Text Locating from Natural Scene Images Using Image Intensities

JiSoo Kim, SangCheol Park, and SooHyung Kim
Computer Science Dept., Chonnam National University
{*kimjisoo, scpark, shkim*}*@iip.chonnam.ac.kr*

Abstract

In this paper, we propose three text extraction methods based on intensity information for natural scene images. The first method is composed of gray value stretching and binarization by an average intensity of the image. This method is appropriate to extract texts from complex backgrounds. The second method is a Split and Merge approach which is one of well-known algorithms for image segmentation. The third one is a combination of the two. Experimental results show that the proposed approaches are superior to conventional methods both in simple and complex images.

1. Introduction

In recent years, we have created many still-images and videos using digital cameras, digital camcorders and cellular-phone cameras. Texts in these images contain very important information about locations and road signs. If we could recognize these texts accurately in real time, we can design artificial vision systems for assisting auto-navigation of vehicles and vision-impaired, video indexing and retrieval systems, text translation systems, and spam-mail filtering systems, and so on.

Previous works on text extraction in natural scene can be briefly classified into gray level information-based and color information-based methods[1-9]. In the following a brief overview is given for gray level information methods. [2] used a spatial variance on the converted gray-level image(still images, web images). [3] presented a method with several restrictions on the characters. After a local binarization, a relaxation algorithm is used for the component merging. They tested the proposed method on several images taken from license number plates, sign boards of shops, and road signs which include characters of various sizes, and the extraction results are dependent on the character slant and tilt. [4] proposed an algorithm for the automatic extraction of characters from digital images. Under the hypothesis of the character monochromaticity, a segmentation of the image is performed by means of a split and merge algorithm.

In the following, a brief overview is given for color information-based methods. [1] presented two methods for the localization of text in complex color images. The first one is based on a color segmentation step performed by searching for prototype colors as local maxima on the color histogram. The second considers the local spatial variance on a gray-level image computed over horizontal lines. [7] proposed two methods using a differential top-hats morphological operator and a direction filter. These methods show robustness to the light changes in images. [8] presented a method for the automatic extraction of text lines from color video frames, and the characters are assumed to lying horizontally with similar colors, and sizes.

We propose three text extraction methods based on intensity information from natural scene images. The first method is composed of gray value stretching and binarization by an average intensity of the image, and this method is appropriate to extract texts from complex backgrounds. The second method is a Split and Merge approach which is one of well-known algorithms for image segmentation. The third one is a combination of the two methods.

2. Proposed System

2.1 Gray-level Information Analysis(GIA)

As illustrated in Figure 1, the GIA algorithm consists of two steps: preprocessing and text region extraction.

2.1.1. Preprocessing

The preprocessing consists of gray image transformation, contrast stretching, median filtering, high pass filtering, and Laplacian operator. First, the input color image is transformed to a gray image[10]. Second, a contrast stretching and a median filter are applied to the gray image. Third, a high pass filtering and an opening operation are applied to that image. Last, an edge image is extracted by Laplacian operator.

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

The GIA algorithm is very robust to extract text whose color is quite similar to the background.

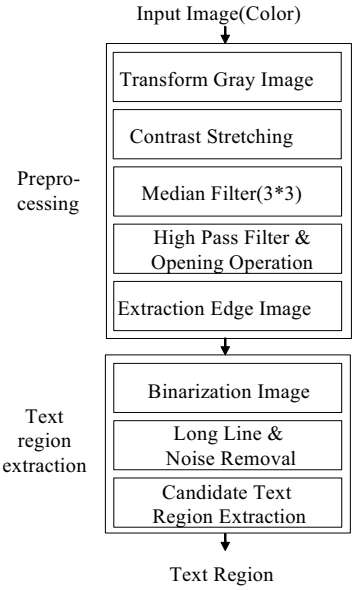


Figure 1. Algorithm of GIA method.

2.1.2. Text Region Extraction

The text region extraction consists of binarization, long line and noise removal, and candidate text region extraction. In the first step, the gray image is transformed to binarization image with Equation (1) which defines the image binarization function.

$$Mean = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}F(i,j) \qquad (1)$$

(M = height of image, N = width of image)
We binarize each pixel as follows:

$G_{ij} \geq Mean$, then 0 (edge pixel), G : Gray Image

$G_{ij} < Mean$, then 255 (background pixel)

The result of this step is a binary image containing edges.

