pixels detection

As shown in Fig. 2-A, the text pixels have high variation around its neighbor pixels, therefor a technique based on edge detection will be applied. As shown in Fig. 2-B the edges found in high frequencies therefor is used in the next step for detection of text region. A wavelet edge pixel $E(i,j)$ at pixel in location i and j can be define by taking the average of corresponding pixels in three high frequency subbands (LH, HL and HH) according to Equation (2).

$$\mathrm{E}(\mathrm{i},\mathrm{j}) = \sum_{k=1}^{3}\left(\mathrm{D}_k(\mathrm{i},\mathrm{j})^2/3\right) \tag{2}$$

Then each pixel in $E(i,j)$ array will be a candidate text pixel if its value is larger than threshold (α) as follows:

$$C_p(i,j) = \begin{cases} 1 E(i,j) \geq \alpha \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

where $C_p(i,j)$ is an array of candidate text pixels, α is a threshold value, which its value is determined depending on the statistical measurements of $C_p(i,j)$ as follows:

$$\alpha = m + k \times \sigma \tag{4}$$

where m, and σ represent mean and standard deviation of candidate text pixels array respectively. The k is a parameter its value is selected depend on the local statistics such as mean and standard deviation of an image. Fig. 3 shows candidate text pixel detection.

The value of $\alpha = 43.89$ for image in (a) that is applied to get an image in (c), while the value of $\alpha = 22.83$ for image in (b) that is applied to get an image in (d).

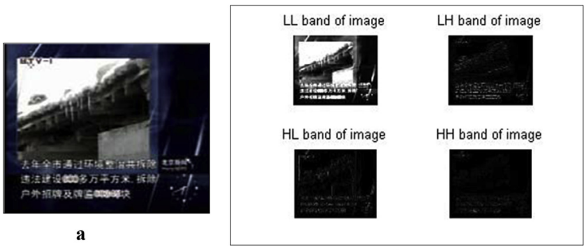
3.4. Projection

Vertical and horizontal projections are one of the amplitude segmentation methods. They convert image contents into one-dimensional representations. They computed parallel to the coordinate axis, so, they are useful methods. Vertical and horizontal projections of a binary candidate text pixels image $C_p(i,j)$, with $0 \leq i \leq M$, $0 \leq j \leq N$, are defined as [17]:

$$P_{\mathrm{hor}}(i_0) = \sum_{j=0}^{N} C_p(i_0,j) \quad \text{for } 0 \leq i_0 \leq M \tag{5}$$

$$P_{ver}(j_0) = \sum_{i=0}^{M} C_p(i,j_0) \quad for\ 0 \leq j_0 \leq N \tag{6}$$

where M and N represent height and width of image respectively. Equation (5) computes horizontal projection P_{hor} of row (i_0) as the sum of pixel values in that row and all the columns in the image. While vertical projection P_{ver} of column (j_0) is computed as the sum of pixel values in that column and all the rows in the image. To segment text regions, two thresholds are used horizontal and vertical thresholds, which denoted as T_h and T_v respectively. They are defined as:



a

b

Fig. 2. Single level decomposition of DWT (a) original text image (b) wavelet decomposition subbands.

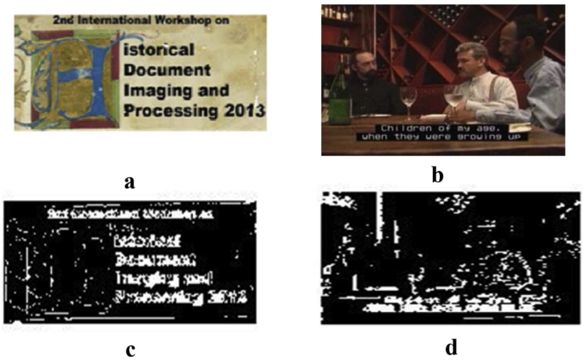


Fig. 3. Candidate text region detection (a–b) original text images (c–d) images with candidate text pixels images.

$$T_h = \frac{\text{mean}(P_{\text{hor}}) + \min(P_{\text{hor}})}{2} \quad (7)$$

$$T_v = \frac{mean(P_{ver}) + min(P_{ver})}{2} \quad (8)$$

If P_{hor} (i_0) is greater than horizontal threshold T_h then row (i_0) can be considered as a part of a candidate text region; otherwise this row is suppress. If P_{ver} (j_0) is greater than vertical threshold T_v then column (j_0) can be considered as a part of a candidate text region. Fig. 4 shows projection of candidate text image that is obtained from Fig. 3-D.

