A New Text Extraction Method Incorporating Local Information

Yang.Zhang Chunheng.Wang Baihua.Xiao Cunzhao.Shi
Institute of Automation Chinese Academy of Sciences, People's Republic of China
yang.zhang@ia.ac.cn

***Abstract*—Text detection and extraction in images with complex background can provide useful information for video annotation and indexing. More attention is paid to text detection for its importance, but text extraction is necessary for the text recognition, and it can test the validity of text detection. In this paper, we conclude text extraction is to segment the image and to remove noises, and then a robust text extraction method incorporating local information is proposed. First, we get the gray image from the original image and reprocess the gray image with edge enhancement. Then a binarization method incorporating local information is used to segment the gray image, by which the text-noises are removed and a binary image is obtained. Finally, the connected component analysis based on the character's density and geometric feature is performed on the binary image, by which background-noises are removed. The preliminary experiments show some promising results.**

Keywords-component; text extraction, text segmentation, binarization, local information, CCA

I. INTRODUCTION

With the development of digital devices and internet, multimedia such as image and video has become one of the most important information carriers. The need to efficiently index, browse, retrieve the multimedia information is increasing. As text in image or video frame can provide useful information for image or video indexing and summarization, text detection and extraction in image have attracted more and more researchers' attentions [1]. But text extraction in image with complex background is a challenging problem because of many possible sources of variations: such as color, contrast, resolution, text font and size, and orientation etc. The main component for text extraction in image is to segment text from background, which is preprocessed by standard optical character recognition (OCR) software. In order to obtain satisfactory results when applying the OCR, the text must be separated from the background and transformed to a binary image with black characters in white background.

To solve this challenging problem, many methods have been proposed. The common proposal is to separate the text and background by binarization method based on global [2]or local (adaptive) [3] threshold. But no matter global threshold or local threshold method, it is difficult to get satisfactory result because they can hardly remove all the background noise. To better separate the text and background, Liu et. al [4] proposed a text segmentation method based on stroke filter which they assume can describe the intrinsic characteristic of text in terms of scale, orientation and response. But this method is unstable when text size is small and the region is mixed with many different colors; because it is hard to find the feature which can totally represent all kinds of text in image with complex background. Lyu et. al [5] proposed an effective text segmentation method by adaptive threshloding, dam point labeling and inward filling. But this method must normalize the original image to a small size to label the dam point, this may lead to more overlap noises which is not good for character recognition.

We conclude the text extraction is to filter the noises of image which can be classified to two kinds. One kind is the text-noise which is caused by low image quality. And another is the background noise caused by complex background in the real image. In this paper, we propose a robust approach to extract text from the image with complex background. Firstly a binarization incorporating local information is used to separate text and complex background which also remove the text-noise, then a method based on connected component analysis (CCA) is used to filter the background-noise. Experiments on the set of video frames show that the promising results can be achieved by our proposed method.

The rest of paper is organized as follows. Section 2 describes the detailed algorithm. Experiment results on the experimental image and real image dataset are shown in section 3.Section 4 concludes the paper and outlines areas for future research.

II. DETAILS FOR TEXT EXTRACTION

Our proposed method is robust to segment text in image with complex background, which can be in arbitrary font, color, orientations and size (there is a minimal height limit,

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.164

which is reasonable as too small text also cannot be distinguished by human). This method is mainly divided into three steps described in details below.

A. Edge Enhancement

In order to separate the text and background into different connected components, it is necessary to enhance the text edges in the image. Here we adopt the simple high-pass filtering by spatial convolution, and then the filtered image is added to the original image to get the enhanced image. The spatial convolution mask is as Figure 1.

0	-0.2	0
-0.2	0.8	-0.2
0	-0.2	0

Figure 1 the convolution mask

B. Binarization incorporating local information

After the edge enhancement on the image, a binarization method incorporating local information is used to binarized the gray image and remove the text-noise. Here we adopt the local similarity measure S_{ij}, which is proposed in [9], instead of fuzzy c-means clustering, global threshold method [2] (OTSU) is used to separate the weighted image.

