A Text Localization Algorithm in Color Image via New Projection Profile

G. Aghajari, J. Shanbehzadeh, and A. Sarrafzadeh

***Abstract*—Text data in image present useful information for automatic annotation, indexing and structuring of images. In this paper, we propose an approach to automatically localize horizontally texts appearing in color and complex images. First, an edge detection method using a wavelet transform is used to finding text in image. Second, the image is binarized. Third, a new filter is applying for removing disperses pixels and non text area. After that, a new projection profile is applying for estimating text regions. The experimental results show that the proposed method achieves a much higher accuracy than existing methods. The advantage of this algorithm is low computation for finding text.**

***Index Terms*— Projection Profile, Text Localization, Wavelet Transform, Machine Vision, Image Processing.**

I. Introduction

Content-based image indexing refers to the process of attaching labels to images based on their content. Image content can be divided into two main categories: perceptual content and semantic content. Perceptual content includes attributes such as color, intensity, shape, texture, and their temporal changes, whereas semantic content consists of objects, events, and their relations.

Among semantic image content, text within an image is of particular interest as (i) it is very useful to describe the contents of an image; (ii) it can be easily extracted comparing with the other semantic contents, and (iii) it enables applications such as keyword-based image search, automatic video logging, and text-based image indexing [1].

This paper presents a new approach that localizes text in complex images automatically. First, a wavelet-based edge extraction scheme is applied on gray-level image. Second the gray level edge image is converted into binary image by using suitable global thresholding. After that, a new filter is applied on binary image to remove noise and non text areas. The text locations are determined using new projection profile. Several heuristic methods are applied to improve the system performance. Bounding boxes are generated two last steps. Figure 1 shows the structure of the system.

Manuscript received August 25, 2009.This work is part of MSc thesis in Islamic Azad University Science and Research Branch which is supported by the educational center of the university.

G. Aghajari is a postgraduate student in the field of artificial intelligence in Islamic Azad University Research and Science Branch, Tehran, I. R. Iran. He is interested in machine vision and image processing and its application on emotion detection for e-learning. (e-mail: g.aghajari@gmail.com).

J. Shanbehzadeh is an Associate Professor with the Department of Computer Engineering, Tarbiat Moallem University (TMU). (Mobile: +989121484177; e-mail: jamshid@tmu.ac.ir).

A. Sarrafzadeh is an Associate Professor and Head of Department of Computing at Unitec, New Zealand. (Corresponding author. e-mail: hsarrafzadeh @unitec.ac.nz).

The remainder of the paper is organized as follows. Section 2 provides a brief overview of related work in the field. Section 3 introduces new projection profile. Section 4 presents the individual steps of our approach to text localization in detail. Section 5 describes the comparative experimental results obtained for a set of images. Section 6 concludes the paper and outlines areas for future research.

II. Related Work

Text localization methods can be categorized into three types: region-based, texture-based and hybrid approaches. Region-based schemes use the properties of color or gray-scale in a text region or their differences with corresponding properties of background. These methods can be divided further into two sub-approaches: connected component (CC) and edge-based.

CC-based methods apply a bottom-up approach by grouping small components into successively larger ones until all regions are identified in the image. A geometrical analysis is required to merge the text components using the spatial arrangement of the components so as to filter out non text components and mark the boundaries of text regions.

Among the several textual properties in an image, edge-based methods focus on the 'high contrast between the text and the background'. The edges of the text boundary are identified and merged, and then several heuristics are performed to filter out the non text regions.

Texture-based methods use the observation that text in images has distinct textural properties that distinguish them from the background. The techniques based on Gabor filters, Wavelet, FFT, spatial variance, etc. can be performed to detect the textural properties of a text region in an image [1].

First class of localization methods can be found in some works [2] – [5]. Gllavata *et al.* [2] have presented a method to localize and extract text automatically from color images. First, they transform the color image into greyscale image and then only the Y component is used. The text candidates are found by analyzing the projection profile of the edge image. Finally, a binarized text image is generated using a simple binarization algorithm based on a global threshold. They [3] also have applied the same idea of the previously mentioned paper to localize text; in addition the algorithm has been extended with a local thresholding technique.

