2011 International Conference on Document Analysis and Recognition

A Novel Method for Embedded Text Segmentation Based on Stroke and Color

Xiufei Wang*, Lei Huang*, and Changping Liu*
*Character Recognition Engineering Center ,
Institute of Automation, Chinese Academy of Science
95 Zhongguancun East Road, 100190, Beijing, China
Email: {xiufei.wang, lei.huang, changping.liu}@ia.ac.cn

***Abstract*—In this paper, a novel method for embedded text segmentation is proposed. The basic idea of our method is based on two properties of embedded texts: a) the color of text pixels is subject to gaussian distribution; b) the local part and the global part of embedded text shares the same color distribution. Inspired by this two characteristics, we develop a two-step text segmentation approach: in the coarse segmentation step, a 1-D gaussian function is adopted to model the color distribution of text pixels. To get the model parameters, a stroke operator is utilized to extract confident text region, and then a heuristic process is developed to estimate the parameters. The coarse segmentation can be carried out by the color model. In the noise elimination step, a color distribution homogeneity based method with connected component analysis is introduced. Preliminary experimental results show that our method performs well on complex background.**

***Keywords*-embedded text; text segment; stroke; color;**

I. Introduction

With the development of information technology, a quantity of multimedia comes forth, which leads to an urgent demand for content based browsing and retrieving system. Text embedded in images and videos always carries rich useful information, which can help the computer to understand the content of images and videos. A variety of text extraction approaches have been proposed and many applications have been investigated [6][7]. To extract the embedded texts, the text segmentation is a key step.

Compared with the segmentation of texts in scanned documents, the research of embedded texts segmentation is always challenged by low resolution and complex background. The height of text in images and videos is usually beyond 50 pixels, while the resolution of scanned documents is no less than 300 dpi. What's more, the background of embedded texts is more complex than the scanned documents.

Threshold-based methods were first developed to segment texts in scanned document images with simple background. The related thresholds are selected based on the intensity contrast between text and background. Otsu [1] presents an adaptive thresholding method through minimizing the intra-class variance, which has been widely used for text segmentation in scanned document images. After that, more threshold-based method such as local and global thresholding, Niblack's [2] method and etc. are proposed. Although these methods have been proved to be effective for document image segmentation, they may fail for embedded texts because of the low resolution and complex background.

Some authors have also considered the texture information of embedded text in their segmentation algorithms, such as color, edge, strokes, corner and etc. [5][9][10]. Fu et al.[3] first utilize the K-means algorithm to cluster a detected rectangle text block into K binary image layers, and then an effective post-processing procedure is applied to obtain more complete and accurate segmentation results by multi-constraints on color, edge and stroke thickness. Ye et al. [4] employ the "edge couple" characteristic of text to sample text pixels, and then a GMM is trained online to model the distribution of the hue and intensity values of text pixels. These methods may perform well when the text background is simple with less noise, but the selection of color layer number remains a big problem.

In this paper, a new method for embedded text segmentation is proposed. The basic idea of our method is based on two properties of embedded texts: a) the color of text pixels is subject to gaussian distribution; b) the local part and the global part of embedded text shares the same color distribution. Inspired by this two characteristics, we develop a two-step text segmentation approach: in the coarse segmentation step, a 1-D Gaussian function is adopted to model the color distribution of text pixels. To get the model parameters, a stroke operator is utilized to extract confident text region, and then a heuristic process is developed to estimate the parameters. The coarse segmentation can be carried out by the color model. In the noise elimination step, a color distribution homogeneity based method with connected component analysis is introduced. Preliminary experimental results show that our method performs well on complex background.

The rest of the paper is organized as follows: the color distribution characteristics of embedded texts and the algorithm overview are presented in Section 2; the details of our method are illustrated in Section 3; Section 4 are the related experiments and results; finally we draw our conclusions in Section 5.