Step 1: Compute the local similarity measures S_{ij} using Eq. (1) for neighbor windows over the image.

$$S_{ij} = \begin{cases} S_{s\_ij} \times S_{g\_ij}, & j \neq i, \\ 0, & j = i, \end{cases} \quad (1)$$

Where the *i*th pixel is the center of the local window (3×3), and *j*th pixel are the set of the neighbors falling into a window around the *i*th pixel. S_{s_ij} and S_{g_ij} is computed using Eq. (2) and (3).

$$S_{s\_ij} = \exp\left(\frac{-\max(|p_j - p_i|, |q_j - q_i|)}{\lambda_s} \right) \quad (2)$$

Where (p_i, q_i) is a spatial coordinate of *i*th pixel, λ_s is fixed to 3.

$$S_{g\_ij} = \exp\left(\frac{-\|x_i - x_j\|^2}{\lambda_g \times \sigma^2_{g\_i}} \right) \quad (3)$$

Where σ_{g_i} is defined by Eq. (4).

$$\sigma_{g\_i} = \sqrt{\frac{\sum_{j \in N_i} \|x_i - x_j\|^2}{N_R}} \quad (4)$$

Where x_i is gray value of the central pixel within a special window, x_j is gray value of the *j*th pixel in the same window, N_i denotes the neighborhoods of *i*th pixels falling into the local window and N_R is its cardinality. In our experiments, we fix the parameter λ_g and λ_s to 3.

Step 2: Compute linearly-weighted summed image ζ in terms of Eq. (5).

$$\xi_i = \frac{\sum_{j \in N_i} S_{ij} x_j}{\sum_{j \in N_i} S_{ij}} \quad (5)$$

Where ξ_i denotes the gray value of the *i*th pixel of image ζ, x_j and N_i are as above.

Step 3: An efficient iterative algorithm for finding the optimal threshold[6] that minimize a weighted sum-of-squared-error objective function is used on the linearly-weighted summed image ζ, with which a binary image can be achieved.

Figure 2 Otsu method VS Our method

Figure 2 shows the results of text binarization by original global threshold [2] (OTSU) and binarization method incorporating local information. The binarization method incorporating local information demonstrate the advantages of noise-immunity (remove text-noise) and detail-preserving (good for removing background-noise) for image segmentation.

C. Noise Removal Based CCA

With the results of above binarization, we see that text and background can be separated well when the high contrast between text and background, and most text-noise are filtered. But as the complex background, some background-noises also are classified to text. The main reason is that the background-noises have the similar gray level with the text. That is also why the methods based threshold can not achieve good results. In our proposed method, an important advantage of segmentation based binarization incorporating information is that the text (text strokes) and the background have been divided into different connected components, which are mutually disjoint. As there are obvious difference

in position, density and geometry between the connected components of text and background, so connected component analysis based character stroke's density and shape feature can remove the background noises easily.
Some main rules which determine whether the connected component (CC) is background noise are as below.

(1)、The position of CC is on the edge of original image and the aspect ratio is larger than ε_1 or smaller than ε_2.

(2)、The position of CC is on the edge of original image, the width or height of CC is larger than 85% of total image's width or height and the density of CC is larger than 60% of the area of CC.

(3)、The geometric size of CC is smaller than a minimum value or the density of CC is smaller than a minimum value.

The corresponding result is shown in Figure.3

Figure 3. background noise removal based on CCA

In our experiments, the noise removal based on CCA acts on the step **3.2**' result image and its reverse image. The text polarity can be classified by filtered result image easily.

III. Experiment Results and Comparisons

To evaluate the effectiveness of our approach, contrast experiments are performed on the experimental image and real images. On the experimental image, we compare our proposed binarization method incorporating local information with the other popular binarization methods: global threshold [2] (OTSU) and adaptive threshold [3] (BST). The experimental image is labeled artificially and Guassian noise with different variance is performed on the experimental image. Here the performance of proposed method is evaluated according to the pixel extraction rate (PCR). It is defined as:

$$PCR = N_p / N \qquad (10)$$

Where N_p is the number of pixels which are classified correctly and N is the total number of pixels. The related experimental results are showed in Table 1.

