Journal of Applied Science and Engineering, Vol. 21, No. 4, pp. 595–602 (2018) *DOI: 10.6180/jase.201812_21(4).0012*

Complex Text Location and Segmentation Approach through Region Fusion and Edge Features Extraction Method

Fen Cai[1], Yan-Min Luo[2], Pei-Zhong Liu[3] and Hsuan-Ming Feng[4]*

[1]*College of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, Fujian 362000, P.R. China*
[2]*College of Computer Science & Technology, Huaqiao University, Xiamen 361021, P.R. China*
[3]*College of Engineering, Huaqiao University, Quanzhou, Fujian 362021, P.R. China*
[4]*Department of Computer Science and Information Engineering, National Quemoy University, Kinmen, Taiwan, R.O.C.*

Abstract

Detecting and locating text information in complex background images has become an important technology for marking the image automatically. The traditional text segmentation method is difficult to obtain the ideal result in the text segmentation of complex background images. According to the features of complex background images, this paper puts forward a new segmentation method by fusing the edge and region features. Firstly, the initial localization of text region is extracted by using stroke filter based on the characteristics of Chinese characters. Secondly, the method of projection histogram analysis is used to remove the conglutination between Chinese characters and background, and the connected region analysis method based on region growing is used to recognize the single or multiple active targets. Finally, the final text segmentation is performed by the improved vertical projection segmentation algorithm which based on the structure of Chinese. The experimental results show that the proposed algorithm has good performance in text location, and can adapt to the characteristics of the complicated background and get accurate character image area comparing with the traditional edge detection method.

Key Words**:** Edge Feature, Region Feature, Stroke Filter, Text Segmentation

1. Introduction

Image text extraction and recognization is the important fields of artificial intelligence, which involves image processing, pattern recognition, philosophy, psychology, physiology, computer science and so on [1]. Color images often contain useful information, text information is one of them. The text in the image is important for image understanding. Text extraction, recognition and semantic analysis are of great significance for public opinion control, speech communication and image understanding. A lot of information is spread through the pictures and videos on the Internet, and those pictures and videos will contain many valuable information. The text in the image information is an important clue to help understand image semantic content [2]. Therefore, it is very significance to identify those text exactly which has great significance to the network purification.

With the development of computer technology, the traditional keyword-based retrieval cannot satisfy the user requirements. The technology of retrieving relevant data directly through images or videos is on the rise and development, which have important applications in many digital image library [3]. Therefore detecting and locating text information in complex background images has become an important technology for marking the image automatically using artificial intelligence technology and computer vision technology. This paper discusses the text

*Corresponding author. E-mail: hmfeng@nqu.edu.tw

location and segmentation approach of complex background images and puts forward a new segmentation method by fusing the edge and region features, which has good performance in text location, and can adapt to the characteristics of the complicated background.

2. Introduction to Text Segmentation and Extraction

Image text extraction and recognition is to pick up the text contained in the images, and then recognize the text and transform it into texts [4]. Text extraction includes text location and text segmentation. The segmentation is better, the subsequent recognition rate is higher. There usually are four classes of text extraction methods [5]: edge detection based text extraction [6], the connected domain analysis based text extraction [7], the color analysis based text extraction and texture feature based text extraction [8]. In recent years, people tried to combine multiple methods for text segmentation [9]. Zhang. presents a method for locating text based on a simplified pulse coupled neural network (PCNN) [10]. This method outperforms several traditional methods both in text detection rate and text detection accuracy. However, due to the instability of network memory, the influence of training samples on neural network is relatively large, so it is difficult to adapt to all kinds of complex background images. According to the features of complex web images, Liu proposed a text segmentation method based on the OTSU method, which can improves the segmentation character of simple background image. But for complex background images, the effect is not satisfactory [11].

The main problem of the text extraction is that those methods are too sensitive to the color, size, type of the text and the complex background, and the extracting results are not unsatisfactory when the background color is similar to the text color. To solve these problems, Shi proposed a new method for text extraction based on stroke filter which making use of the intrinsic characters of text [12]. Liu designed the image text extraction experiments based on stroke filter [13]. The experiment results showed that the method has strong robustness. However, in the experiments of Ref [12], the efficiency of the algorithm is low the algorithm due to executing the OCR module to amend the segmentation result. This paper propose a new method which combined the edge feature and region feature for text extraction. We use stroke filter to locate the position of the text, and propose a new method based on empirical value for text segmentation which can improve the segmentation accuracy and reduce the execution time.

