A novel Web image processing algorithm for text area identification that helps commercial OCR engines to improve their Web image recognition efficiency

S. J. Perantonis, B. Gatos and V. Maragos

Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
National Center for Scientific Research ì Demokritosî
153 10 Agia Paraskevi, Greece
sper@iit.demokritos.gr

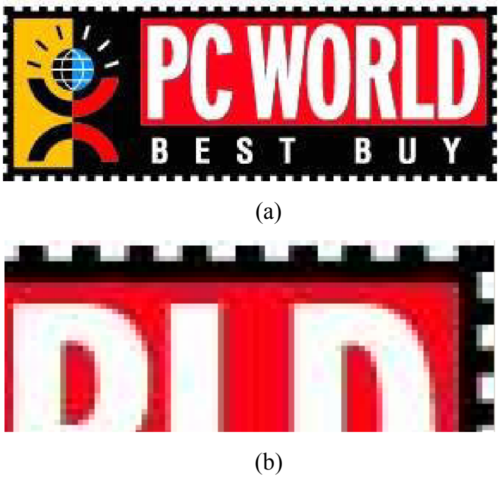
Abstract

In this paper, a novel Web image processing algorithm is presented for text area identification. Statistics show that a significant part of Web text information is encoded in Web images. Since Web images have special characteristics that sometimes distinguish them from other types of images, commercial OCR products often fail to recognize Web images due to their special key characteristics. This paper proposes a novel Web image processing algorithm that aims to locate text areas and prepare them for OCR procedure with best results. According to this algorithm, first the Web color image is converted to gray scale in order to record the transitions of brightness that are perceived by the human eye. Then, an edge extraction technique facilitates the extraction of all objects as well as of all inverted objects. A conditional dilation technique applied with several iterations helps to choose text and inverted text objects among all objects. Experimental results, obtained from a large corpus of Web images, demonstrate the improvement in recognition accuracy after applying the proposed text area identification algorithm.

1 Introduction

With the World Wide Web becoming a major source of information, a growing number of documents are published and accessed on-line. The increase and the rapid changes of this information open new perspectives for developing automatic systems to organize and search this huge and distributed corpus of web documents. The World Wide Web contains lots of information but even modern search engines just index a fraction of this information. This issue poses new challenges for Web Document Analysis and Web Content Extraction. While there has been active research on Web Content Extraction using text-based techniques, documents often include multimedia content. It has been recorded [1][2] that of the total number of words visible on a Web page, 17% are in image form and those words are usually the most semantically important. Besides, of the words in image form, 76% do not appear elsewhere in the encoded text. Furthermore, the textual description of the images is incomplete, wrong or does not exist in 56% of the cases. For these reasons, research into advanced Web multimedia document processing techniques can lead to intelligent information filtering and knowledge extraction tools. Hence, techniques that have been developed for image-based documents could prove valuable for Web documents, while at the same time, new methods for the analysis of the web multimedia content have to be developed.

Unfortunately, commercial OCR engines often fail to recognize Web images due to their special key characteristics. Web images are usually of low resolution, consist mainly of graphic objects, are usually noiseless and have the anti-aliasing property (see Fig. 1).



(a)

(b)

Fig. 1. A Web image example (a) and a zoom in to demonstrate the web image key characteristics.

Several approaches in the literature deal with text locating in color images. In [3], characters are assumed of almost uniform colour. In [4], foreground and background segmentation is achieved by grouping colours into clusters. A resolution enhancement to facilitate text segmentation is proposed in [5]. In [6], texture information is combined with a neural classifier. Recent work in locating text in Web images is based on merging pixels of similar colour into components and selecting text components by using a fuzzy inference mechanism [7]. Another approach is based on information on the way humans perceive colour difference and uses different colour spaces in order to approximate the way human perceive colour [8]. Finally, approaches [9][10] restrict their operations in the RGB colour space and assume text areas of uniform colour.

In this paper, a novel method is proposed for text area identification in Web images. The method has been developed in the framework of the EC-funded R&D project, CROSSMARC, which aims to develop technology for extracting information from domain-specific Web pages. Our approach is based on the transitions of brightness as perceived by the human eye. An image segment is classified as text by the human eye if characters are clearly distinguished from background. This means that the brightness transition from the text body to the foreground exceeds a certain threshold. Additionally, the area of all characters observed by the human eye does not exceed a certain value since text bodies are of restricted thickness. These characteristics of human eye perception are embodied in our approach. According to it, the Web color image is converted to gray scale in order to record the transitions of brightness perceived by the human eye. Then, an edge extraction technique helps the extraction of all objects as well as of all inverted objects. A conditional dilation technique helps to choose text and inverted text objects among all objects. The criterion is the thickness of all objects that in the case of characters is of restricted value.