In the second step, connected components and their bounding boxes are extracted, and their size, locations, and aspect ratio are determined. Some of the candidate text regions are removed. A component is removed when its size is too big or too small, or when the width or height of its bounding box is too large or too small. Then, the neighboring boxes are merged to compose characters or text lines. Finally, the bounding boxes are grown horizontally 2 pixels to the left and 2 pixels to the right in order to compensate for the preprocessing step. Figure 2 shows the results of GIA after preprocessing and text region extraction. Figures 2a shows the input image. Figure 2b shows the image where long line and small noise are removed. Figure 2c shows the input image with superimposed bounding boxes of the connected components.

(a) (b) (c)

Figure 2. The input image (a), the long line and noise removal image (b), the final result (c).

2.2 Split and Merge Analysis(SMA)

As illustrated in Figure 3, the SMA algorithm consists of a Split/Merge method.

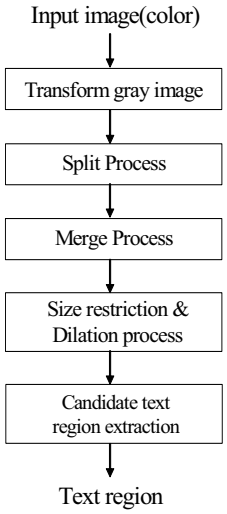


Figure 3. Algorithm of SMA method.

2.2.1. Split and Merge Process

In this section, we present a Split/Merge method to accurately extract text regions in natural scene images.

(1) Split Process

The input color image is transformed to a gray image[10]. The split algorithm consists of the following steps:

Step 1: Start with the entire image as a single region

Step 2: Pick a region R. If $H(R)$ is true, split the region into four subregions

Step 3: Repeat these steps until no further splits take place

$H(R)$: $R_{\max} - R_{\min} > \mathrm{T}$, T : Threshold

$R_{\max}$: a maximum gray value in a region.

$R_{\min}$: a minimum gray value in a region.

(2) Merge Process

The merge algorithm consists of the following steps:

Step 1: Consider any two or more neighboring subregions, $R_1, R_2, \cdots, R_n$, in the image

Step 2: If $H(R_1 \cup R_2 \cup \cdots R_n)$ is true, merge the n regions into a single region

Step 3: Repeat these steps until no further merges take place

$H(R_1 \cup R_2 \cup \cdots R_n)$: $H(R_1 \cup R_2 \cup \cdots R_n)_{\max}$ - $H(R_1 \cup R_2 \cup \cdots R_n)_{\min} < \mathrm{T}$

These methods can be used for a differentation between images. As a result, all homogeneous texts appearing in the image should be contained in some of the homogeneous segments.

2.2.2. Size Restriction and Dilation Process

We segment the input image into homogeneous gray scale segments in a Split/Merge process step. The segmented image now consists of homogeneous segment regions according to their gray tone intensity. The first step is size restrictions; some regions are too large and others are too small to be instances of texts, and therefore homogeneous segment regions whose width and height exceed max_threshold are removed, as are homogeneous segment regions less than min_threshold. Equation (2) defines a Mean value. The second step is image binairzation which is done with following process:

step 1 : if $B_i > Mean$ is true, a candidate Text region

else if $B_i < T_B$ is true, a candidate Text region

else non-candidate Text region

step 2 : Repeat these steps N times

$$Mean = \frac{1}{N}\sum_{i=0}^{N-1} B_i \qquad (2)$$

(N : Number of remaining segment regions in image, B_i : Mean bright value of remaining segment regions in image, T_B : Bright threshold value)

The result of the second step is a binary image containing segment regions. In the third step, a dilation process with a standard dilation algorithm is applied to the second step image. This step closes small holes in the components and connects components separated by small gaps.

2.2.3. Text Region Extraction

After applying size restrictions and dilation process, we use Blob coloring[11] for connected component labeling in a segmented image. It is of interest to assign a unique label to each region in a segmented image. Next, we calculate the mean of each labeling region by a Split and Merge process image above. Then, first candidate text region extraction is performed by Connect Region Mean Analysis. Calculated means of each labeling region are aligned horizontally or vertically.