Many studies have shown that the segmentation algorithm based on edge detection can achieve better segmentation results when the target and background has obvious difference, but when it come to the scene that the edge of the image is fuzzy, or the target and background have similar characteristics, or the segmented regions contain more than one edge, the segmentation algorithm based on edge detection can not get well performance. The region based segmentation approach can make good use of image region information, and make the segmentation to be continuous, and can get good segmentation results for a plurality of target segmentation situation. However, when the image background is similar to the character color, or the text is affected by noise, the region based segmentation approach is difficult to extract the text accurately.

By analysis of the features complex background images, this paper puts forward a new segmentation method by fusing the edge and region features. The proposed method is divided into two parts, text extraction and text segmentation. Based on the edge information of text, this paper use stroke filter which is based on the feature of characters to get the text regions. The text extraction of the proposed method includes image preprocessing, stroke filtering, denoising and connected domain analysis. Before the stroke filtering, we deal with the image by gray processing and gray stretching, which can retain adequate image details and reduce unnecessary interference and save execution time. In text segmentation, a experience value based projection segmentation algorithm is proposed according to the structure of Chinese characters, which can segment the Chinese characters effectively.

3. The Proposed Method

3.1 Text Location Based on Stroke Filtering

For making the image have high-contrast between background and foreground, the proposed algorithm transforms the color image into gray image. Then the stroke filter is utilized to locate the text region roughly.

3.1.1 The Definition of Stroke Filter

Stroke filter is the text filter used in the optical character recognition preprocessing, It can keep the visible parts of the text and filter out the parts with no obvious text features. Ref. [14] defined stroke filter as a straight line or a signal arc used in text segmentation. The text in the image is consist of one or a few stroke. The stroke filter is showed in Figure 1.

The red dot in stroke filter represents a pixel point (x, y), and it locates in the center of the filter. The filter has three rectangles as shown above and the center point is located in the center of the rectangle 1. The other two rectangles locate on two sides of rectangle 1. The size and direction of those rectangles are determined by the symbolic distance parameters d and direction parameter alpha. In this paper, the distance parameter d is determined by prior knowledge of image, the direction parameter alpha has four directions, they are 0°, 45°, 90°, 135°. In addition, the parameter $d1$, $d2$ and w as shown above are defined as followings: $d1 = d2 = d/2$, $w = 2d$.

First, we need to define the stroke response value. For the defined distance d and angle alpha, considering the text pixels might be dark or bright relative to the image background (white characters on a black background or black characters on a white background) ,the response value R of a pixel is defined in formula 2.1 and formula 2.2.

$$R^{B}_{(a,d)}(x,y) = \frac{u1 - u2 + u1 - u3 - |u2 - u3|}{o} \quad (1)$$

$$R^{D}_{(a,d)}(x,y) = \frac{u2 - u1 + u3 - u1 - |u2 - u3|}{o} \quad (2)$$

where $u1$, $u2$, $u3$ represent the number of the pixels of rectangle 1, rectangle 2, and rectangle 3 respectively, and o represents variance of pixels in rectangle 1. Formula 2.2 shows that if the center point is stroke pixel point, the difference values between the number of pixels of rectangle 1 and both of the pixels numbers of rectangle 2 and rectangle 3 will become larger, and also the difference value between the number of pixels of rectangle 2 and the number of pixels of rectangle 3 will become smaller. Therefore, the maximum of response value will be in the position of the stroke pixel. But due to the direction of the current stroke and font size, the response value R will be different, therefore we need to adjust the distance parameter d and direction parameter alpha to find the maximum value.

A pixel response value R is one of the biggest response values in all directions and distance parameter values. And the value is assigned to the original image after filtering. After filtering we set a threshold to attain binary image. By theoretical analysis and comparative experiments, we find that when the threshold is between 80~90, we can get the best effect.

3.1.2 The Comparison of Stroke Filter and Other Filter

We compare the stroke filter with other filters in this part to show the advantage of stroke filter. In Ref. [15] the author find that Canny filter will extract both the fence shape edge and step edge completely at the same time. Gabor filter is better at extracting the fence shape edge, but it takes hundreds of times of other filters. Laplacian filter is better at extract the filament and isolated point, but it is sensitive to noise. Sobel filter uses the fast convolution function to extract the edge which is simple and useful. Therefore sobel filer has been widely applied. But it can not distinguish the foreground and background strictly.