In the sections to follow, we present our Web image processing algorithm for text area identification that helps commercial OCR engines to improve their Web image recognition efficiency, as well as experimental results that demonstrate the improvement in recognition accuracy after applying the proposed algorithm.

2 Text area location algorithm

2.1. Edge extraction

Consider a color Web image *I*. First, we covert it to the gray scale image *Ig*. Then, we define as *e* and e^{-1} the B/W edge and invert edge images that encapsulate the abrupt increase or decrease in image brightness:

$$e(x,y) = \begin{cases} 1, \text{ if } \exists\,(\mathrm{m,n}) : \mathrm{Ig(m,n) - Ig(x,y)} > \mathrm{D} \;\wedge \\ \qquad\qquad |\mathrm{m - x}| <= \mathrm{d} \;\wedge\; |\mathrm{n - y}| <= \mathrm{d} \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

$$e^{-1}(x,y) = \begin{cases} 1, \text{ if } \exists\,(\mathrm{m,n}) : \mathrm{Ig(m,n) - Ig(x,y)} < \mathrm{D} \;\wedge \\ \qquad\qquad |\mathrm{m - x}| <= \mathrm{d} \;\wedge\; |\mathrm{n - y}| <= \mathrm{d} \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

where *D* is the gray level contrast visible by the human eye and *d* defines the window at x,y in which we search for a gray level contrast. Fig. 2 shows an example for *e* and e^{-1} calculation.

0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	2	2	2	2	2	2	1	0	0
0	0	1	2	2	2	2	2	2	1	0	0
0	0	1	2	2	3	3	2	2	1	0	0
0	0	1	2	2	3	3	2	2	1	0	0
0	0	1	2	2	2	2	2	2	1	0	0
0	0	1	2	2	2	2	2	2	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

(a)

0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0
0	0	0	**1**	0	0	0	0	**1**	0	0	0
0	0	0	**1**	0	0	0	0	**1**	0	0	0
0	0	0	**1**	0	0	0	0	**1**	0	0	0
0	0	0	**1**	0	0	0	0	**1**	0	0	0
0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

(b)

0	0	0	0	0	0	0	0	0	0	0	0
0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0
0	**1**	0	0	0	0	0	0	0	0	**1**	0
0	**1**	0	0	0	0	0	0	0	0	**1**	0
0	**1**	0	0	0	0	0	0	0	0	**1**	0
0	**1**	0	0	0	0	0	0	0	0	**1**	0
0	**1**	0	0	0	0	0	0	0	0	**1**	0
0	**1**	0	0	0	0	0	0	0	0	**1**	0
0	**1**	0	0	0	0	0	0	0	0	**1**	0
0	**1**	0	0	0	0	0	0	0	0	**1**	0
0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0
0	0	0	0	0	0	0	0	0	0	0	0

(c)

Fig. 2. (a) Gray scale image Ig, (b) edge image e and (c) invert edge image e-1 (parameters used: D=2, d=2).

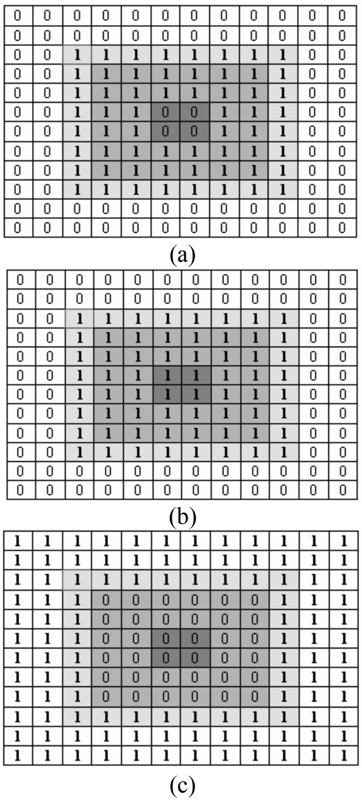
2.2. Object identification

Objects are defined as groups of pixels that neighbor with edge pixels and have similar gray scale value. To calculate image objects, we proceed to a conditional dilation of edge images. A pixel is added only if it has a similar gray scale value in the original image *Ig*. The dimension of the structuring element defines the expected maximum thickness of all objects. Objects O_s and inverted objects O_s^{-1} are defined as follows:

$$\mathrm{O_s}(x,y) = \begin{cases} 1, \text{ if } \exists\,(\mathrm{m,n}) : \mathrm{e(m,n)} = 1 \wedge |\mathrm{m - x}| <= \mathrm{s} \;\wedge\; |\mathrm{n - y}| <= \mathrm{s} \\ \qquad\qquad \wedge\, |\mathrm{Ig(x, y) - Ig(m, n)}| < \mathrm{S} \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