We give each labeling region to the first candidate text region. We assume the colors of the characters in the same text regions are similar, and then characters are aligned horizontally or vertically in this paper. Final candidate text region extraction is performed by Region Fill Factor Analysis. The region fill factors are calculated again for first candidate text regions. If the fill factor is too low, the corresponding regions are discarded. Next, the height-to-width ratio of the labeling regions is calculated. If it exceeds certain limits, i.e. does not lie between min_number_ratio and max_number_ratio, the corresponding regions are also discarded.

Finally, the bounding boxes are grown horizontally 2 pixels to the left and 2 pixels to the right in order to compensate for the preprocessing step. Figure 4 shows the SMA results after the Split/Merge process and text region extraction. Figures 4a and 4b display the input image and the Split/Merge process image. Figure 4c shows the binarized image. Figures 4d and 4e display the size restriction and dilation process image and text region image. Figure 4f shows the input image with superimposed bounding boxes of the connected components.

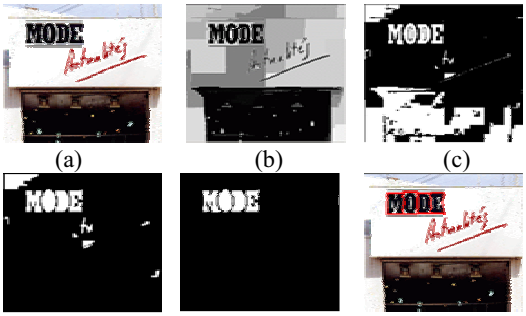


Figure. 4. . The input image (a), the Split/Merge image (b), the binarization image (c), the size restriction and dilation image (d), the text region image (e), the final result (f).

2.3 Hybrid Analysis Method

The GIA method is sensitive to the brick texture with horizontal and vertical lines, and the SMA method is sensitive to the color similarity between foreground and background regions in the image. The former method is rather robust to the weakness of SMA, and the latter robust to complex lines and complex texture. Thus, combining two methods can complement the shortcomings of each method. Hybrid Analysis Method(HAM) is shown in Figure 5. Figure 6a shows false accepted regions with GIA method. Then, Figure 6b shows false accepted regions with SMA method. Next, Figure 6c shows extracted text regions with HAM. We can see the usefulness of the HAM result is better than that of GIA and SMA in Figure 6.

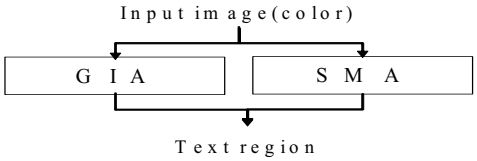


Figure 5. Algorithm of HAM method.

Figure 6. The false accepted region with GIA (a), The false accepted region with SMA (b) and the text region with HAM (c) : red box with GIA and green box with SMA.

3. Experimental Results

To evaluate the performance of the proposed system, we used 120 natural images and 251 images(2003 ICDAR Contest Test images) captured from schools, hospitals, subway stations, and streets. Test images have various sizes and include thousands of colors. Performance of the text extraction is measured in several terms. Sum is the number of total characters in images. True is the number of text regions correctly to detect. Part is the number of text regions correctly to detect part text region. Error is the number of texts failed to detect. False counts the number incorrectly identified as texts. Precision is True/(True+Part+False). Recall is True/Sum. Table 1 summarizes the performance of the proposed three methods and [9] for 120 natural images. Our HAM system works very well on candidate text region in simple and complex images. Also, it can be observed that the our HAM system results in a good tradeoff between the high True number and low Error number.

Table 1. Extraction result(region of character)

Image	System	Sum	True	Part	Error	False
Natural Images (120)	GIA	1045	829	54	162	222
		Precision = 75%		Recall = 79.3%		
	SMA	1045	759	50	236	284
		Precision = 69.4%		Recall = 72.6%		
	HAM	1045	923	34	88	254
		Precision = 76.2%		Recall = 88.3%		
	[9]	1045	866	49	130	450
		Precision = 63.4%		Recall = 82.8%		

Table 2 and 3 summarize the extraction text results for 251(2003-ICDAR Contest test) images. Table 2 compares Precision and Recall with Table 3.

