A new algorithm for text segmentation based on stroke filter

ZHANG Xiu-li , SHI Zhen-gang

College of Information Science and Engineering ,Shenyang Ligong University, Shenyang, 110168,China
E-mail: syluzx@yahoo.cn

Abstract: In order to solve segment text accurately and robustly from a complex background. This paper propose a new algorithm for text segmentation in images by using a stroke filter. In this paper, we design a new algorithm for text segmentation by using the stroke filter. First, we describe the stroke filter briefly based on local region analysis. Second, the determination of text color polarity and local region growing procedures are performed based on the response of the stroke filter. Finally, the feedback procedure by the recognition score from an optical character recognition (OCR) module is used to improve the performance of text segmentation. The proposed algorithm is compared with other algorithms. The experimental results demonstrate that the proposed algorithm obtained satisfactory results.

Key Words: Text segmentation, Complex background, Optical character recognition(OCR), Stroke filter

INTRODUCTION

The extraction of text information is very important because texts in images and videos contain important and useful information. In general, the extraction of text information from images includes three major steps: text localization, text segmentation, and text recognition[1-3]. In this paper, we focus on the text segmentation, which is employed to separate text pixels from the background in a text image.

The text segmentation methods was classified into two main methods. The first method is based on the difference in contrast between the foreground and background, for example, the fixed thresholding method[4], Otsu's adaptive thresholding method[5], global and local thresholding method[6]. In general, these methods are simple and fast; however, they tend to fail when the foreground and background are similar. The second method is the similarity-based method clusters pixels with similar intensities. For example, Lienhart[7] used the split and merge algorithm, and Wang et al.[8] used a method in which edge detection, watershed transform, and clustering were combined. However, these methods are unstable because they exploit many intuitive rules for the text shape. The main problem in most of the existing methods is that they are sensitive to text color, size, font, and background clutter; this is because these methods simply exploit the general segmentation methods or the prior knowledge about the text shape.

In this paper, we attempt to consider the intrinsic characteristics of text and design a algorithm particularly for text segmentation. The paper is organized as follows: in Section 2, we explain the stroke filter in brief. In Section 3, we describe the text color polarity determination and local region growing procedures and the feedback procedure by the recognition score from an OCR module. Experimental results are given in Section 4, followed by concluding remarks in Section 5.

This work is supported by National Nature Science Foundation under Grant ×××××××

Design of stroke filter

A stroke is defined as a straight line or arc used as a segment of text, and the texts in images comprise one or several strokes. An image is defined as a text, if and only if, several stroke-like structures exist in it. A stroke filter is designed based on this definition using the local region analysis. In order to design the stroke filter, we first define a local image region to be a stroke-like structure, if and only if: (1) it is different from its lateral regions, (2) intensities of its lateral regions are similar, and (3) it is nearly homogenous with respect to its intensities.

We compute stroke filter response of each pixel in the source image as shown in Fig. 1.

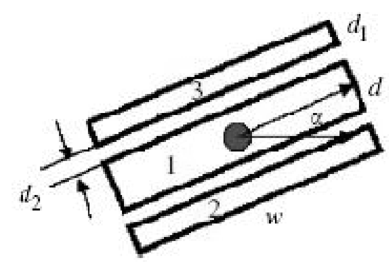

Fig. 1 A stroke filter

The central point of a stroke filter denotes an image pixel (x, y) around which three rectangular regions can be observed. Let index 1 denote the central region and indices 2 and 3 denote each of the lateral regions. The orientation and scale of these local regions are determined by the parameters α and d, where d, the width of the rectangular region, is determined based on prior knowledge of the text obtained by conducting experiments on text images. The distance, d2, between the central region and lateral regions in the stroke filter is due to the fact that dark or bright lines are often embedded around some texts to convey their meanings efficiently, and blurred edges appear around texts as a result of their compression, d1=d2=d/2, w=2d. According to the definition of a stroke-like structure, we