3.5. Connected components

In this step, horizontal and vertical images are combined to be one image, and then non-text regions which are false detected are eliminated using geometrical features of regions. To get image regions a 4-connected component is applied on the combined projection image. Area geometric feature of each region is computed and compared with threshold (*thr1*) which computed according to Equation (9).

$$thr1 = \frac{\text{Maxarea}}{10} \quad (9)$$

where Maxarea represents a largest region area.

If the area of each region is less than threshold, then it is considered as non-text region and its pixels should be discarded. The second condition that applies to discard non-text region is to compare the ratio (*R*) between height and width of each region with second threshold (*thr2*). The value of *thr2* is selected empirically and it is set to (12) in all experiments. If (*R*) of region larger than (*thr2*) then discarded region elements. Fig. 5 shows the result of applying this step.

3.6. Morphology operations

As shown in Fig. 5-B, the resultant image may contain non-text pixels. So a morphology operation; dilation using structure element with size (9 × 10) followed by opening operation using structure element with size (11 × 35) are applied to further remove non-text pixels. Fig. 6 shows the detection of text regions after applying morphology operations on Fig. 5-B.

4. Experimental results

Several experiment tests were carried out on text images. Every color image must be converted into gray scale image before applying text localization method. Fig. 7 shows several performed experiments.

5. Conclusions and suggestion for future works

Nowadays, the detecting and localization of text regions in scene images is open problem. In this paper a new method based on combinations of edge detection method and connected components (CC) method is proposed for text detection and localization in image. The experimental tests show the following conclusions:

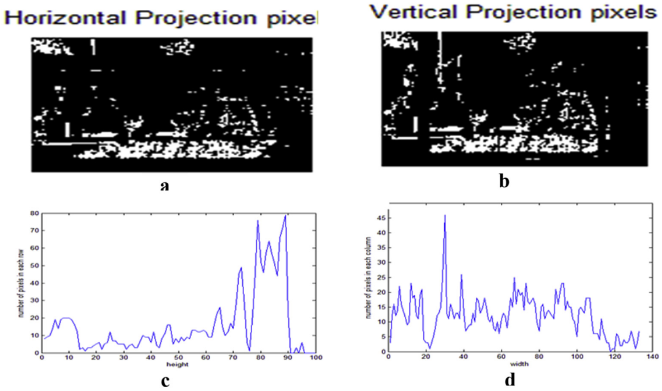


Fig. 4. Projection of candidate text image from Fig. 3-D (a) horizontal projection image (b) vertical projection image (c−d) graphical representation of horizontal and vertical projections respectively.

Fig. 5. Connected components step (a) combined projection image (b) result image after discard non-text regions.

Fig. 6. Text region detection (a) original image (b) detected text regions image.

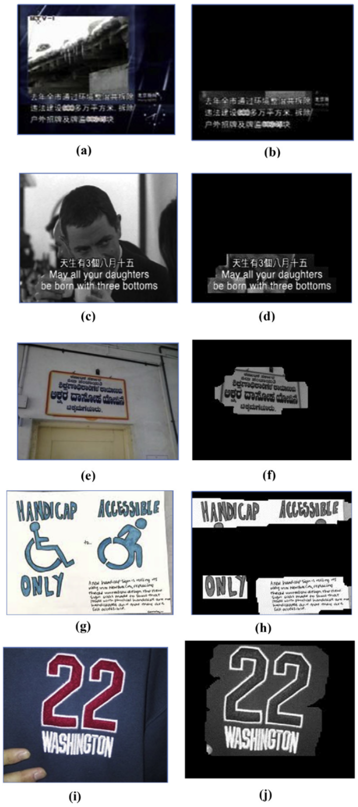


Fig. 7. Original text images and its detected text regions images.