Method \ Var / PCR	0.01	0.05	0.1
OTSU	91.36%	85.93%	78.71%
BST	94.56%	92.78%	90.06%
Our approach	98.28%	97.89%	97.11%

Table 1. Performance comparison of three algorithms

In the table 1, Var means the variance of Gussian noise which is added to the original gray image. From the results of experiments, three methods have the similar performance on image with simple background, and our method demonstrates better performance when the image with more noises and complex background.

We also experiment our approach on the real image dataset, collected from web images and TV news video frames which contain text regions. The embedded text has different sizes, and complex background. The characters in the dataset involve Chinese and Japanese. The details of dataset are tabulated in Table 2. Some of the samples are shown in Fig 4.

Fig 4. Some samples of the dataset

Language	Text regions	Characters
Chinese	200	1268
Japanese	100	479

Table 2. Composition of test dataset

In the real test dataset, we evaluate our approach according to the character extraction rate (CER) and the character recognition rate (CRR). They are defined as:

$$CER = N_e / N, CRR = N_r / N \qquad (11)$$

Where Ne is the number of characters completely extracted without obviously lost strokes or connected background, Nr is the number of character which are correctly recognized by an OCR engine, and N is the number of all the characters. The Chinese character recognition is carried out by commercial OCR software TH-OCR 9.0.

We compare the performance of our method with Lyu *et al's* approach [5], which use adaptive thresholding, dam point labeling and inward filling to segment the text regions. The related experimental results are summarized in Table 3.

	CER	CRR(only Chinese)
Our approach	92.4%	87.6%
Lyu' approach	91.5%	83.3%

Table 3. Performance comparison

The results show that our method has better performance than Lyu's method according to both CER and CRR. Since incorporating the local spatial and gray information, our approach is not sensitive to various types of noise and complex background, and be more effective to remove background noise based on CCA.

IV. CONCLUSION AND FUTURE WORK

In this paper, we proposed a text extraction method incorporating local information. First a binarization method incorporating local information is used to separate the text and background and remove, then a noise removal based CCA eliminates the noise in the text pixels. Experiments demonstrate the good performance of our method on the text extraction from the image with complex background.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China under Grant No. 60835001, No. 60933010 and No.60802055.

REFERENCES

[1] K. Jung, K. In Kim, and A. K. Jain, "Text information extraction in images and video: a survey," Pattern Recognition, vol. 37, pp. 977-997, 2004.

[2] "A Threshold Selection Method from Gray-Level Histograms," Systems, Man and Cybernetics, IEEE Transactions on, vol. 9, pp. 62-66, 1979.

[3] M. Seeger and C. Dance, "Binarising camera images for OCR," in Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on, 2001, pp. 54-58.

[4] L. Qifeng, J. Cheolkon, and M. Youngsu, "Text segmentation based on stroke filter," in Proceedings of the 14th annual ACM international conference on Multimedia Santa Barbara, CA, USA: ACM, 2006.

[5] M. R. Lyu, J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," IEEE Trans. Circuit and Systems for Video Technology, vol. 15, no. 2, pp. 243–255, Feb. 2005.

[6] L. Dong, G. Yu, P. Ogunbona, and W. Li, "An efficient iterative algorithm for image thresholding," *Pattern Recognition Letters,* vol. 29, pp. 1311-1316, 2008.

[7] Y. J. Song, K. C. Kim, Y. W. Choi, H. R. Byun, S. H. Kim, S. Y. Chi, D. K. Jang, and Y. K. Chung, "Text region extraction and text segmentation on camera-captured document style images," in Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on, 2005, pp. 172-176 Vol. 1.

[8] C. Thillou and B. Gosselin, "Combination of binarization and character segmentation using color information," in Signal Processing and Information Technology, 2004. Proceedings of the Fourth IEEE International Symposium on, 2004, pp. 107-110.

[9] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," Pattern Recognition, vol. 40, pp. 825-838, 2007.

[10] L. Szilagyi, Z. Benyo, S. M. Szilagyi, and H. S. Adam, "MR brain image segmentation using an enhanced fuzzy C-means algorithm," in Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE, 2003, pp. 724-726 Vol.1.