978-1-4244-5182-1/10/$26.00 ©2010 IEEE

define the bright and dark stroke filter responses of the pixel (x, y), RB and RD, respectively, as follows:

$$R_{\alpha,d}^{B}(x,y)=\frac{\mu_1-\mu_2+\mu_1-\mu_3-|\mu_2-\mu_3|}{\sigma}$$

(1)

$$R_{\alpha,d}^{D}(x,y)=\frac{\mu_2-\mu_1+\mu_3-\mu_1-|\mu_2-\mu_3|}{\sigma}$$

(2)

where μ_i denotes the estimated mean of the intensities in the region i, where i = 1, 2, 3. The parameter σ denotes the standard deviation of the intensities in region 1 and is a measure of the extent to which the intensities of the region are spread out. By means of stroke filtering, we extract the stroke features (RB, OB, SB, RD, OD, SD) of any pixel (x, y) as follows :

$$R^{B}(x,y)=\max_{(\alpha,d)} R_{\alpha,d}^{B}(x,y)$$

(3)

$$O^{B}(x,y)=\arg_{(\alpha)}\max R_{\alpha,d}^{B}(x,y)$$

(4) $$S^{B}(x,y)=\arg_{(d)}\max R_{\alpha,d}^{B}(x,y)$$

(5)

$$R^{D}(x,y)=\max_{(\alpha,d)} R_{\alpha,d}^{D}(x,y)$$

(6)

$$O^{D}(x,y)=\arg_{(\alpha)}\max R_{\alpha,d}^{D}(x,y)$$

(7) $$S^{D}(x,y)=\arg_{(d)}\max R_{\alpha,d}^{D}(x,y)$$

(8)

where R, O, and S, respectively, denote the response, orientation, and scale of the stroke filter, whereas B and D denote the bright and dark stroke filters, respectively. Next we should introduce the algorithm for text segmentation based on the stroke filter.

The algorithm for text segmentation based on stroke filter

3.1 Determine the color polarity of text

It is very important to determine the color polarity of text for the correct text segmentation[9-10]. In order to determine the text color polarity automatically, we use two features as follows. First we perform bright and dark stroke filtering to obtain RB and RD. Then, two features for the determination of text color polarity can be obtained, the first one FR of which is the ratio of the sums of the magnitude of the bright and dark stroke filter responses:

$$F_R=\frac{{}_{(x,y)}R^{(B)}(x,y)}{{}_{(x,y)}R^{(D)}(x,y)}$$

(9)

The second feature FE, inspired by Ref. [9], is the ratio of the sums of the number of edge points in the binary response maps of the bright and dark stroke filters:

$$F_E=\frac{N^{(B)}}{N^{(D)}}$$

(10)

where N(B) and N(D) denote the number of edge points in the binary map of the bright and dark stroke filter responses, respectively.

FE is useful for the case of bright texts on a bright background or dark texts on a dark background. The bright (dark) texts have few edge points in the binary bright (dark) stroke filter response map.

3.2 The method of local region growing

According to section 3.1 we can obtain the initial segmentation result. But the result has many text pixels tend to be missed in the binary stroke filter response map, and the missed text pixels should be recalled for accurate text segmentation. In order to recall these pixels, we perform a local region growing procedure. The binary response map is regarded as the initial segmentation result for the local region growing procedure. From the initial segmentation result and text image, we estimate the global probability density function (PDF). Then, a non-text pixel is changed to the text pixel if the following three conditions are simultaneously satisfied: (1) the number of textpixels is greater than 3 in the local region (3×3 neighbors of the non-text pixel), (2) the probability of the non-text pixel's intensity, Pr(s), is greater than a threshold Th1, and (3) the difference of intensity between the non-text pixel and its neighbors is lower than a value Th2. The local region growing procedure is repeated until no pixel is changed in the segmentation map.Therefore, the novelty of the local region growing procedure is that it is based on the stroke filter response and combines the global PDF and local similarities to achieve a reliable performance.

In this paper, the local region growing procedure will be used, and the procedure is summarized as follows:

The white and black pixels represent the text and non-text pixels, respectively. Input: I—initial segmentation result ; S—source text image.

