Automatic Text Detection Using Multi-Layer Color Quantization in Complex Color Images

Soo-Chang Pei, and Yu-Ting Chuang

Department of Electrical Engineering, National Taiwan University
No.1, Sec. 4 Roosevelt Rd., Taipei, 10617 Taiwan, R.O.C
TEL: 886-2-23635251 ext. 321 FAX: 886-2-23671909
Email: pei@cc.ee.ntu.edu.tw, r90942058@ms90.ntu.edu.tw

ABSTRACT

News, magazines, Web pages, etc in this modern life always contain much text information. In this paper, we propose a novel approach to detect text in images with very low false alarm rate. First of all, neural network color quantization is used to compact text color. Second, 3D histogram analysis chooses several color candidates, and then extracted each of these color candidates to obtain several bi-level images. Moreover, for each bi-level image, connectivity analysis and some morphological operators are fed to produce character candidates. Furthermore, we calculate some spatial features and relationships of each text candidate. At last, we can localize text regions by authentication from L.O.G edge detector. Meanwhile, in complex color images, multi-quantization layers can be integrated to reject non-text parts and reduce false alarm rate.

1. INTRODUCTION

In modern multimedia times, News, Web pages, magazines, and advertisements are everywhere in our lives. Among them, text absolutely, is the most important information. For example, when people surf Web pages, they always care about scores in a baseball game, or the price and name of products they like. Therefore, text detection is becoming a popular research nowadays. In related works, Jain and Yu [2,8] use color reduction to decompose an input image to several individual foreground images and then put them to connected component analysis to localize text regions. This approach has two drawbacks. First, in the low contrast color image, false alarm rate might increase rapidly due to color quantization. Second, as number of quantized color increases, the system has to pay higher computing complexity or more memory space. Lienhart and Wernicke [3] decimate the input image to multiple resolution layers. By utilizing edge features in each individual layer, text can be located after integrating all resolution layers. But in general compound images and video score bar always contain many characters with small font size so that characters after decimation are almost invisible. Gao, and Yang [5], Cai, Song, and Lyu [4] suppose that edge strength and density of characters are always stronger than other objects in color images. Therefore, after edge filtering, text candidates could be easily found. Unfortunately, systems with above assumptions could never work very well in complex color images. Zhong, Karu and Jain [7] compute the spatial variance along each horizontal line over the whole image and text lines can then be found by extracting the rows between two sharp edges of the spatial variance – one edge rising and the other falling. In their approach, if the background is complex, an appropriate threshold could not easily be identified. By our approach, we could get fewer foreground images by means of 3D histogram analysis and raise detecting rate by integrating all single quantization layers. In addition, we use two morphological operators to compensate text fractions resulted from color quantization.

Remainders of this paper are organized as follows. Section.2 describe details of our algorithm. Section.3 shows our performance of this algorithm and some experiment results. Final conclusion is made in Section.4

2. ALGORITHM

The fundamental building block of our whole text detection system is illustrated in Fig.1. First, the input image is quantized to several quantized images with different number of quantized color. For each quantized image, it was put to 3D histogram analysis to find some specific colors, which are probable text candidates. Furthermore, each bi-level image relative to its color candidate could be produced. By calculating some spatial features and relationships of characters, text candidates would be identified. Final, we combine all single quantization layers so that we could localize text regions accurately. In the following sections, we will first explain details of single quantization layer, and than multi-layer combination approach will be described later.

2.1 Text Contents of Single Quantization Layer

Before explaining steps of single quantization layer, we have to first make some assumptions for general text in color images as follows.

a. Text within an image is meaningful if and only if people can recognize easily.
b. For the same sentence or word, character size is similar to each other.
c. The words need to be big enough for human eyes to recognize.
d. Characters of the same word or sentence within an image always have similar colors.