We compare sobel filter and stroke filter with testing results. The experimental results are showed as followings:

Figure 2 shows that the stroke filter extracts all the text region and remove most of the background. However, the sobel filter extracts both the edge of text and the edge of background, It can not distinguish the text region and the non-text region effectively. This paper uses the stroke filter to extract text region.

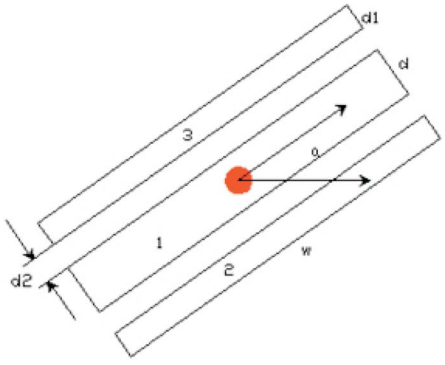
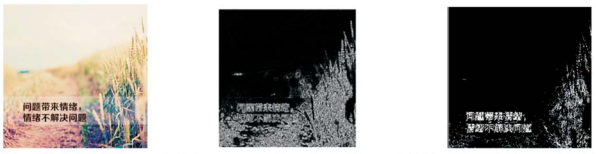


Figure 1. Stroke filter.



(a) original image (b) image with sobel filter (c) image with stroke filter

Figure 2. The comparison for sobel filter and stroke filter.

3.2 Denoising and Connected Domain Analysis Based on Image Region Information

After the stroke filter, there are always some noise and misjudgement region. In order to ensure the segmentation veracity, we need to deal with the image by denoising and removing the misjudgement region.

3.2.1 Denoising by Projection Histogram Analysis

The binary image just has two pixel value, one is the background pixel value and the other is the foreground pixel value. The background pixel value is 255 and the foreground pixel value is 0. We use the method of projection histogram to remove the noise and part of non-target region.

First, we set an array V to store the number of target pixel in each line. And then we scan image pixel in line-by-line, if the pixel value is 0, then the V plus one, if the pixel value is 255, the V plus zero. Thus the projection of text region is higher, the projection of background is zero in theory. Because of the noise, the background would have some pixel of value is 0. But comparing with the text region, the sum of noise value is very much smaller. So, we can set a threshold to remove noise interference. Through theoretical analysis and experimental verification, the implementation steps are showed in formula 2.3.

$$lpSrc = \begin{cases} 0 & V[i]/V[i+1] < 3 \\ 255 & V[i]/V[i+1] \geq 3 \end{cases} \tag{2.3}$$

where $lpSrc$ represent the pixel value after newly calculated.

3.2.2 Connected Domain Analysis

Connected domain analysis [16] is based on region growing algorithm, and its purpose is to recognize single or multiple target in images by calculating the number of connected domain. Connected domain analysis algorithm scans image pixel by pixel, and then analyzes the eight-neighbors of each pixel to get the connected domain division. The concrete realization steps are showed as followings.

Step 1. Check whether the pixel is the target pixel in the image, if yes, then insert it to the target segment list, if not, check the next pixel.

Step 2. Check whether there has unmarked pixel in eight-neighbor pixels. If yes, mark all the unmarked target pixels, and create new nodes and insert it into the list of the current position. If all the eight-neighbor pixels have be marked, move to the next node of the list.

Step 3. Judge whether the current pixel is the root node. If not, go to step (2), else, the algorithm is end.

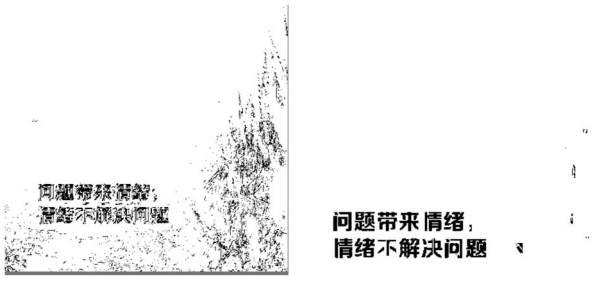
By means of denoising and connected component analysis, the impulse noise is completely removed and the text region become clear. Figure 3 shows the result.

In Figure 3, picture (a) is the image after stroke filter but before denoising and connected component analysis. Picture (b) is the image after denoising and connected component analysis. It is clearly to see that the effect after denoising is better than before.

