行政院國家科學委員會專題研究計畫　期中進度報告

子計畫一:視訊的智慧型高階處理(2/3)

計畫類別： 整合型計畫
計畫編號： NSC92-2219-E-002-016-
執行期間： 92 年 08 月 01 日至 93 年 07 月 31 日
執行單位： 國立臺灣大學電機工程學系暨研究所

計畫主持人： 貝蘇章

報告類型： 精簡報告
報告附件： 出席國際會議研究心得報告及發表論文
處理方式： 本計畫可公開查詢

中　華　民　國 93 年 5 月 4 日

智慧型音視訊和傳輸技術及多媒體應用-子計畫一：視訊的智慧型高階處理(II)
Intelligent High level Video Processing (II)

計畫編號：NSC-92-2219-E-002-016
執行期限：92 年 8 月 1 日至 93 年 7 月 31 日
主持人：貝蘇章　台灣大學電機系教授

摘要

本論文為了達到完美的語言翻譯，我們提出了一個新的文字偵測技術，可達到一個極低的假警報率。首先，以類神經網路的彩色量化法使得顏色類似的文字可被量化成相同的顏色，接著我們使用三維的統計長條圖分析法來選擇幾個可能的文字候選色，如此，對於每個可能的文字候選色我們可以分別萃取出它們的相對雙色調圖，之後我們使用相連物件分析法與兩個型態學的運算子於每一張雙色調圖來找出可能的文字區域，最後，我們使用高斯的拉普拉邊緣檢測器來對可能的文字區域作更進一步的確認，同時，多層量化的技術可以讓我們大大的降低假警報率。
關鍵字: 文字偵測、色彩量化、彩色影像、類神經網路

ABSTRACT

In order to achieve good translating performance, we propose a novel approach to detect text in color 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 colors candidates, and then extracted each of these color candidates to obtain several bi-level images. For each extracted bi-level image, connected component analysis and several morphological operators are fed to hold some boxes that are possible text regions. At last, we can use L.O.G edge detector to authenticate accurate text regions from each possible text regions. Meanwhile, in complex color images, multiquantization layers can be integrated to reject non-text parts and reduce false alarm rate.
Keyword: Text detection、color quantization、color images、neural network

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. Characters of the same word or sentence within an image always have similar colors.