II. Color distribution characteristics of embedded text

Text embedded in images and videos is a special kind of texture which contains high-level semantic information. It

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.39

Figure 1. Color distribution of embedded text: (a) images with embedded text, (b) labeled binary text images, (c) gray-level histogram of text pixels

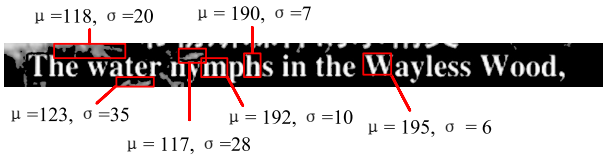

Figure 2. Color distribution homogeneity of embedded text

can be observed that the embedded texts show the following properties:

A. Gaussian model for the color of text pixels

The image with embedded texts can be regarded as a combination of text and non-text region. For an image I with embedded text, taken the text region as foreground and the non-text region background, I can be written as:

$$I(x) = B(x) + F(x) \tag{1}$$

Here, B and F are the background and the foreground. In this paper, the symbol $X(\cdot)$ refers to the pixel value of image X in the specified point. As the embedded text is added to express high semantic information, it is reasonable to hypothesis that the original text pixels have homogeneous color for better visual effects. However, during the image/video storing and transmission process, the image could be polluted by noises. We use a constant C to denote the color of original added text, F_N the transmission and storing noises. Thus the foreground F can be written as:

$$F(x) = C + F_N(x) \tag{2}$$

Let T be the embedded text of I, as T is regarded as the foreground, it can be written as:

$$T(x) = F(x) = C + F_N(x) \tag{3}$$

We hypothesize that the foreground noise obeys gaussian distribution as:

$$F_N \sim N(\mu, \sigma) \tag{4}$$

According to Euq.3, the embedded text T also obeys gaussian distribution:

$$T \sim N(\mu + C, \sigma) \tag{5}$$

Figure 1 shows some images with embedded texts. Figure 1(a)(b) are input images and the binary text images obtained by manually labeling. Figure 1(c) are the gray-level histogram of the text pixels. Considered the pixel level labeling error, it can be seen that the distribution of text pixels can be similarly modeled by a 1-D gaussian function as:

$$f(p(i)) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(p(i)-\mu)^2}{2\sigma^2}) \tag{6}$$

where $p(i)$ denotes the gray scale value of the i-th pixel in the image, and $f(p(i))$ is the probability of the i-th pixel to be in the text region. Based on this model, if we could obtain the gaussian parameters μ and σ, the text pixels can be then segmented from the background successfully.

B. Color distribution homogeneity of embedded text

Given an embedded text image, let T be the text region and $T_i \in T$ be a part of the text. As described above, we use 1-D gaussian function to model the color of text pixels, thus the color of text pixels in T and T_i should both obey the gaussian distribution as follows:

$$T \sim N(\mu, \sigma) \quad T_i \sim N(\mu_i, \sigma_i) \tag{7}$$

The color distribution homogeneity means that the local region T_i should have homogenous color distribution with its neighbors and the whole text region T.

In this paper, we use the global and local color distribution homogeneity to demonstrate this characteristic of the embedded text.

1) Global color distribution homogeneity (GCDH): GCDH means that T_i should have homogeneous color distribution with the whole text region T. We define global color homogeneous probability $P_g(i)$ to measure the color distribution similarity of T_i and T. Here $P_g(i)$ is calculated by:

$$P_g(i) = \frac{F_{\mu,\sigma}(\mu_i) + F_{\mu_i,\sigma_i}(\mu)}{2} \tag{8}$$

where $F_{\mu,\sigma}(\mu_i)$ and $F_{\mu_i,\sigma_i}(\mu)$ are calculated by:

$$F_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{9}$$

2) Local color distribution homogeneity (LCDH): LCDH means that T_i should have homogeneous color with its neighbors. Like $P_g(i)$ for GCDH, we define local color homogeneous probability $P_l(i)$ to measure the color distribution similarity of T_i and its neighbors. $P_l(i)$ is calculated by:

$$P_l(i) = \frac{F_{\mu_i,\sigma_i}(\mu_i^*) + F_{\mu_i^*,\sigma_i^*}(\mu_i)}{2} \tag{10}$$

Here $F_{\mu_i,\sigma_i}(\mu_i^*)$ and $F_{\mu_i^*,\sigma_i^*}(\mu_i)$ are calculated by Equ.9. μ_i^* and σ_i^* are the mean and standard variance of T_i's neighbors. μ_i^* and σ_i^* are calculated by:

$$\mu_i^* = \frac{1}{M}\sum_{j \in N_i} \mu_j \quad \sigma_i^* = \frac{1}{M}\sum_{j \in N_i} \sigma_j \tag{11}$$