0-7803-8603-5/04/$20.00 ©2004 IEEE

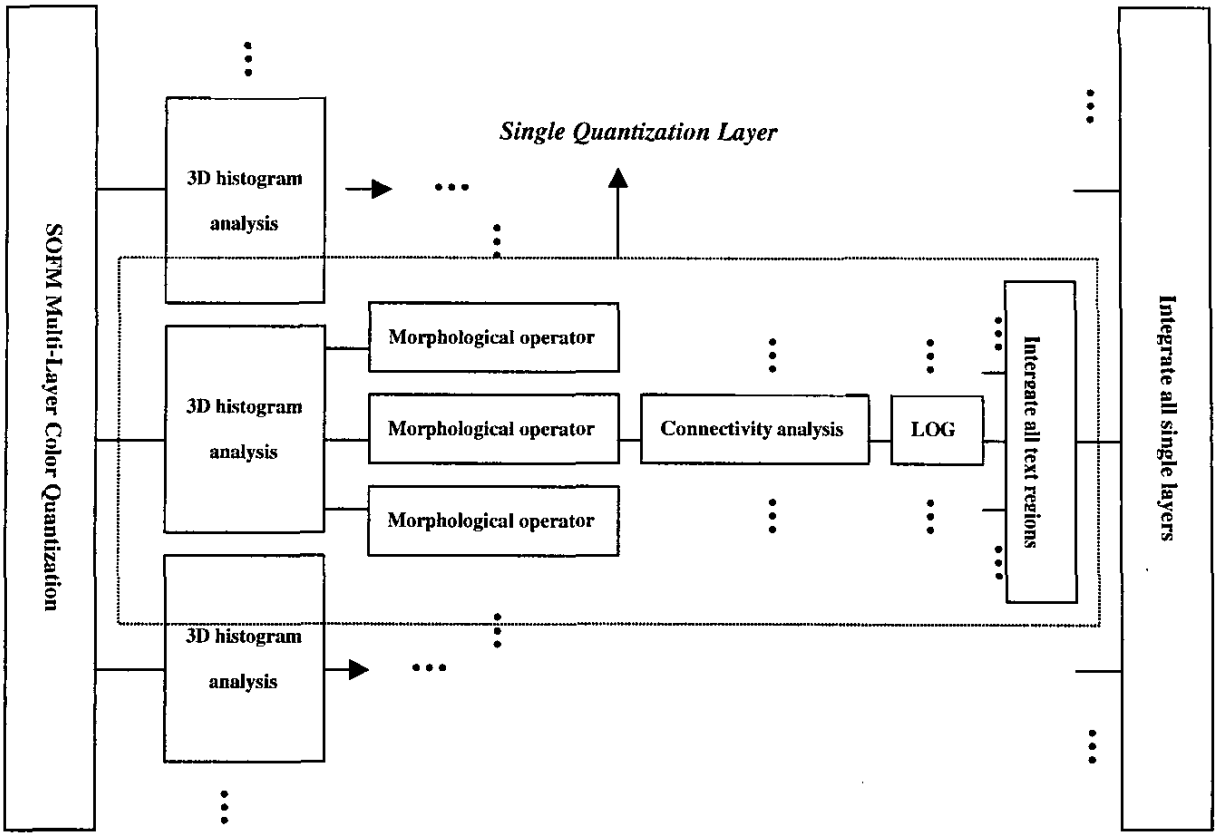

Fig. 1. Building blocks of whole text detection system

2.2 Color Quantization

It is essential that text regions have to be merged together first if we would like to localize it by using color information. In accordance with above points, we choose SOFM neural network [1] to quantize input image. First of all, we select an appropriate neural structure and small color table size (4 to 10 colors were recommended), then assigned uniformly distributed color to initial neurons. Final, Butterfly-Jumping sample sequence from original image was fed in SOFM training to obtain the color palette.