3.3 The Improved Vertical Projection Segmentation Algorithm

3.3.1 Traditional Characters Projection Segmentation Method

Traditional characters projection method usually converts the original image into binary image, and then perform the projection segmentation on the binary image is used in binary image [17]. First, this method scans the image pixel in line-by-line, and adds the pixel point whose value is 0 up, and sets the final value as the projection in this column. After scanning all the columns of the image,



(a) before (b) after

Figure 3. Comparison diagram of denosing.

the vertical projection histogram of whole image can be achieved. Due to the number of pixel whose value is 0 in the parting of the characters is less, this part is at a low ebb in vertical projection histogram, the column where the number of pixels is 0 is the boundaries of character segmentation.

However, there are two kinds of font structures in Chinese characters: left and right structures and upper and lower structures. Therefore, in the traditional projection method, the projection results of these two structures will result in that not all the pixels value in target region is 0, and a background area with a pixel value of 255 would appear inside the character area. Therefore if we simply divide the characters based on the valley of the traditional character projection, it is easy to produce a mis-segmentation phenomenon, and separates individual characters into multiple parts. Comparing with the vertical projection segmentation method, top and bottom edge projection segmentation has a better single crest characteristics which can be more easy to segment out the whole character region.

3.3.2 Projection Segmentation Algorithm Based on Experience Value

In this paper, by analysis of the structure of Chinese characters, we introduce the experience value as the segmentation condition to screen out the character components. And for the up and down or so edge projection segmentation method has a good single wave characteristics, and can be easier to segment the whole regional characters, we perform the Chinese characters segmentation based on this algorithm.

The mean of top and bottom edge projection is calculating the distance between the top edge and bottom edge of the character. Besides, we add the left and right edge projection to the method. That means we also need to calculate the distance between the left edge and the right edge of the character.

First, we need to find the initial position of top edge and bottom edge. The steps are showed as followings:

Step 1. Starting from the first pixel of the image, scan the pixels in line-by-line and mark the first pixel whose value is 0 as *t*.

Step 2. Keep scanning in line-by-line until finding the first pixel whose value is 255 and mark it as *b*.

Thus we haply find the top edge and bottom edge of a character. Then we need to find the left edge and right edge of the character. The method is similar to the steps above. which is scanning the pixels in row-by-row from the first pixel of the image and mark the first pixel whose value is 0 as *l*, and keep scanning until finding the first pixel whose value is 255 and mark it as *r*. At this point the four boundaries of the character are found. Then we repeat the above method and continue searching for the boundary of next character.

Because Chinese characters are composed of multiple parts, there will be more than one wave crest in the projection result. It is possible that a complete Chinese characters will be divided into more than one part. In order to avoiding this, on the basis of projection segmentation, we add the empirical value as screening condition. According to the characteristics of Chinese characters, for the common typeface, the difference value of left and right boundary is greater than two-thirds of the difference value of top and bottom boundary. So, if the condition is not met, we need to keep scanning towards the right. Also in general, the difference value of left and right boundary is less than four-thirds of the difference value of top and bottom boundary. If the condition is not met, it indicates that the segmented character is more than one character, and we need to scan again within the borders. In terms of visual effects, if the character is too small, it is hard to recognize it. By analyzing hundreds of pictures, we set the minimum threshold of the difference value of left and right boundary as 10 and set the maximum threshold of the difference value of left and right boundary as 64. In other words, the difference value of left and right boundary is more than 10 and less than 64. The threshold of the top and bottom boundary is the same. Figure 4 shows the process of segmentation algorithm.

In the process of characters segmentation, we find the approximate location of the characters by using the projection analysis method at the first time. To confirm the character boundary further, we use the method again within the border which was confirmed before. Thus we get the accurate segmentation result eventually.

4. Experimental Results and Analysis

In order to verity the feasibility and accuracy of the

The experience value based projection segmentation algorithm
CharSegment(){

```
begin
  scan each pixel of image from top to bottom, if the pixel of line i whose value is 0 then vertical [i]++;
  while k does not exceed the scope of image line
    Finding the top bounds of the character border: line t;(finding the first k which satisfies vertical[k]=0, let t=k)
    Finding the bottom bounds of the character border: line b;(finding the last k which satisfies vertical[k]=0;let b=k)
    if (t<b) then
      scan each pixel of image from left to right, if the pixel of row i whose value is 0 then horizon [i]++;
      Finding the left bounds of the character border: row l;(finding the first j which satisfies horizon [j]=0;let l=j)
      Finding the right bounds of the character border: row r;(finding the last j which satisfies horizon [j]=0;let r=j)
      if  (r-l)<2*(b-t)/3   then
        add the border to linked list;
        if  (r-l)>3*(b-t)/4   then
            scan and segment again in the frame;
        else
            set the pixel value as 0;
        endif
      endif
    else
      set the pixel value as 255;
    endif
  endwhile
end
```

Figure 4. Process of the experience value based projection segmentation algorithm.

proposed algorithm, we select 80 images from web page about life, entertainment, news and so on to carry out experiment. Experimental platform is the computer with 3.3 GHz dual core processor, 4 GB RAM.

