MVA'98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

8—13

Orientation and Scale Invariant Text Region Extraction in WWW Images

Taehoon Park* Dongsung Kim Kyusik Chung
School of Electronic Engineering
Soongsil University

Abstract

Text extraction from a web image is important for web indexing because the text can contain a key information of the web. This paper presents a method to detect a text with various orientation and multi-font sizes in a web image. The proposed method consists of three steps; 1) color reduction of low resolution web image by a spatial merging algorithm, 2) extraction of character candidates by finding connected components corresponding to strokes, and 3) detection of texts in an arbitrary direction by filtering and combining character candidates based on a potential field. This method allows to detect text orientation and scale invariantly. Experiments with 200 web images are performed to verify the validity of the proposed method.

1 Introduction

The number of web pages has been increased so rapidly[1] that effective indexing and efficient retrieval of them become important issues. Most recent web pages contain images with graphical texts for better visual appearances and rich information content. Since such graphical texts can represent keywords of web pages, their extraction is essential for web indexing.

Detection of a text region from a web image is difficult due to the following reasons. Firstly, a text region can appear on a complex background and/or have a color similar with a background color. This makes segmentation of a text region very difficult. Secondly, a web image can have very large number of colors, for instance, a JPEG image can represent up to 16 million colors. This makes color reduction required for efficient processing. Thirdly, the text region can have various alignments and character sizes within the region. This makes extraction of strings more difficult.

Previous research on text detection from color images can be divided into two based on the number of images used: detection from sequence of images and from a single image. The former can utilize interframe information to detect texts from TV news or movie[2-6]. For example, a TV news caption is static along several image sequences while other backgrounds change. The latter can be further classified into two based on resolutions of images: scanned images(high resolution) and WWW images(low resolution).

Jain[2] and Zhong *et al.*[7] proposed methods to locate text in CD cover images. Huang *et al.*[8] proposed a method by grouping colors into clusters for foreground/background segmentation of color images. Doermann*et al.*[9] developed a method for extracting text from logs and trademark images. Zhou *et al.*[10] suggested a method to reduce the number of colors globally and extract text regions by the shape of connected components. A connected component can be considered as a character depending on their shape parameters such as size, width, number of holes, and branches of the component. Jain *et al.*[2] proposed a method to reduce the number of colors by a clustering algorithm and extract text regions using projection profiles of color components. Those methods proposed so far have limitations in that they do not deal with extraction of a text region whose font size is changed inside the region, and whose alignment is in an arbitrary direction. In order to overcome these limitations, this paper proposes a new method to detect text region orientation and scale invariantly.

The proposed method consists of three steps; 1) color reduction by a spatial merging algorithm, 2) extraction of character candidates by finding connected components corresponding to strokes, and 3) detection of texts in an arbitrary direction by filtering and combining character candidates based on a potential field.

2 Color Reduction

Most web images can represent very large number of colors, for instance, a JPEG image can represent up to 16 million colors. The large number of colors should be reduced for efficient text extraction. Conventional color reduction methods for high resolution images are not directly applicable to low resolution web images which has anti-aliasing filtering effects. The filtering is performed for comfortable discrimination of web image contents. Figure 1

*Address: 1, Sangdo-Dong, Dongjak-Ku, Seoul, Korea E-mail: spondge@q.soongsil.ac.kr

shows an example of a web image where anti-aliasing filtering effects are shown on the boundaries of its zoomed characters.

As seen in Figure 1, a stroke consists of a major component and its surround minor components. These minor components may be lost by simple bit dropping or global thresholding methods because color differences between a major component and minor components can vary for each character In order to conglomerate such minor components, a color reduction method using spatial merging is proposed. The method reduces number of colors in two steps: color quantization and spatial merging.