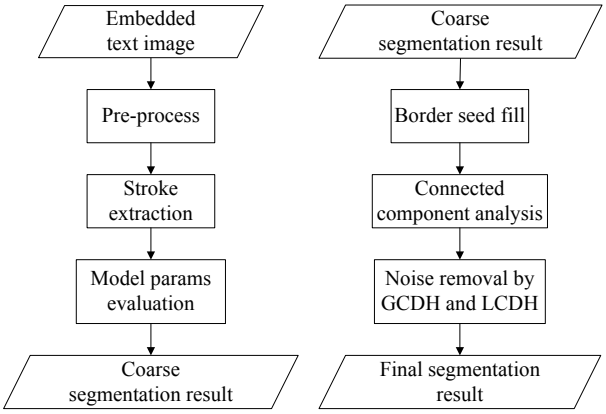

Figure 3. Flow chart of the proposed method: (a) coarse segmentation, (b) noise elimination

Here N_i refers to the neighbors of T_i and M is the number of selected neighbors. In this paper, we select 4 nearest neighbors of T_i to calculate μ_i^* and σ_i^*.

For an unknown region T_i, if it belongs to the text region, the $P_g(i)$ and $P_l(i)$ of T_i should satisfy:

$$P_g(i) > \varepsilon_0 \quad P_l(i) > \varepsilon_1 \tag{12}$$

Here, $\varepsilon_0 = 0.75$ and $\varepsilon_1 = 0.85$ in this paper.

An example is showed in Figure 2. It can be seen that the color of text regions tend to be similar while that of the text and non-text regions are quite different.

III. Proposed Method

Motivated by the analysis above, we propose a two-step method for the embedded text segmentation based on stroke and color. In the coarse segmentation step, a 1-D Gaussian function is adopted to model the color distribution of text pixels. To get the model parameters, a stroke operator is utilized to extract confident text region, and then a heuristic process is developed to estimate the parameters. The coarse segmentation can be carried out by the color model. In the noise elimination step, the color distribution homogeneity based constraints and the connected component analysis is introduced. The flow chart of the proposed method is showed in Figure 3.

A. Stroke-based coarse segmentation

According to the analysis in section II-A, the color distribution of text pixels in the embedded text can be modeled by a 1-D gaussian function. If we could obtain the gaussian parameters μ and σ, the text region can then be segmented from the background. Stroke is an important and useful feature for the text. Ye et al. [8] proposed a method to extract the stroke map of the input text image. In this paper, Ye's method is used to extract high-confident text regions, which are then utilized to evaluate the color model of the embedded text.

Table I
Parameters evaluation algorithm

Input: Gray image I, Stroke map S
Output: μ and σ

1. Get binary image B of S by OTSU Thresholding
2. Let i be the heuristic variant, $B^{(i)}$, $\mu^{(i)}$ and $\sigma^{(i)}$ be the results of i-th step.
(1) Initialize: $i = 1$, $B^{(0)} = B$, $\mu^{(0)} = \mu^{(1)} = 0$, $\sigma^{(0)} = \sigma^{(1)} = 0$.
(2) Let $\mu^{(i-1)} = \mu^{(i)}$, $\sigma^{(i-1)} = \sigma^{(i)}$, update $\mu^{(i)}$ and $\sigma^{(i)}$ by:
$$\mu^{(i)} = \frac{1}{N}\sum_{B^{(i-1)}(p)=255} I(p)$$
$$\sigma^{(i)} = \sqrt{\frac{1}{N}\sum_{B^{(i-1)}(p)=255} (I(p)-\mu^{(i)})^2}$$
(3) Update $B^{(i)}$:
$$B^{(i)}(p) = \begin{cases} 255 & \text{if } B^{(i-1)}(p) = 255 \text{ and } |I(p)-\mu^{(i)}| < \alpha\sigma^{(i)} \\ 0 & \text{otherwise} \end{cases}$$
Here $\alpha = 2.5$ in this paper.
(4). If $\mu^{(i-1)}$,$\sigma^{(i-1)}$,$\sigma^{(i)}$ and $\mu^{(i)}$ satisfy:
$|\mu^{(i)} - \mu^{(i-1)}| < t_0$ and $|\sigma^{(i)} - \sigma^{(i-1)}| < t_1$
Here t_0=0.001, t_1=0.1. Goto Step 3.
else:
$i = i + 1$, goto Step 2(2).
3. $\mu = \mu^{(i)}$, $\sigma = \sigma^{(i)}$, end