First we use the proposed method to test multiple different types of images. We set the recognition rate R and accuracy rate E as the benchmark. The calculation of R and E are showed in formulas 5.1 and 5.2.

Set N as the total number of all the characters in the image. C presents the total number of recognized characters. D presents the total number of characters which were finally segmented accurately.

$$R = D/N \times 100\% \quad (5.1)$$

$$E = D/C \times 100\% \quad (5.2)$$

The experiment is with 50 pieces of simple images and 30 pieces of complex images which is mixed of text and graphics. The test results are showed in Table 1.

The experimental results are showed in Figure 5.

In the pictures above, the picture (a), picture (d), picture (f) are original images, picture (b), picture (e), picture (h) are the text extraction results, picture (c), picture (f), picture (I) are the text segmentation results. From the experimental results it can be found that if the back-

Table 1. The results of different kinds of images

Image type	Recognition rate R				Average recognition rate	Accuracy rate E
	> 95%	95%-80%	80%-70%	< 70%		
Simple image	48	2	0	0	97%	99%
Complex image	14	10	6	0	90%	95%

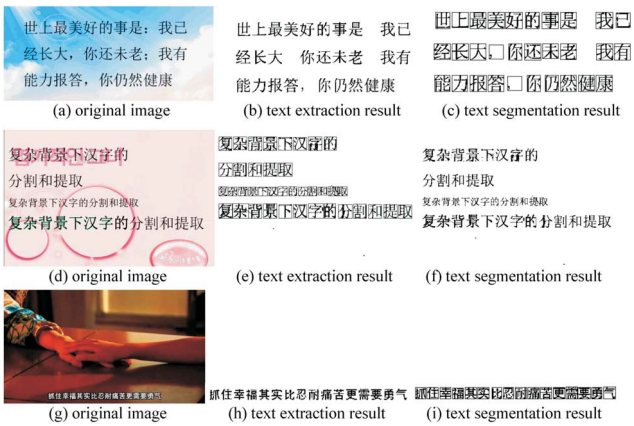
ground is simpler, the segmentation result is better. And the proposed method have a good recognition rate and segmentation rate in different kinds of images.

Besides, we compare the proposed method with the method of Ref. [12] and Ref. [18]. The comparing results are showed in Table 2. Figure 6 shows the results of different methods.

It can be found by comparing the results in Figure 6 that the characters stroke and the characters outline in picture (e) are more clear than that in picture (c) and picture (d). It indicates that the method we proposed has a better extraction effect which is important for the text segmentation. The Table 2 shows the segmentation accuracy rate of different method. It can be seen obviously that the method we proposed has the highest accuracy rate among three methods.

5. Conclusions

Text extraction and segmentation is the key to character recognition in complex background images. In order to effectively detecting and locating text regions in complex background images, This paper proposed a complex text location and segmentation method by fusing the edge and region features. On which the stroke filter is used to extract the text region. The vertical projection and connected component analysis is carried out to remove the interference of noise and non-target region. After text extraction, we proposed a simple and useful method based on empirical value to segment out single Chinese character. For the structural diversity of Chinese characters. The experience value based projection segmentation algorithm was carried out to segment the Chinese characters effectively. The experiment results show that the method can get the accurate text region and have a high segmentation accuracy rate. In the future research we will further improve the robustness of the algorithm and widely spread it in the field of text segmentation in more scenarios.



(a) original image (b) text extraction result (c) text segmentation result

(d) original image (e) text extraction result (f) text segmentation result

(g) original image (h) text extraction result (i) text segmentation result

Figure 5. The result of segmentation.

Table 2. The segmentation accuracy rate of different method

Method	Segmentation accuracy rate
Ref [12]	96.8%
Ref [18]	86.3%
Proposed method	97.5%

Acknowledgements

This work was supported by the Quanzhou scientific and technological planning projects of Fujian, China (No. 2015Z40), the Natural Science Foundation of Fujian Province, China (No. 2015J01256), and the Talent project of Huaqiao University (No. 14BS215).