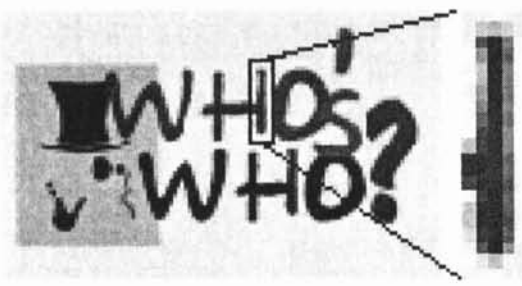

Figure 1: Original web image and anti-aliasing's distortion

The number of color is initially reduced by the RGB color quantization. The quantization equation is given in Equation 1. Results of color quantization for Figure 1 are shown in Figure 2.

$$C_i = (R_i, G_i, B_i) \rightarrow C_q = (\left|\frac{R_i}{64}\right|, \left|\frac{G_i}{64}\right|, \left|\frac{B_i}{64}\right|) \quad (1)$$

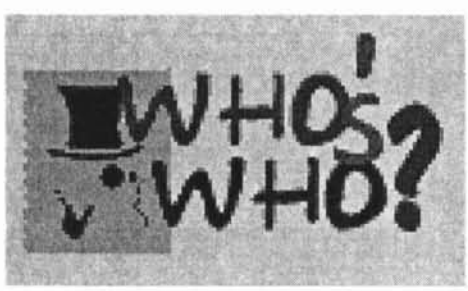

Figure 2: Quantized image

Next, spatial merging is performed. Once the number of colors is reduced, connected components with the same color are found. Then they are divided into major components and minor components based on their population. A component whose population is greater than α, where $\alpha = 7$, is categorized as a major one. For each major component, adjacent minor components are merged if color difference between them is within a threshold. While merging, major components whose population is more than $\beta\%$ of total number of pixels consisting the image are categorized as backgrounds and then removed. The β is set as 10 for the experiments. The results of color reduction are given in Figure 3.

Figure 3: Color reduction image

3 Character Extraction

Several characteristics of a character may be exploited for character extractions. Some of them are language-dependent and the others are language-independent. For our cases to extract texts with multiple languages, language-indepenendent characteristics need to be exploited during character extraction. One of them is that a character is composed of a set of strokes.

In stead of detecting a character directly, strokes are detected using their characteristics. The current implementation uses two characteristics: A stroke has a relatively fixed width along its skeleton and the ratio between width and length is within some ranges.

To detect strokes, connected components are found from a color reduced image. For each connected component, its width and length are computed. The computation is performed in a thinning process. Initially, an iteration count is set as 1. Then boundaries of the stroke are peeled off at every iteration if there exist inside pixels. The peeled pixels are marked with the iteration count. The iteration count is increased by one after each iteration. This process is repeated until no more inside pixel exists. Once the thinning is finished, a resulting skeleton contains information on width and length of the stroke. The number of the pixels of the skeleton becomes the length of the stroke, and the number marked at a skeleton pixel becomes the width of the stroke at that location. Average width is computed by averaging the widths of all skeleton pixels.

Among the detected strokes, there may exist false ones. Those can be removed by potential field text region detection. A stroke not constituting a text region is removed.

4 Potential Field Text Region Detection

A string can be defined as characters linked each other. This definition does not restrict orientation of a string, or sizes of characters in a string.

The definition can be implemented with a potential filed approach. In the potential field approach, a stroke is considered as an object having a potential

filed that is proportional to its size. A potential field for a character is computed with Equation 2.

$$PF(r) = \frac{r \times Character\ size}{Average\ width \times Length} \quad (2)$$

where r is distance from the center of the character.

Two adjacent potential fields can be merged into new potential field if their attraction forces are larger than a threshold. The potential field of new one is a sum of the two potential fields.

In merging adjacent potential fields, direction alignment is examined if it has a convincible structure. A text string has a consistent direction, so connection angles between adjacent characters should not change abruptly compared with other elements consisting of the string. Thus, characters giving convincible angles are selected as adjacent characters. Results of text region extraction for Figure 1 are given in Figure 4.

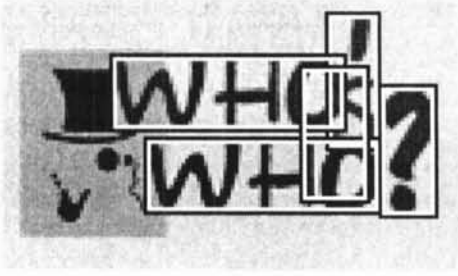

Figure 4: Text region extracted

5 Results

The proposed method is implemented in Visual C++ on Pentium PC, and tested on about 200 web images collected from internet web pages. Average detection rate is 88.8%. The details are summarized in Table 1.