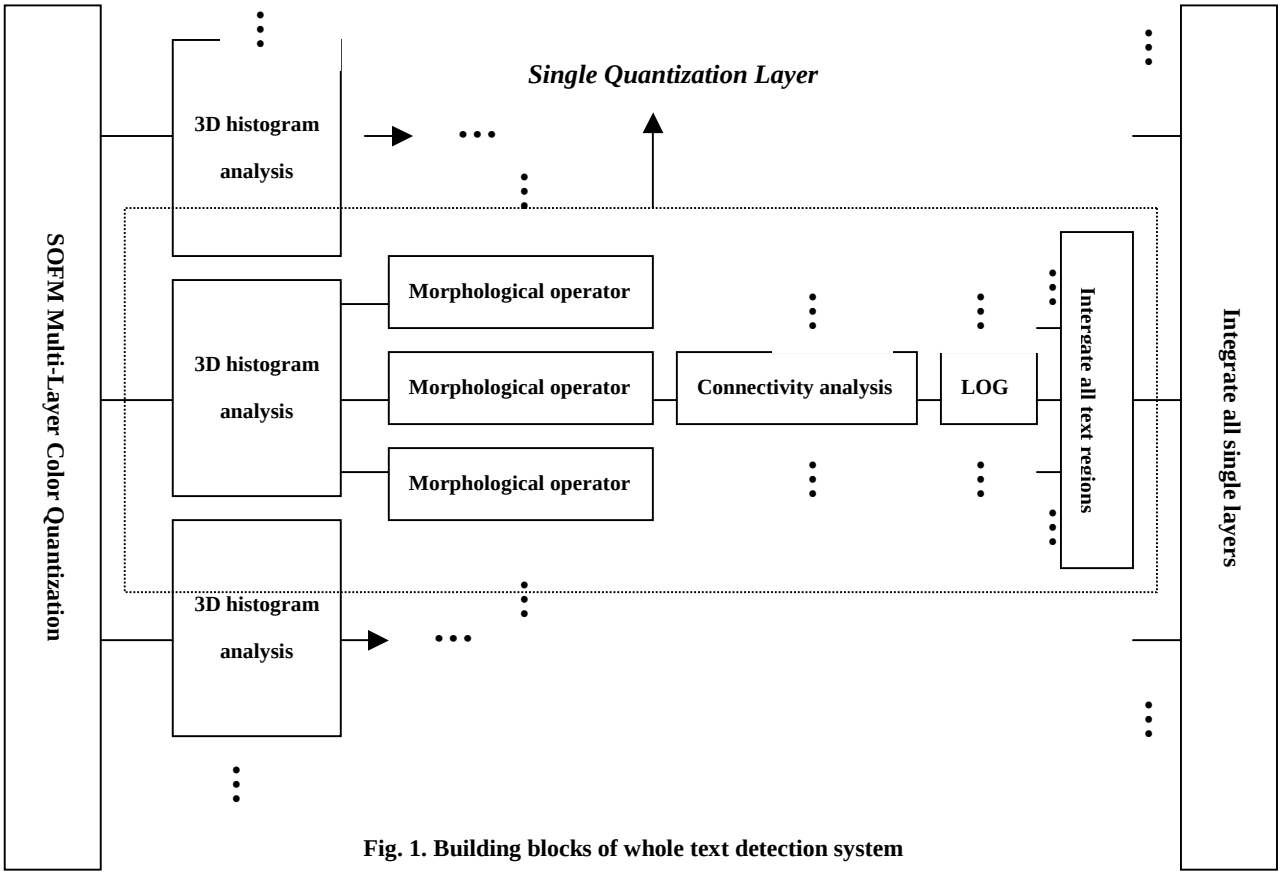


Fig. 1. Building blocks of whole text detection system

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 $\vec{v}_q$, *where* $\vec{v}_q \in \{\vec{v}_{q1}, \vec{v}_{q2}, \ldots \vec{v}_{qm}\}$, *and m=number of quantized color*, we estimate its histogram gradient *E(*$\vec{v}_q$*)* 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.

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

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

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

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

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 and lead to false localization in output image. In order to solve this problem, we utilize morphological dilation to merge these “co-character” regions (see Fig.2.b) so that it can work well in connected component analysis. Unfortunately, a serious problem might be followed by this operation, if two characters are very close to each other, this two characters might also be merged together due to this operation. Consequently, morphological erosion with different structuring element such as bar shape must be used to solve this problem (Fig2.c). In addition, when color quantization works on low contrast images, characters sometimes might be divided to several fractions so that even English characters would be broken also. Morphological dilation can also compensate these effects owing to color quantization in low contrast images. An example of this condition is shown in Fig.2.d.and Fig.2.e.

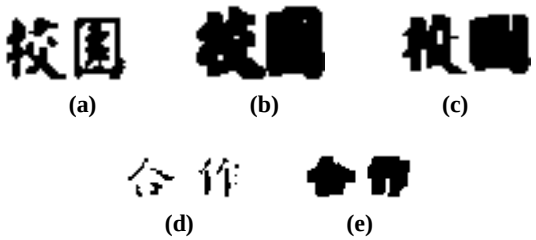


Fig. 2. (a) Original characters, (b) dilation on (a), (c) erosion on (b), (d) color quantization in low contrast image, (e) dilation on (d)

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 (here we set the threshold 0.28).