2.3 3D Histogram Analysis

Generally speaking, pixels in text region are often more compact than in other objects. According to this point, we calculate the 3D color histogram as shown in Equation (1), however, with respect to each quantized color $\bar{v}_q$, *where* $\bar{v}_q \in \{\bar{v}_{q1}, \bar{v}_{q2}, \ldots \bar{v}_{qm}\}$, *and m=number of quantized color*, we estimate its histogram gradient $E(\bar{v}_q)$ by equation (2). Thus, some quantized colors whose gradient is higher than a threshold would be considered as textual color candidates. After determining dominant colors, several non-important colors could be rejected to reduce computation complexity.

$$H(\bar{v}) = \#\{(r,c) \mid I(r,c) = \bar{v}\} \quad (1)$$

where $\bar{v} \in (r,g,b)$

$$E(\bar{v}_q) = \frac{1}{125} \sum_{i=-2}^{2} \sum_{j=-2}^{2} \sum_{k=-2}^{2} \left| H(\bar{v}_q) - H(\bar{v}_q + \bar{v}_{bias}) \right| \quad (2)$$

where $\bar{v}_{bias} = (i,j,k)$

2.4 Morphological Operating

Substantially, English character consists of only one connected region (except "i", "j"), but in other languages, a character always includes two or three regions (see Fig.2.a). Thus, if we put this kind of "non-single region" characters to connected component analysis without doing any preprocessing, it is more likely that connected component analysis might make the wrong decision. So, we have to merge these regions first. Here, we utilize morphological dilation [10] to merge the "co-character" regions (see Fig.2.b). Unfortunately, a serious problem might be followed by this operation, if two characters are very close to each other, they might also be merged together due to this operation. Consequently, morphological erosion [10] with different structuring element such as bar shape must be used to solve this problem (Fig2.c). Also, it is exciting that Morphological dilation can also compensate the effect that when color quantization works on low contrast images, a character sometimes might be

broken to pieces, an example is shown in Fig2.d and Fig2.e, where the left character of Fig2.d is broken to four pieces and it is compensated in Fig2.e. In parentheses, it is known that using both dilation and erosion will result in granularities in images, but in our application, they are always discarded, because granularities are always resulted from very small and independent connected region, i.e., a small connected region is difficult for human eyes to recognize it even if it is actually a character.

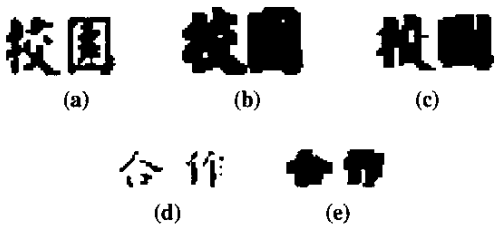

Fig. 2. (a) Original characters, (b) dilation on (a), (c) erosion on (b), (d) color quantization in low contrast image, (e) dilation on (d)

2.5 Connectivity Analysis

Details of this step are similar to other existing text detection methods. In general, characters in an image always appear in groups; therefore, using some features such as width, height and distance can identify text candidates.

2.6 Authentication from L.O.G Edge Filter

Instead of adopting edge filter to find text candidates, we make use of L.O.G (Laplacian of Gaussian) edge filter to only confirm each text candidate. By calculating the ratio of edge and non-edge points, we could make sure whether it is text region or not, if the ratio of this bounding box is higher than a threshold.

2.7 Multi-Layer Combination

For simple background images, single quantization layer may work very well. But in complex background images single quantization layer may fail to detect the accurate text regions, meanwhile, it may also produce many false boxes at the same time. Fortunately, in different single layers although many false boxes might happen, they would neither appear in similar location nor have same box size. Therefore, we could solve this serious problem by integrating several different single quantization layers that have different false boxes as shown in Fig.3. It is clear that if we hold boxes that always not only appear in the fixed location but also have similar box size, and reject the other boxes, we could detect real text boxes with low false alarm rate in complex background images.