The flow chart of stroke-based coarse segmentation is showed in Fig.3(a). For an embedded text image T, the proposed coarse segmentation method is carried out as follows:

1) Pre-process: The pre-process aims at normalizing the size of the input image. For horizontal embedded text images, the height is normalized to 64 pixels; for vertical text images, the width is normalized to 64 pixels.

2) Stroke extraction: Calculate the stroke map of the normalized image. Let $f(\cdot)$ denote the gray scale value of the normalized input image and W the stroke width of the embedded text, the stroke map S is calculated as:

$$S = max_{d=0}^{3}\{S_d\} \tag{13}$$

where $d = 0, 1, 2, 3$ denote the four directions on 0, $\pi/4$, $\pi/2$, $3\pi/4$, and S_d is the stroke response on the direction d, which is calculated by:

$$S_d(p) = \begin{cases} S_d^*(p) & \text{if } S_d^*(p) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$S_d^*(p) = max_{i=1}^{W-1}\{\min\{f_d(p-i), f_d(p-i+W)\}\} - f(p) \tag{15}$$

Here $f_d(p-i)$ denotes the gray scale value of the pixel at a distance of i pixels from point p in the d-th direction.

3) Model parameters evaluation: By stroke extraction, the text regions are strengthened and the noises are depressed. So the stroke map can be utilized to evaluate the text color distribution parameters μ and σ. In this paper, a heuristic process is adopted to obtain more precise evaluation results. The proposed heuristic model parameters evaluation algorithm is listed in Tab.I.

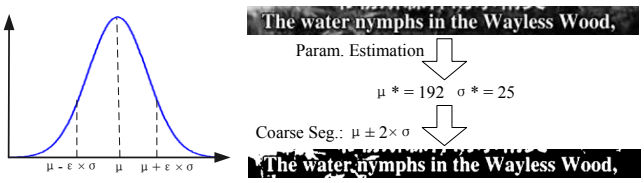

Figure 4. Coarse segmentation by gaussian model

4) Coarse segmentation: For the normalized input image I, once the model parameters μ and σ are obtained, the coarse segmentation can be carried out by:

$$B_c(p) = \begin{cases} 255 & \text{if } |I(p) - \mu| < \varepsilon\sigma \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Here B_c refers to the result image of coarse segmentation, and ε is the segmentation threshold. In this paper, $\varepsilon = 2.0$.

B. Color-based noise elimination

Most of the text pixels can be extracted by the stroke-based coarse segmentation. However, there also would be non-text noises in the coarse segment results, as the color of non-text noises might fall into the color band of text. Thus the noise elimination step is needed.

For the coarse segmentation result image B_c, the noise elimination process is carried out as follows:

1) Border seed filling: The method of border seed fill is proposed by Lienhart et al. [6]. This process aims at removing the non-text regions connected to the border.

2) Connected component analysis(CCA): After the process of border seed filling, the CCA is adopted to extract the connected components.

Let $C = \{C_0, C_1, \cdots, C_N\}$ denote the connected component (CC) set, for each component C_i, we calculate the local mean μ_i and standard variance σ_i as follows:

$$\mu_i = \frac{1}{N_i}\sum_{j \in C_i} p(j) \quad \sigma_i = \sqrt{\frac{1}{N_i}\sum_{j \in C_i}(p(j) - \mu_i)^2} \quad (17)$$

where j is the point in C_i and $p(j)$ denotes the gray value of the j-th pixel in the normalized image. N_i is the number of pixels in C_i.

3) Color ditribution homogeneity based noise elimination: Remove the CCs which don't satisfy color distribution homogeneity constraint. For the i-th CC C_i, it should be eliminated if C_i doesn't obey the conditions in Equation 12.