Step 1: Estimate PDF of text color from I and S.

Step 2: For each white pixel in I, if the number of white pixels in its 3×3 neighbors is within [3,9], then go to Step 3, else Step 2.

Step 3: For each black pixel in the 3×3 regions, if it is: (1) similar to its text neighbors and (2) of high probability according to PDF, then it is marked as text. Repeat Steps 3 and 2, until no pixel is changed.

Output: Refined segmentation result.

3.3 Adjustment of the segmentation result

Usually, during the determination of the text color polarity, it occurs 3, 4% errors, and these errors will have a detrimental effect on the recognition results. In order to improve the accuracy of the determination of text color polarity and the performance of text segmentation, we apply an additional verification step for the segmentation result of the local region growing procedure by an OCR module. We use the feedback of an average recognition score of characters from the OCR module, and the average recognition score is determined as follows.

In general, there are several characters in the segmented text image, and a character separation function that can decompose the segmented text image into sub-images of individual symbols is required. Character separation is one of the most important processes in the OCR module because the performance of character separation significantly affects the accuracy of character recognition. Our character separation function uses the feedback from the recognition score of the character recognition function; furthermore, this method employs a split-merge strategy in which the split segments are merged into a character by using the recognition score obtained by the character recognition function[11]. The optimal separation path is determined by using two scores: (1) the geometric score that is used to estimate the likelihood of "being a character" by geometric features and (2) the recognition score obtained by the character recognition function. The geometric score is calculated by using two character evaluators, the squareness (SQU) and the internal gap (GAP), which are estimated by the Parzen window[12]. If the geometric score is low, the sub-image is eliminated in the separation paths and not recognized, by which the overhead of recognition can be reduced. The recognition score is determined by the distance between the recognition model and a separated sub-image. Here, the extracted feature is the angular directional feature (ADF)[13], and each character is recognized by linear discriminant analysis (LDA) in the character recognition function.

The average recognition score, SA, of the segmented text image is represented by these two scores as follows:

$$S_A = \frac{1}{N} \sum_{i=1}^{N} S_i$$

(11)

$$S_i = K_G \times S_G + K_R \times S_R$$

(12)

where i and N denote the index and the number of characters, KG and KR are constants and Si, SG, and SR denote the total, geometric, and recognition scores of each character, respectively. KG and KR, are usually assigned as 0.3 and 5, respectively. If SA is smaller than a predefined threshold ThR, the color polarity inversion and subsequent procedures are performed, where the threshold ThR is determined by analyzing the statistical distribution of the average recognition scores of the training samples .

Experimental results and analysis

In this section, we test the performance of the proposed approach. We compare our text color polarity determination method with the state-of-the-art method proposed by Song [9] , which is based on the statistic analysis. The accuracy of each method is evaluated as the relative frequency of the correctly determined text color polarities as follows:

$$\text{Accuracy} = \frac{\text{Number of correctly determined color polarity}}{\text{Number of text image}}$$

(13)

Table 1 shows the experimental results of our method and the Song method. From this table, we can see that our stroke filter based determination method achieves a considerably higher accuracy than the statistic based Song method. However, it can be observed that the processing time for our method is slower than that for the Song method. In order to evaluate the performance of our text segmentation method, the character error rate (CER) is calculated for another test. Here, CER is defined as

$$CER = \frac{N_e}{N}$$

(14)

where Ne denotes the number of characters wrongly recognized by the OCR module and N denotes the total number of characters. In this experiment, the proposed text segmentation method is compared with the Otsu [5] thresholding method. The Otsu method is a simple but a classic solution that is employed by many text segmentation schemes. Since this method does not deal with the determination of color polarity, we used our proposed color polarity determination algorithm. The evaluation result of the CER of these two methods is summarized in Table 2. From this table, we can see that our method achieves significantly lower error rates for both English and Chinese characters. This is because our method possesses the capability of handling multilanguage and complex backgrounds by using the stroke filter. Notice that in most cases, the CERs of Chinese characters are larger than that of English characters and this seems to be due to the fact that Chinese characters are structurally more complex than English characters.