Table 2. Extraction result of HAM(region of word)

Image	System	Sum	True	Part	Error	False
ICDAR Images (251)	HAM	721	464	76	181	284
		Precision = 56.3%		Recall = 64.3%		
	[9]	687	474	29	184	346
		Precision = 55.8%		Recall = 68.9%		

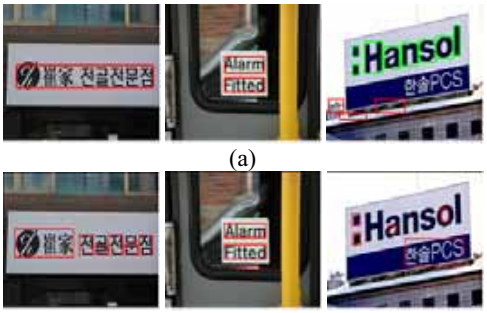
Our experimental results show that the proposed approaches are superior to conventional methods both in simple and complex images. Figure 7a and 7b show comparing HAM system results with [9] system.

Table 3. Extraction result of [12](region of word)

System	Precision	Recall
Ashida	55%	46%
HWDavid	44%	46%
Wolf	30%	44%
Todoran	19%	18%
Ours(HAM)	56.3%	64.3%

4. Conclusions

Text locating in natural scene image with complex background is a difficult, challenging, and important problem. In this paper, we have presented an accurate text region extraction algorithm based on three methods with gray-information. The proposed methods work very well on text region in simple and complex images. Texts in natural scene images contain very important information about location information and road signs. One of the further studies is to design the verifying extraction text region by SVM and HMM, and then to design the recognizer system for extraction text regions.



(a)

(b)

Figure 7. (a) results of HAM, (b) results of [9].

Acknowledgement

This work was supported by the Korea Research Foundation Grant (2004-041-D00631).

References

[1] Y. Zhong, H. Zhang and A. K. Jain, "Automatic Caption Localization in Compressed Video," *IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, No. 4*, pp. 385 - 392, 2000.

[2] Anil K. Jain, B Yu, "Automatic Text Location in Images and Video Frames," Pattern Recognition, Vol. 31, No. 12, pp. 2055 - 2076, 1998.

[3] J. Hoya, A. Shio and S. Akamatsu, "Recognizing Characters in Scene Images," *IEEE Trans. Pattern Analysis and Machine IntelligenceVol.16. No.2,* pp. 67 - 82, 1995.

[4] R. Lienhart, F. Stuber, "Automatic Text Recognition in Digital Videos*," Image and Video Proc. IV, SPIE*, 1996.

[5] S. Messelodi and C. M Modena, "Automatic Identification and Skew Estimation of Test Lines in Real Scene Images," *Pattern Recognition, Vol. 32, No 5*, pp. 701 - 810, 1999

[6] Y. Zhong, K. Karu and A. K. Jain, "Locating Text in Complex Color Images*," Pattern Recognition, Vol. 28. No. 10*, pp. 1523 - 1535, 1995.

[7] L. Gu and T. Kaneko, "Robust Extraction of Characters from Color Scene Images Using Mathematical Morphology," *Proc. of 14th International Conference on Pattern Recognition*, pp.1002 - 1004, 1998.

[8] H.K. Kim, "Efficient automatic text location method and content-based indexing and structuring of video database," *J.Visual Commun. Image Representation* 7(4), pp. 336 – 344, 1996.

[9] Y. Choi, "Scene Text Extraction in Natural Images Using Hierarchical Feature Combining and Verification," *The 2nd KAIST-Tsinghua Joint Workshop on Pattern Recognition*, pp. 76 - 102, Daejeon, Korea, 2003.

[10] R. Crane, *A simplified approach to Image Processing*, Prentice-Hall, 1997.

[11] D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, 1982.

[12] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, "ICDAR 2003 Robust Reading Competitions," *Proc. of 7th International Conference on Document Analysis and Recognition*, pp. 682-687, 2003.