IV. Experiments and analysis

To evaluate the effectiveness of our method, we grab 1200 text blocks including 11789 characters from images and video frames. The characters in the dataset involve Chinese, English, and digits. All the experiments are done on the computer with a CPU of Pentium IV 2.8GHZ.

The proposed approach is evaluated by the recognition results and the process time cost per image. In this paper,

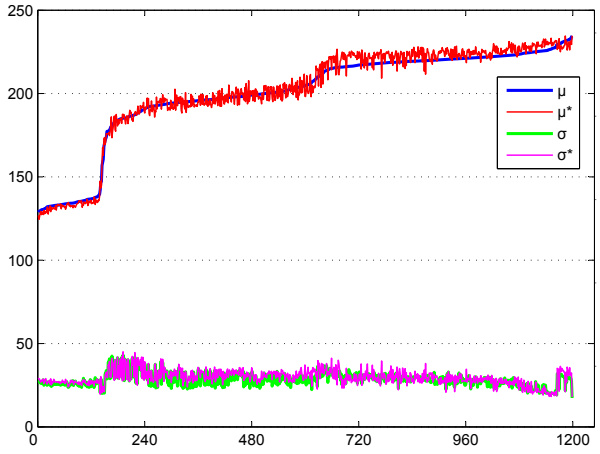

Figure 5. Parameters estimation results on the test dataset: μ and σ are calculated by labeled ground truth image; μ^* and σ^* are calculated by the stroke-based evaluation method.

Table II
The performance of the noise elimination process

Alg.	RPR	RRC	LRR
Coarse segmentation	71.45%	70.53%	39.97%
After noise elimination	97.27%	97.88%	85.22%

the recognition precision rate (RPR), recognition recall rate (RRC) and line recognition rate (LRR) are adopted as evaluation criteria. RPR, RRC and LRR are calculated as follows:

$$RPR = \frac{N_C}{N_R} \quad RRC = \frac{N_C}{N_G} \quad LRR = \frac{L_C}{L_G} \quad (18)$$

Here, N_R and N_C denote the number of totally recognized and correctly recognized characters. N_G is the ground truth number of characters. L_C is the number of correctly recognized text lines. L_G is the ground truth number of text lines. The text recognition in this paper is implemented by commercial OCR engine from Hanvon Tech., Co. Ltd..

In this paper, we utilize a heuristic method to evaluate the color distribution parameters μ and σ by the stroke map (see Table I). To prove the effectiveness of our approach, we compare the evaluated parameters with the ground truth value. The results are showed in Figure 6, where μ and σ are calculated by labeled ground truth binary image, μ^* and σ^* are obtained by the stroke-based evaluation method. It can be seen that the evaluated parameters match the ground truth well.

The proposed method includes a coarse segmentation and noise elimination process. To test the performance of this two process, we do comparison experiments on the test data. The results are showed in Table II. It can be seen that the performance is highly increased by the noise elimination process. The RPR and RRC have been increased by nearly

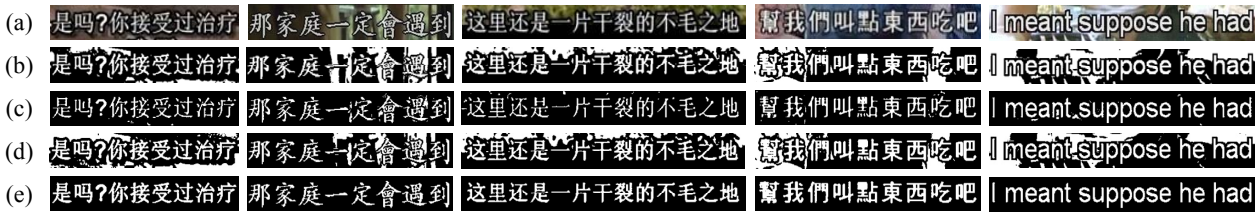

Figure 6. Segmentation results of different methods: (a) Input image, (b) Otsu's method, (c) K-Means, (d) coarse segmentation of the proposed method, (e) noise elimination of (d).