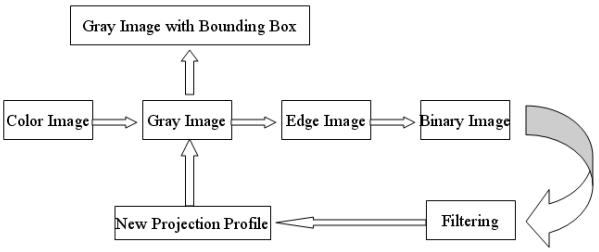

Fig. 1 Flow diagram of the system

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Cai *et al*. [4] have presented a text detection approach which uses character features like edge strength, edge density and horizontal alignment. First, they apply a color edge detection algorithm in YUV color space and filter out non text edges using a low threshold. Then, a local thresholding technique is employed to keep low-contrast text and further simplify the background. An image enhancement process using two different convolution kernels follows. Finally, projection profiles are analyzed to localize the text regions.

Jain and Yu [5] first employ color reduction by bit dropping and color clustering quantization, and afterwards a multi-value image decomposition algorithm is applied to decompose the input image into multiple foreground and background images. Then, CC analysis is performed on each of them to localize text candidates.

Second class of localization methods can be found in other papers [6] – [12]. For example, Q. Ye *et al*. [6] have proposed an algorithm based on wavelet to localize text from images and videos. By applying wavelet, they propose a coarse-to-fine algorithm that is able to locate text lines even under complex background. First, in the coarse detection, after the wavelet energy feature is calculated to locate all possible text pixels, a density-growing method is developed to connect these pixels into regions which are further separated into candidate text lines by structural information. Secondly, in the fine detection based on four kinds of texture features, a forward search algorithm is applied to select the most effective features. Finally, a SVM classifier is employed to identify the true text from the candidates based on the selected features.

Gllavata *et al*. [7] have proposed an unsupervised learning method for text detection and localization in images with complex backgrounds. The standard deviation of histograms of high-frequency wavelet coefficients was used as the main feature for the subsequent classification process. The classification was performed by a slightly modified k-means algorithm. The text candidates undergo a projection profile analysis in order to refine localization.

Xi. Li *et al*. [8] have presented an algorithm for text detection. This algorithm uses the stroke filter to calculate the stroke maps in horizontal, vertical, left-diagonal, right-diagonal directions. Then a 24-dimensional feature is extracted for each sliding window and a SVM classifier obtains rough text regions. The rough text regions are further refined through a group of rules. The candidate text lines were localized more accurately by projection profile of the refined text regions. Finally another SVM classifier based on a 6-dimensional feature is used to verify the candidate text lines.

Some authors use hybrid algorithms for text localization. For example, Gllavata *et al*. [13] have proposed a hybrid approach to automatically localize, segment and binarize text which is superimposed over complex images. Texture-based and CC-based methods are combined in their paper. First, an unsupervised texture-based method is used to detect text regions. Texture features are extracted from the high-frequency wavelet coefficients. Then, the text candidates are localized via CC-based filtering using predefined geometric text properties, and some false alarms are discarded. An unsupervised learning method is applied to achieve accurate character segmentation and binarization. Here, the text and background color are determined using a color vector quantizer and line histograms. The estimated text color and the standard deviation of the wavelet transformed image are used to classify the pixels into text and non text pixels.

III. New Projection Profile

In this section, the new projection profile is introduced. This method is being performed on binary images. It starts scanning from left side of the every line and records a change in case of facing the pixel change from zero to one and again returns to zero. Counting the changes of every line doesn't depend on number of pixels in this method. Figure 2 and 3 show the new projection profile of the images of letters 'T' and 'H' respectively. Although the higher part of letter T has more pixels in comparison with other lines, the new projection profile allocates 1 to all lines in figure 2. Whereas applying the old projection profile leads to 5 for the line of higher part and 1 for the others. Figure 3 is the same too. Studying English letters illustrates that new projection profile considers 1, 2, 3 or 4 for a letter with small font's size in the absence of cuts and noise. Therefore, an advantage of this method is the low deviation in the concerned number for the rows of text. Robustness in noisy condition is another advantage of this method. It is shown in figure 4.