Table 1: Results of experiments

Image	No. of Image	Accuracy(%)
Commercial	150	88.8
Non-commercial	50	92

Figure 5 shows results for a text with cursive alignment direction, where upper left, upper right, lower left, and lower right figures indicate original image, quantized image, character candidates, and detected text regions, respectively. Note that our method detects the whole cursive text.

Figure 6 shows results for a cursive text on complex background, where original image, quantized image, character candidates, and extracted text regions are shown in the same sequence as in Figure 5.

Figure 7 shows results for a text with a special visual effect in strokes of text "OPEN." The proposed

Figure 5: Experiment results 1

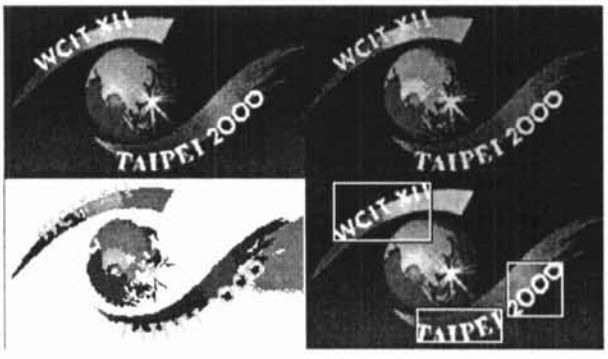

Figure 6: Experiment results 2

method fails to detect the test "OPEN" because it is composed of very thin and long strokes, which breaks width and length ratio constraint of a stroke.

Figure 7: Experiment results 3

6 Conclustion

For indexing web page images, a text region extraction method is proposed. The method uses a spatial merging algorithm to reduce color in low resolution WWW images. A stroke-based character extraction and a potential field based text string extraction are used for various orientation and multi-font size texts. Future research will focus on more robust stroke-based character extraction and intelligent merging of adjacent potential fields to follow a right direction when text strings themselves are adjacent.

References

[1] M. Gray Internet statistics: Growth and usage of the Web and the Internet, *http://www.mit.edu/people/mkgray/net/*

[2] Anil K. Jain, Bin Yu, "Automatic Text Location in Images and Video Frames," *Michigan State University Technical Report* MSUCPS : TR97-33

[3] Rainer Lienhar, "Automatic Text Recognition for Video Indexing," *The Fourth ACM International Multimedia Conference*, Multi- media96, pp.11-20, November 1996

[4] Hae-Kwang Kim, "Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database," *Journal of Visual Communication and Image Representation*, Vol 7, No 1-4, pp.336-344, 1996

[5] Rainer Lienhar, "Automatic Text Recognition for Video Indexing," *The Fourth ACM International Multimedia Conference*, Multi- media96, pp.11-20, November 1996

[6] Shoji Kurakake, Hidetaka Kuwano, "Recognition and visual feature matching of text region in video for conceptual indexing," *The International Society for Optical Engineering, Storage and Retrieval for Image and Video Databases V*, pp.368-379, February 1997

[7] Yu Zhong, Kalle Karu and Anil K. Jain, "Locating Text in Complex Color Images," *Pattern Recognition*, Vol 28, No. 10, pp.1523-1535, 1995

[8] Q. Huang, B. Dom, D. Steele, J. Ashley, and W. Niblack, "Foreground/background Segmentation of color images by integration of multiple cues," *In Proceedings of Computer Vision and Pattern Recognition*, pp246-249, 1995.

[9] D. Doermann, E. Rivlin, and I. Weiss. "Logo Recognition," *University of Maryland Technical Report* CAR-TR-688, 1993.

[10] Jiangying Zhou, Daniel Lopresti, "Extracting Text from WWW Images," *Proc. of Fourth International Conference on Document Analysis and Recognition*, pp248-252, August 18, 1997.