3. EXPERIMENT RESULTS

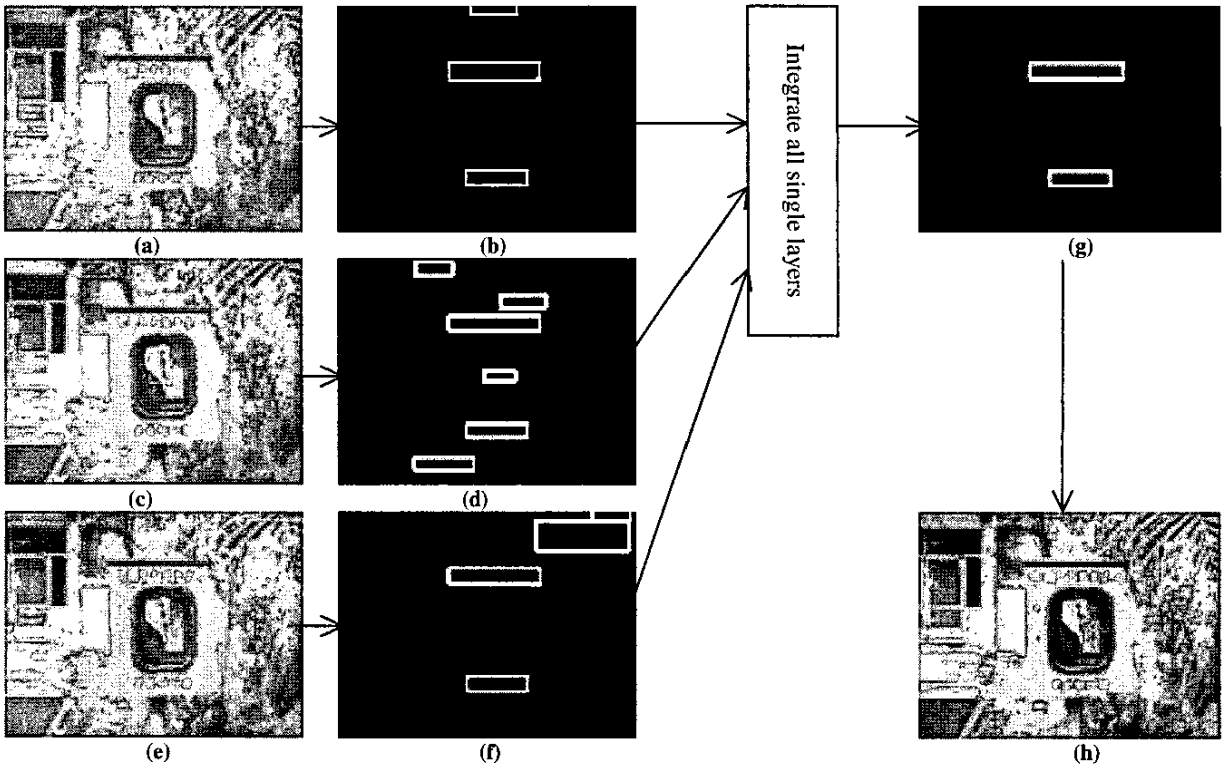

Fig. 3. (a, b) Single layer with CQ = 6 and its own text boxes, (c, d) Single layer with CQ = 8 and its own text boxes, (e, f) Single layer with CQ = 10 and its own text boxes, (g) output text boxes after integrating all single layers, (h) output image (where CQ is the number of quantized color)

To prove the robustness of our system, we test some color images, and three video sequences included sports and news *without adjusting any parameter of our system* by changing conditions such as resolution, font size, languages, and complexity of background. The overall performance of the algorithm is listed in Table 1, and some test image and video frames are shown in Fig. 4.

As listed in Table 1.we define the hit rate, false alarm rate, and miss rate to evaluate our system as follows:

$$\textit{hit rate} = \frac{\#\{\text{boxes detected as text and they are indeed text}\}}{\#\{\text{boxes detected as text}\}} \times 100$$

$$\textit{miss rate} = 100 - \textit{hit rate}$$

$$\textit{false alarm rate} = \frac{\#\{\text{boxes detected as text but they are not text}\}}{\#\{\text{boxes detected as text}\}} \times 100$$

where the symbol " # " represents the number of the set.

Total boxes in database	Number of detected boxes	Hit Rate	False alarm rate	Miss rate
518	455(3 wrongs)	87.26%	2.38%	17.95%

Table.1.

4. CONCLUSION

In our algorithm, instead of concentrating on choosing some "tied" parameters to avoid false text localization, we use a multiple color quantization layer approach to localize correct text with very loose parameters. In addition, several morphological operations are used to compensate some shortcomings. For detection performance, we indeed get a low false alarm rate, and high hit rate.