In the image of letters with large fonts, new projection profile allocates greater numbers in neighbor rows, because the edges are not connected to each other in adjacent rows. However the difference between the allocated numbers to the adjacent lines would be small.

It is concluded from the studies that by new projection profile, we can estimate the possibility of the presence of the text in every line especially for texts with normal font size via thresholding.

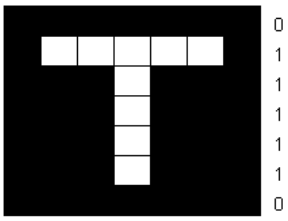

Fig. 2 Result of New projection profile on letter 'T'.

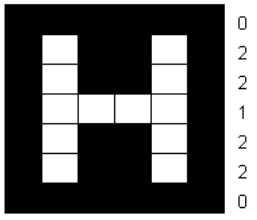

Fig. 3 Result of New projection profile on letter 'H'.

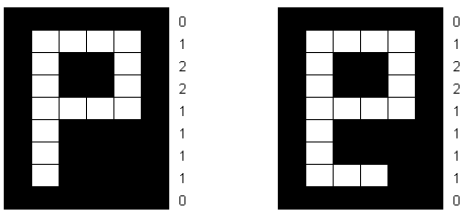

Fig. 4 Result of New Projection Profile for Letter P in noisy and without noise condition.

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

IV. Proposed Text Localization

The proposed approach for text detection in color image can divide into following steps:

Step 1: Image Preprocessing. The input Image in this step is the color Image in RGB color space. It is converted to three channels R, G and B. Then, channel G is selected for next processing. Figure 5 shows a channel G of a picture.

Step 2: Edge Detection. This step focuses on areas where text may occur. This method is similar to the method in [15]. We apply a wavelet-based method to obtain the edges of the gray-level image. First we apply the two-dimensional db4 wavelet transform to G channel from previous step.

Information about the edges of an image is contained in V, H and D. Second, we convert all the values in block B to zero and then compare the absolute values in V, H and D with a threshold and consider zero for the values less than this threshold. After that, we apply the inverse of db4 wavelet transform on modified image to obtain a matrix for comprising only the vertical, horizontal, and diagonal differences between main input image of this step and block B. The main reason of applying wavelet transform for edge detection is that wavelet transform can remove the noise whereas Sobel detector identifies noisy pixels as edge pixels. The other problems of other edge detectors are also considered. In figure 6, you can view the edge image of the image in figure 5.

Step 3: Binarizing Image. In this step, the gray level image is converted into binary image with suitable threshold. Subsequence steps for finding location of text use this binary image. Experimentally, we found that the suitable threshold is get from division of maximum values of pixels in gray edged image by four.

Step 4: Applying Filter on image. Because of global thresholding in previous step and extra edges, it is possible disperse pixels exist in the picture which make difficulties future processes. Because of this, we use a mask for omitting these pixels. We have two views in designing this mask.

First, low number pixels in a big area are considered as non text area or noise, second the possibility of text existence in area with disperse pixels is low. Input of this filter is a binary image. In general, this filter is a matrix 3*3 which each element of this matrix is a square matrix. The size of these blocks is equal. This filter passes from the picture and calculates one pixels under these square blocks. After calculating the one pixels, we compare the totals with two thresholds. If the amounts of one pixels under the blocks
aren't between a low and a high threshold, all the pixels under the middle blocks are considered as non text area and all element of this block get zero. You can view the general

Fig. 5 Channel G selected from a picture.