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

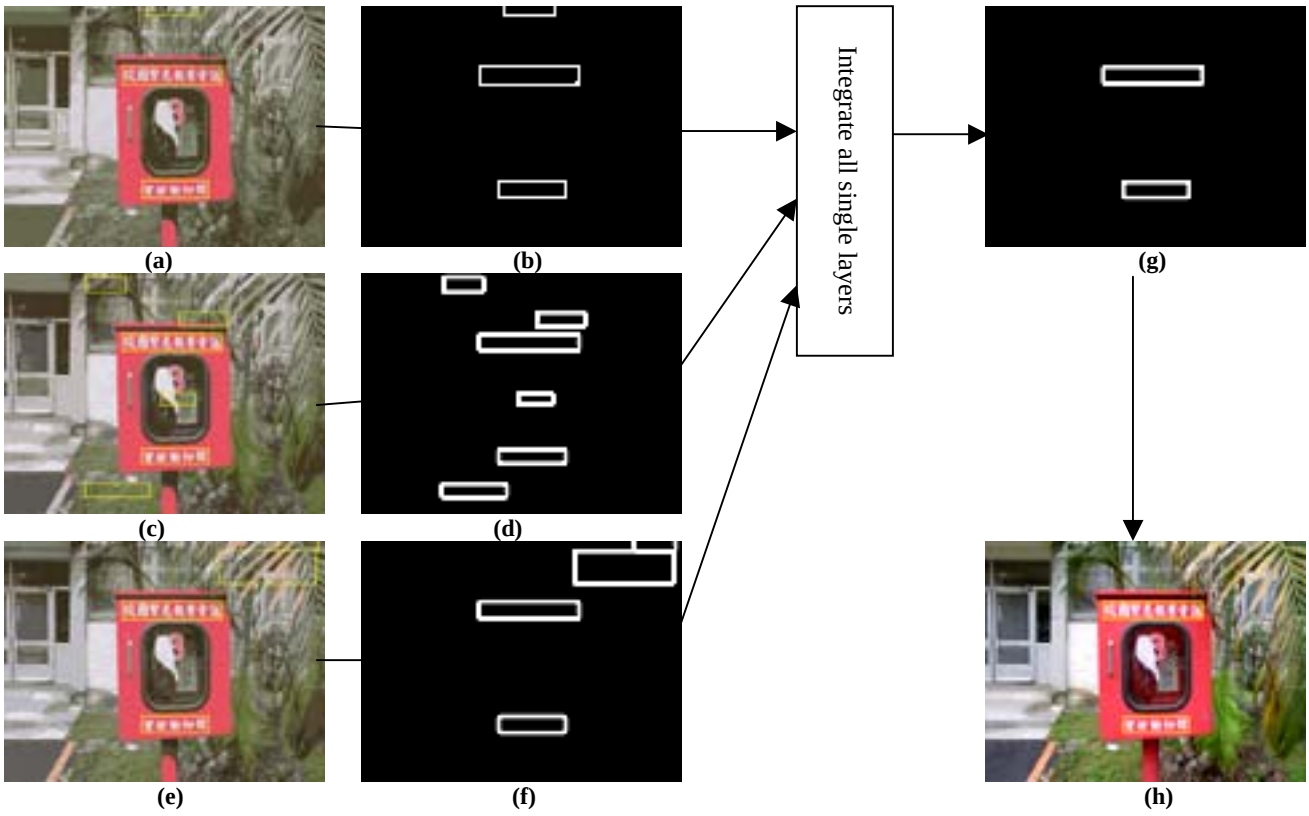


Fig. 3. (a, b) Single layer with CQ = 6 and its own text boxes, (c, d) Single layer with CQ = 8 and its own text boxes, (e, f) Single layer with CQ = 10 and its own text boxes, (g) output text boxes after integrating all single layers, (h) output image (where CQ is the number of quantized color)

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:

hit rate =

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

miss rate = 100 — *hit rate*

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.*

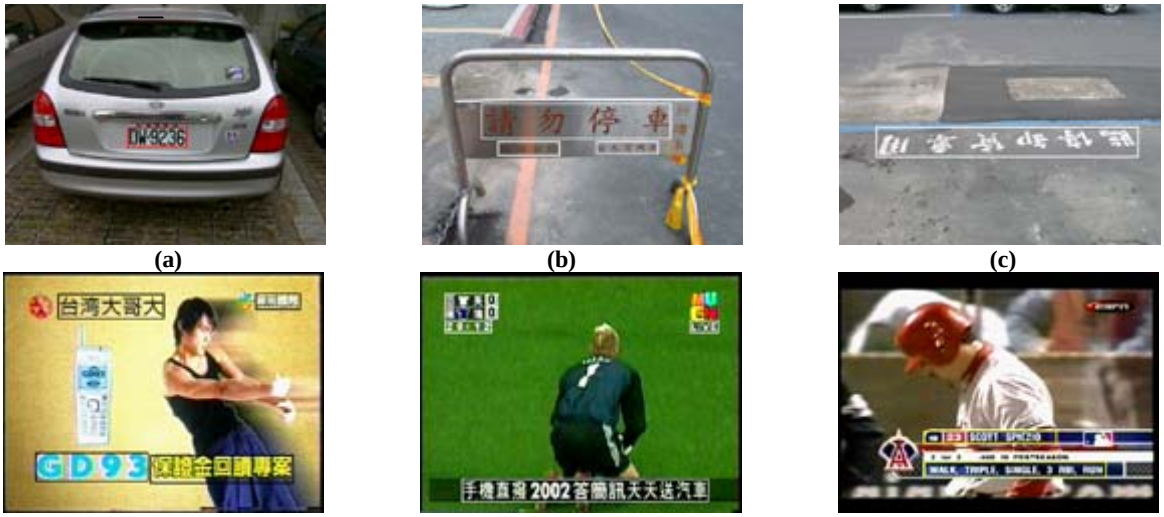
[5] J. Gao, and J. Yang, "An Adaptive Algorithm for Text Detection from Natural Scenes", *Proceedings of Computer Vision and Pattern Recognition (CVPR), pp.84-89, 2001.*

[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*



(a) (b) (c)

(d) (e) (f)

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