Table III
PERFORMANCE COMPARISON OF OUR METHOD WITH THE OTHER TWO ALGORITHMS

Alg.	RPR	RRC	LRR	Time cost(ms)
OTSU	69.56%	68.34%	33.87%	32
K-Means	90.25%	87.56%	75.36%	75
GMM	95.55%	94.45%	74.98%	228
Our method	97.27%	97.88%	85.22%	36

26%, and LRR about 46%, which proves the effectiveness of our noise elimination process.

To prove the advantage of our method, we compare the performance of our method three widely used text segmentation methods: Otsu's [1] adaptive thresholding method, K-Means based method and GMM based method. Here the initial number of color layer $K = 3$ for the K-Means based method and the initial number of gaussians $N = 3$ for the GMM based method. The experimental results are summarized in Table III. The time cost of each method is measured by the average processing time. The results show that our approach has better performance than the other approaches according to both recognition results and the time cost. Since our approach utilizes both stroke and color information of the embedded text, it is less sensitive to the complexity of background and can extract the text in images efficiently. Compared with the high computation complexity of K-Means and GMM based methods, the proposed method is faster and more effective.

V. CONCLUSION

We present a two-step text segmentation approach based on stroke and color in this paper. In the coarse segmentation step, a 1-D Gaussian function is adopted to model the color distribution of text pixels. To get the model parameters, a stroke operator is utilized to extract confident text region, and then a heuristic process is developed to estimate the parameters. The coarse segmentation can be carried out by the color model. In the noise elimination step, a color distribution homogeneity based method with connected component analysis is introduced. Experiments show that promising results have been achieved by the proposed method.

ACKNOWLEDGMENT

The work of this paper is supported by the National Natural Science Foundation of China under grant NO. 60933010.

REFERENCES

[1] N. Otsu, "A Threshold Selection Method From Gray-Level Histograms", IEEE Trans. Sys., Man., Cyber, vol. 9, Jan. 1979, pp. 62-66, doi:10.1109/TSMC.1979.4310076.

[2] W. Niblack, An Introduction to Digital Image Processing, Englewood Cliffs, NJ: Prentice Hall, 1986, pp. 115-116.

[3] L.B. Fu, W.Q. Wang, and Y.W. Zhan, "A Robust Text Segmentation Approach in Complex Background Based on Multiple Constraints", Lecture Notes in Computer Science (LNCS), 2005, vol. 3767/2005, pp. 594-605, doi:10.1007/11581772_52.

[4] Q.X. Ye, W. Gao, and Q.M. Huang, "Automatic text segmentation from complex background", Proc. 2004 International Conference on Image Processing (ICIP'04), vol.5, Oct. 2004, pp. 2905-2908, doi: 10.1109/ICIP.2004.1421720.

[5] X.J. Li, W.Q Wang, Q.M. Huang, W. Gao, and L.Y. Qing, "A hybrid text segmentation approach", Proc. 2009 IEEE International Conference on Multimedia and Expo (ICME'09), June 2009, pp. 510-513, doi: 978-1-4244-4291-1/09.

[6] Rainer Lienhart, "Video OCR : A Survey and Practitioner's Guide", In Video Mining, Kluwer Academic Publisher, Oct. 2003, pp. 155-184.

[7] K. Jung, K.I. Kim, A.K. Jain, "Text information extraction in images and video: a survey", Pattern Recognition, vol. 37, May 2004, pp. 977-997, doi:10.1016/j.patcog.2003.10.012.

[8] X. Ye, M.Cheriet, C.Y. Suen, "Stroke-model-based character extraction from gray-level document images", IEEE Transaction Image Processing, vol. 10, Aug. 2001, pp. 1152-1161, doi: 10.1109/83.935031.

[9] D.T. Chen, J.M. Odobez, H. Bourlard, "Text detection and recognition in images and video frames", Pattern Recognition, vol. 37, March 2004, pp. 595-608, doi:10.1016/j.patcog.2003.06.001.

[10] R. Lienhart, A. Wernicke, "Localizing and segmenting text in images and videos", IEEE Transaction on Circuits and System for Video Technology, vol. 12, April 2002, pp. 256-268, doi:10.1109/76.999203.