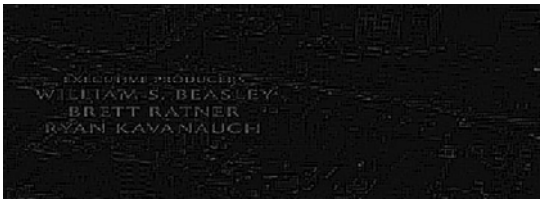

Fig. 6 Edge detection with wavelet transform

design of these filters in figure 7 and You can view the pseudo code of this filter in figure 8.

Step 5: Detection of Text Regions. For detecting the text area, first we apply new projection profile on the image and calculate the number of changes for each row. If the allocated amount of changes for each row is between two thresholds (low and high thresholds), the row potentially would be considered as a text area and the up and down of this row would be specified. Next, we search vertically for finding the exact location of the text and ignoring these rows as a text. For finding the exact location of texts or ignoring it, we use some heuristic. These heuristics include height and length of text and the ratio of height to length and enough number of pixels in this horizontal area. In figure 9, you can see the resulting bounding box and text area of the algorithm.

HighLeft	HighMid	HighRight
MidLeft	Mid	MidRight
LowLeft	LowMid	LowRight

Fig. 7 General design of the filter.

```
BinaryImage = Function RemoveNoise (Binary Image)
// SquareLength is the Length of Square that Centered in the mask
// CThresh is the threshold for the sum of pixels in the center square
// SThresh is the threshold for the sum of pixels in each surrounding squares of the center
// square
for I=(SquareLength+1): Length: Row
      for J=(SquareLength+1) :Length: Column
        HighLeft = # of Pixels in the High and Left Square of Center Square
        MidLeft = # of Pixels in the Left Square of Center Square
        LowLeft = # of Pixels in the Low and Left Square of Center Square
        HighMid = # of Pixels in the High Square of Center Square
        Mid = # of Pixels in the Center Square
        LowMid = # of Pixels in the Low Square of Center Square
        HighRight = # of Pixels in the High and Right Square of Center Square
        MidRight = # of Pixels in the Right Square of Center Square
        LowRight = # of Pixels in the Low and Right Square of Center Square
        if ( (HighLeft<=CThresh) & (MidLeft<=SThresh) & (LowLeft<=SThresh) &
           (HighMid<=SThresh) &  (Mid<=CThresh) & (LowMid<=SThresh) &
           (HighRight<=SThresh) & (MidRight<=SThresh) & (LowRight <=SThresh) )
           OutImage (I:I+((2*SquareLength)-1),J:J+((2*SquareLength)-1))=0;
      End // of if
  End // of for J
End  // of for I
End // of Function
```

Fig. 8 Pseudo code for the filter in step 4.

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Fig. 9 Result of bounding box by our method.

V. Experimental Result

In this section, we will show experiment results produced by the algorithm. The proposed approach has been evaluated a data set containing different type of images. The performance of the system is evaluated on the text box level in term of precision and recall.

The results for the experiments are summarized in table 1 where the number of existing text lines, the number of detected text lines, the number of false alarms and the corresponding values for recall and precision are listed. Recall is defined as:

Table I Experimental result for the proposed algorithm.

#images	200
# textlines	480
#correct detected	437
#false positives	20
Recall (%)	91.77%
Precision (%)	96%

$$\text{Recall} = \frac{\text{Correct Detected}}{(\text{Correct Detected} + \text{Missed Text Lines})}$$

Whereas precision is defined as:

$$\text{Precision} = \frac{\text{Correct Detected}}{(\text{Correct Detected} + \text{False Positives})}$$

A text box is considered as detected correctly, if a text is completely surrounded by a box, while a detected text box is considered as a false alarm, if no text appears in that box.

The text localization algorithm achieved a recall of 91.77% and a precision of 96%.

VI. Conclusion

In this paper, we have proposed an approach to automatically localize text appearing in images with complex backgrounds.

There are several areas for further research. We see that the edge based method have good performance for localization of text in image with complex background. This is notable that we can combine this method with texture method for better performance. Further more, to improve the performance, it is possible to combine this algorithm with mathematical function.