Tab1. Performance comparison of two algorithms

Method	Accuracy (%)	Time cost per frame (ms)
Our method	95.1	16.2
Song method	90.3	11.3

Tab2. CER evaluation of two text segmentation methods

Method	English CER (%)	Chinese CER (%)
Our method	12.6	15.7
Otsu method	15.9	61.2

Conclusions

A new method of text segmentation in images using a stroke filter is proposed in this paper. The benefit of the stroke filter is that it is capable of discovering the intrinsic characteristics of the text by making use of the relationship among the local regions. By using the response of the stroke filter, we could determine the text color polarity robustly with the accuracy of 95.1%, which exceeds that of the Song method by 4.8%. When the feedback procedure from an OCR module is included, the accuracy of our method even more improves up to 97.9%. We also make use of the response of the stroke filter as the initial seed region for the local region growing procedure. In this procedure, we obtain a reliable performance by combining the global PDF and local similarities. Our method is capable of robustly handling texts from complex backgrounds by using the stroke filter. In addition, we believe that the stroke filter will play an important role in fields other than text segmentation. For example, the stroke filter could be applied to text verification using the features of the filter, or extended to the fields of line-like structure detection, such as road detection from remote images.

Our future study will mostly be focused on improving the OCR module. Although current OCR module is sufficiently efficient, it still leaves quite a room for improvement in the recognition rate. Currently, tensor techniques are newly proposed for image representation and object recognition, and we will continue developing our algorithm,and developing OCR module by using them .

REFERENCES

[1] N. Dimitrova, H.J. Zhang, B. Shahraray, I. Sezan, A. Zakhor, T. Huang, Applications of video content analysis and retrieval, IEEE Multimedia 9 (3) (2002): 43–55.

[2] Y. Wang, Y. Liu, J.C. Huang, Multimedia content analysis using both audio and visual clues, IEEE Signal Process. Mag. 17 (6) (2000): 12–36.

[3] R. Liehart, A. Wernicke, Localizing and segmenting text in images and videos, IEEE Trans. Circuits Syst. Video Technol. 12 (4) (2002): 256–268.

[4] T. Sato, T. Kanade, E.K. Hughes, M.A. Smith, S. Satoh, Video OCR: indexing digital news libraries by recognition of superimposed caption, ACM Multimedia Syst. Special Issue Video Libr. 7 (5) (1998): 385–395.

[5] N. Otsu, A threshold selection method from gray-scale histogram, IEEE Trans. Syst. Man Cybern. 9 (1979): 62–66.

[6] F. Chang, G.C. Chen, C.C. Lin, W.H. Lin, Caption analysis and recognition for building video indexing system, Multimedia Syst. 10 (4) (2005): 344–355.

[7] R. Lienhart, Automatic text recognition in digital videos, in: Proceedings of SPIE Image Video Processing, vol. 4, 1996, pp. 2666–2675.

[8] K. Wang, J.A. Kangas, W. Li, Character segmentation of color images from digital camera, in: Proceedings of the International Conference on Document Analysis and Recognition, 2001, pp. 210–214.

[9] J. Song, M. Cai, M.R. Lyu, A robust statistic method for classifying color polarity of video text, in: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, 2003, pp. 581–584.

[10] S. Antani, D. Crandall, R. Kasturi, Robust extraction of text in video, in: Proceedings of the International Conference on Pattern Recognition, 2000, pp. 831–834.

[11] Y.H. Tseng, H.J. Lee, Recognition-based handwritten Chinese character segmentation using a probabilistic Viteribi algorithm, Pattern Recognition Lett. 20 (1999): 791–806.

[12] M.S. Kim, K.T. Cho, H.K. Kwag, J.H. Kim, Segmentation of handwritten characters for digitalizing Korean historical documents, in: Proceedings of the International Conference on Document Analysis and Recognition, 2004, pp. 114–124.

[13] K.T. Lim, A study on machine printed character recognition based on character type classification, J. Inst. Electron. Eng. Korea 40 (2003): 26–39.