1. The combinations of edge detection method and connected components (CC) method make the proposed method more robust one.
2. Using wavelet transform gives good tools for detecting candidate text regions.
3. The proposed method is more robust in dealing with the low contrast image, images that have different sizes and fonts, and text images with different languages as shown from Fig. 7.

The suggestion for future works can be summarized as follows:

1 In this paper, we made a text detection and localization method only. In order to get a good recognition result for the text characters, a text should be clearly extracted from its background.
2 Efficient method should be investigated to segment the characters from their background before putting them into OCR software.
3 A text extraction and recognition should be integrated with text localization to complete the need of text information extraction.

References

[1] Shivananda V. Seeri, J.D. Pujari, P.S. Hiremath, Multilingual text localization in natural scene images using wavelet based edge features and fuzzy classification, Int. J. Emerg. Trends Technol. Comput. Sci. 4 (1) (January-February 2015) 210–218.
[2] T. Kumuda, L. Basavaraj, Novel technique for text detection and localization in natural scene images, Int. J. Eng. Res. Technol. (IJERT) 3 (4) (April – 2014) 2675–2680.
[3] Xiaoqian Liu, Ke lu, Weiqiang Wang, Effectively localize text in natural scene images, in: 21st International Conference on Pattern Recognition, November 11–15, 2012, pp. 1197–1200. Tsukuba, Japan.
[4] L. Neumann, J. Matas, Scene text localization and recognition with oriented stroke detection, in: IEEE International Conference on Computer Vision (ICCV), 2013, pp. 97–104.
[5] Keechul Jung, Kwang In Kim, Anil K. Jain, Text information extraction in images and video: a survey, Pattern Recognit. 37 (5) (2004) 977–997.
[6] Pooja B. Chavre, Archana Ghotkar, A survey on text localization method in natural scene image, Int. J. Comput. Appl. 112 (13) (February 2015) 0975–8887.
[7] Divya gera, Neelu Jain, Comparison of text extraction techniques- a review, Int. J. Innovative Res. Comput. Commun. Eng. 3 (2) (February 2015) 621–626.
[8] Nikita Aher, Text localization and extraction in natural scene images, Int. J. Recent Innov. Trends Comput. Commun. 3 (2) (2015) 512–516.
[9] Mohieddin Moradi, Saeed Mozaffari, Ali Asghar Orouji, Farsi/Arabic text localization from video images by corner detection, in: 6th IEEE, Iranian Conference on Machine Vision and Image Processing, 2010, p. 674. Isfahan, Iran.
[10] S.A. Angadi, M.M. Kodabagi, Text region extraction from low resolution natural scene images using texture features, in: IEEE 2nd International Advance Computing Conference, 2010, pp. 121–128.
[11] B. Epshtein, E. Ofek, Y. Wexler, Detecting text in natural scenes with stroke width transform, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2963–2970.
[12] C. Yi, Y. Tian, Text string detection from natural scenes by structure-based partition and grouping, IEEE Trans. Image Process 20 (9) (2011) 2594–2605.
[13] S.V. Seeri, S. Giraddi, B.M. Prashant, A novel approach for Kannada text extraction, in: Proceedings of the International Conference IEEE on Pattern Recognition, Informatics and Medical Engineering, 2012, pp. 444–448.

[14] H. Koo, D.H. Kim, Scene text detection via connected component clustering and non text filtering, IEEE Trans. Image Process 22 (6) (2013) 2296−2305.
[15] H. Raj, R. Ghosh, Devanagari text extraction from natural scene images, in: International Conference on Advances in Computing. Communications and Informatics (ICACCI), IEEE, 2014, pp. 513−517.
[16] Tejashvini R. Hiremath, Shivanand M. Patil, V.S. Malemath, Detection and extraction of text in images using DWT, Int. J. Adv. Res. Comput. Commun. Eng. 4 (6) (2015) 533−537.
[17] Wilhelm Burger, Mark J. Burge, Principles of Digital Image Processing: Core Algorithms, Springer Publishing Company, 2009.