References

[1] K. Jung, K.I. Kim, and A.K. Jain, “Text Information Extraction in Images and Video: A Survey,” Pattern Recognition, vol. 37, 2004, pp. 977-997.

[2] J. Gllavata, R. Ewerth, and B. Freisleben, “A Robust Algorithm for Text Detection in Images”, Proc. of 3rd Int'l Symposium on Image and Signal Processing and Analysis, Rome, 2003, pp. 611-616.

[3] J. Gllavata, R. Ewerth and B. Freisleben, “Finding Text in Images via Local Thresholding”, International Symposium on Signal Processing and Information Technology, Darmstadt, Germany, 2003, pp. 539-542.

[4] M. Cai, J. Song, and M. R. Lyu, “A New Approach for Video Text Detection”, Proc. of IEEE Int’l Conference on Image Processing, Rochester, New York, USA , 2002, pp. 117-120.

[5] A. K. Jain, and B. Yu, “Automatic Text Location in Images and Video Frames”, Pattern Recognition 31(12), 1998, pp. 2055-2076.

[6] Q. Ye, Q. Huang, W. Gao and D. Zhao, “Fast and robust text detection in images and video frames”, Image Vision Comput. 23 (2005), pp. 565–576.

[7] J. Gllavata, R. Ewerth, and B. Freisleben, “Text Detection in Images Based on Unsupervised Classification of High-Frequency Wavelet Coefficients”, Proc. of Int’l Conf. on Pattern Recognition, Cambridge, UK, 2004, pp. 425-428.

[8] Xiaojun Li, Weiqiang Wang, Shuqiang Jiang, Qingming Huang, Wen Gao, “Fast and Effective Text Detection”, IEEE International Conference on Image Processing, Oct 12-15, 2008, San Diego, USA, 969-972.

[9] Y. Hao, Z. Yi, H. Zengguang, and T. Min, “Automatic Text Detection In Video Frames Based on Bootstrap Artificial Neural Network and CED”, Journal of WSCG Vol. 11, No.1, Plzen, Czech Republic, 2003, ISSN 1213-6972.

[10] H. Li, D. Doermann, and O. Kia, “Automatic Text Detection and Tracking in Digital Videos”, IEEE Transact. on Image Processing, Vol. 9, Nr. 1, 2000, pp. 147-156.

[11] R. Lienhart, and A. Wernicke, “Localizing and Segmenting Text in Images and Videos”, IEEE Transact. on Circuits and Systems for Video Technology, Vol. 12, Nr. 4, 2002, pp. 256-268.

[12] V. Wu, R. Manmatha, and E.M. Riseman, “Textfinder: An Automatic System to Detect and Recognize Text in Images”, IEEE Transact. on Pattern Analysis and Machine Intelligence, Vol. 21, Issue 11, 1999, pp. 1224-1229.

[13] J.Gllavata, R. Ewerth and B. Freisleben. “A Text Detection, Localization and Segmentation System for OCR in Images”, in Proc. of Int'l Symposium on Multimedia Software Engineering (MSE'04), pp. 310-317, IEEE Press, Miami USA, 2004.

[14] Patrick J. Van Felet, 2008," Discrete Wavelet Transform: An Elementary Approach with Application", Wiley.

[15] J. Gllavata, R. Ewerth, T. Stefi, and B. Freisleben, “Unsupervised Text Segmentation Using Color and Wavelet Features”, Proc. of 3rd Int’l Conf. on Image and Video Retrieval, Dublin, Ireland, 2004, pp. 216-224.

[16] Maria Petrou, Pedro Garcia Sevilla, 2006, "Image Processing Dealing with Texture", Wiley.

[17] Rafael C. Gonzalez and Richard E. Woods, 2002," Digital Image Processing", Second Edition, Prentice Hall.

[18] K.P. Soman, K.I. Ramachandran, 2006, " Insight into Wavelets: From Theory to Practice", Second Edition, Prentice-Hall.

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)