(a) original image

东风雪铁龙发布新爱丽舍两厢上市

(b) text image to be segmented

东风雪铁龙发布新爱丽舍两厢上市

(c) the result of Ref [3]

东风雪铁龙发布新爱丽舍两厢上市

(d) the result of Ref [13]

东风雪铁龙发布新爱丽舍两厢上市

(e) the result of proposed method

东风雪铁龙发布新爱丽舍两厢上市

(f) the final segmentation result

Figure 6. Experimental comparison.

References

[1] Wang, Y., H. Zheng, and D. W. Hu (2004) The Technology to Acquire Text in Image and Vedio, *Journal of Image and Graphic* 9(12), 532–538.

[2] Zhong, Y., K. Karu, and A. K. Jain (1995) Locating Text in Complex Color Images, Proceedings of the 3rd International Conference on Document Analysis and Recognition 28(10), 1523–1535. doi: 10.1109/ICDAR.1995.598963

[3] Botz-Bornstein, T., and M. M. Mostafa (2017) A Corpus-based Computational Analysis of Philosophical Texts: Comparing Analytic and Continental Philosophy, *International Journal of Social and Humanistic Computing* 2(3), 230–246. doi: 10.1504/IJSHC.2017.10005737

[4] Zhu, C. J., C. Li, and Z. Xiong (2007) Text Detection and Recognition Technique in Video, *Computer Engineering* 33(10), 218–219.

[5] Keechul, J., K. Kwang, and K. J. Anil (2004) Text Information Extraction in Images and Video: a Survey, *Pattern Recognition* 37(5), 977–997. doi: 10.1016/j.patcog.2003.10.012

[6] Lienhart, R., and A. Wernicke (2002) Localizing and Segmenting Text in Images and Videos, *IEEE Transactions on Circuits and Systems for Video Technology* 12(4), 256–268. doi: 10.1109/76.999203

[7] Mariano, V. Y., and R. Kasturi (2000) Locating Uniform-colored Text in Video Frames, IEEE Proceedings of the 15th International Conference on Pattern Recognition 539–542. doi: 10.1109/ICPR.2000.902976

[8] Zhong, Y., H. J. Zhang., and A. K. Jain (2000) Automatic Caption Localization in Compressed Video, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(4), 385–392. doi: 10.1109/34.845381

[9] Pan, Y. F., X. Hou, and C. L. Liu (2011) A Hybrid Approach to Detect and Localize Texts in Natural Scene Images, *IEEE Transactions on Image Processing* 20(3), 800–813. doi: 10.1109/TIP.2010.2070803

[10] Zhang, X., and F. C. Sun (2011) Pulse Coupled Neural Network Edge-based Algorithm for Image Text Locating, *Tsinghua Science & Technology* 1(16), 22–30. doi: 10.1016/S1007-0214(11)70004-9

[11] Liu, P. Z., X. Ning, and W. J. Li (2014) Approach for Text Segmentation in Web Image, *Computer Engineering and Applications* 50(10), 180–184.

[12] Shi, Z. G., and L. Q. Gao (2010) New Algorithm for Text Segmentation Based on Stroke Filter, *Computer Science* 37(1), 287–289.

[13] Liu, C. P., G. Liu, C. H. Wang, and Y. Q. Zhu (2013) Design of Image Text Extracting Experiment Based on Stroke Filter, *Research and Exploration in Laboratory* 11, 89–92.

[14] Liu, Q., C. Jung, and Y. Moon (2006) Text Segmentation Based on Stroke Filter, Proceedings of ACM International Conference on Multimedia, USA, 129–132.

[15] Liu, Q., C. Jung, S. Kim, et al. (2006) Stroke Filter for Text Localization in Video Images, Image Processing, 2006 IEEE International Conference on. IEEE, 1473–1476. doi: 10.1109/ICIP.2006.312560

[16] Yao, J. L., L. B. Weng, and X. H. Wang (2012) A Text Region Location Method Based on Connected Component, *Pattern Recognition and Artificial Intelligence* 25(2), 325–331.

[17] Cao, W. (2010) Character Vertical Projection Segmentation Based on Dichotomy, *Software Cuide* 10, 71–72.

[18] Shi, J., and J. Malik (2000) Normalized Cuts and Image Segmentation, *IEEE Transactions on Patten Analysis Machine Intelligence* 22(8), 888–905. doi: 10.1109/34.868688

Manuscript Received: Nov. 30, 2017
Accepted: Sep. 11, 2018