5.REFERENCES

[1] S. C. Pei, and Y. S. Lo, "Color Image Compression and Limited Display Using Self-Organization Kohonen Map", *IEEE Trans. Circuits and Systems for Video Technology, pp.191-205, Apr. 1998.*

[2] Anil K. Jain, and Bin Yu "Automatic Text Location in Images and Video Frames", *IEEE, Intl. Conf. Pattern Recognition, pp.1497-1499, Aug. 1998*

[3] R.Lienhart, and A. Wernicke "Localizing and Segmenting Text in Images and Videos", *IEEE Trans. Circuits and Systems for Video Technology, pp.256-68, Apr. 2002.*

[4] M. Cai, J. Song, and M. R. Lyu, "A New Approach for Video Text Detection", *IEEE, Intl. Conf. Image Processing, pp.117-120, 2002.*

[5] J. Gao, and J. Yang, "An Adaptive Algorithm for Text Detection from Natural Scenes", *Proceedings of Computer Vision and Pattern Recognition (CVPR), pp.84-89, 2001.*

[6] J. Yang, X. Chen, J. Zhang, Y. Zhang, and A. Waibel, "Automatic Detection and Translation of Text from Natural Scenes", *IEEE, Intl. Conf. Acoustics, Speech, and Signal Processing (ICASSP), pp.2101-2104, May, 2002.*

[7] Y. Zhong, K. Karu, and A. K. Jain, "Locating Text in Complex Color Images", *Pattern Recognition, 28:1523-1535, pp.146-149, 1995.*

[8] A. K. Jain, and Bin Yu, "Automatic text location in images and video frames", *Pattern Recognition, vol. 31, no. 12, pp.2055-2076, 1998.*

[9] S. Prabhakar, H. Cheng, John C. Handley, Z. Fan, and Y. W. Lin, "Picture-Graphics Color Image Classification", *IEEE, Intl. Conf. on Image Processing (ICIP), pp.785-788, 2002*

[10] Robert M. Haralick, and Linda G. Shapiro, *"Computer and Robot Vision" vol 1*

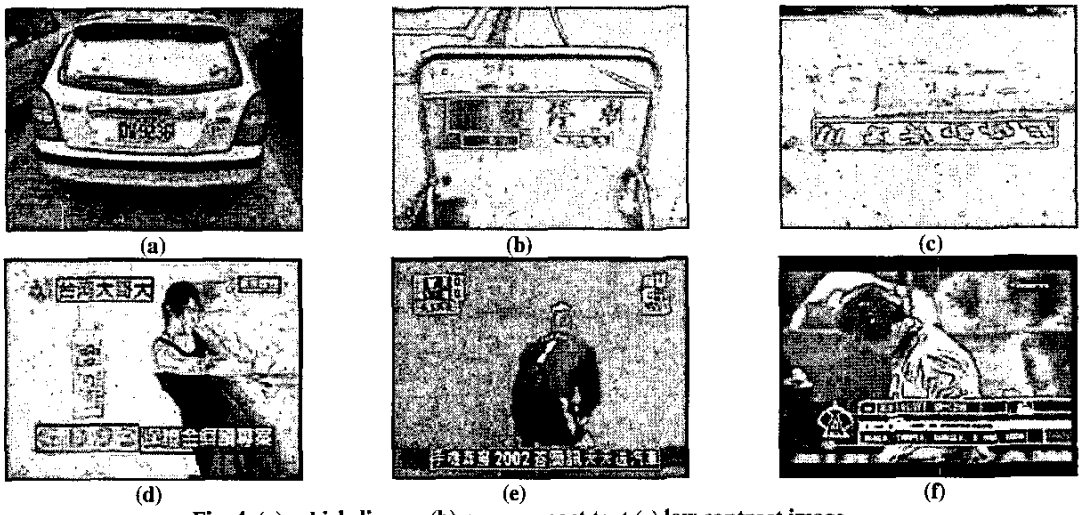

Fig. 4. (a) vehicle license (b) non-compact text (c) low contrast image (d) TV advertisements (e) soccer (f) baseball