$$\mathrm{O_s^{-1}}(x,y) = \begin{cases} 1, \text{ if } \exists\,(\mathrm{m,n}) : \mathrm{e^{-1}(m,n)} = 1 \wedge |\mathrm{m - x}| <= \mathrm{s} \;\wedge\; |\mathrm{n - y}| <= \mathrm{s} \\ \qquad\qquad \wedge\, |\mathrm{Ig(x, y) - Ig(m, n)}| < \mathrm{S} \\ 0, \text{ otherwise} \end{cases} \tag{4}$$

where *s* the dimension of the structuring element and *S* is the expected maximum difference in gray scale values within the same object. Fig. 3 shows an example for O_s and O_s^{-1} calculation.



0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	0	0	1	1	1	0	0
0	0	1	1	1	0	0	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

(a)

0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

(b)

1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	0	0	0	0	0	0	1	1	1
1	1	1	0	0	0	0	0	0	1	1	1
1	1	1	0	0	0	0	0	0	1	1	1
1	1	1	0	0	0	0	0	0	1	1	1
1	1	1	0	0	0	0	0	0	1	1	1
1	1	1	0	0	0	0	0	0	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1

(c)

Fig. 3. For the example of fig. 2 we calculate object O_1 (a), object O_n, n>1 (b) and object O_1^{-1} (c) (parameters used: S=1).

2.3. Text identification

The above conditional dilation technique applied with several iterations (several values for the structuring elements) helps to chose text and inverted text objects among all objects. The criterion is the thickness of all objects that in the case of characters is of restricted value.

Let *P(f)*, the set of points of a b/w image *f*:

$$O(f) = \{(x,y):f(x,y)=1\} \qquad (5)$$

$p_i(f)$, the set of points of all the connected components that comprise image *f*:

$$P(f) = \cup\, p_i(f) \qquad (6)$$

$S(p_i(f))$, the number of pixels of the connected component, $E(p_i(f))$, the set of background points that have a 4-connected relation with the connected component, $S(E(p_i(f)))$, the number of pixels of $E(p_i(f))$, and $C(p_i(f))$, the category a connected component belongs to:

$$C(p_i(f)) = \text{TEXT or OTHER CATEGORY} \qquad (7)$$

A connected component of image object O_n is classified as text region if while increasing n the set of background pixels that have a 4-connected relation with the connected component remains almost the same (see the example of Fig. 3b where object O_n remains the same for n>1):

$$\begin{aligned} &C(p_i(O_n)) = \text{TEXT if} \\ &\quad \exists j\text{: } (p_i(O_n) \subseteq p_j(O_{n+1}) \text{ AND} \\ &\quad S(E(p_i(O_n)) \cap E(p_j(O_{n+1}))) / S(E(p_i(O_n)) < s) \\ &\quad \text{AND } n<N \qquad (8) \end{aligned}$$

where N depends on the maximum expected letter thickness and s is the allowed tolerance in changes of the 4-connected background pixel set. The reason we trace the changes to the 4-connected background pixels and not to the foreground pixels is that due to dilation with a larger structuring element, the connected components may be joined together. In the same way, we define the condition for locating inverse text objects.

At Fig. 4 the flowchart of the proposed method is demonstrated.

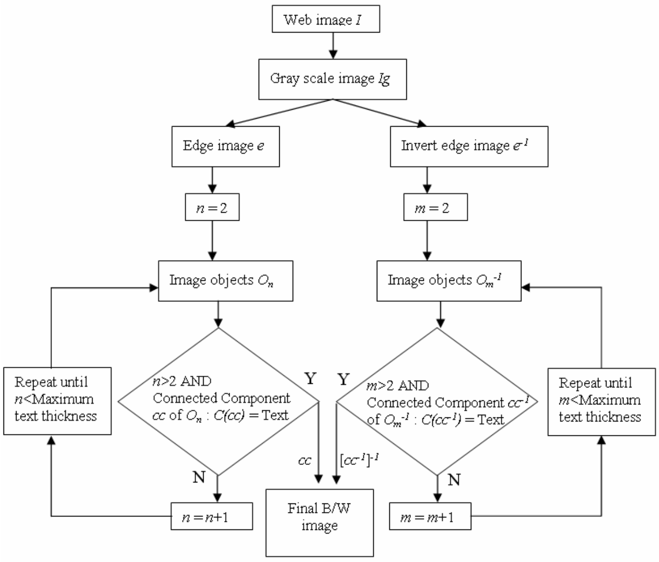
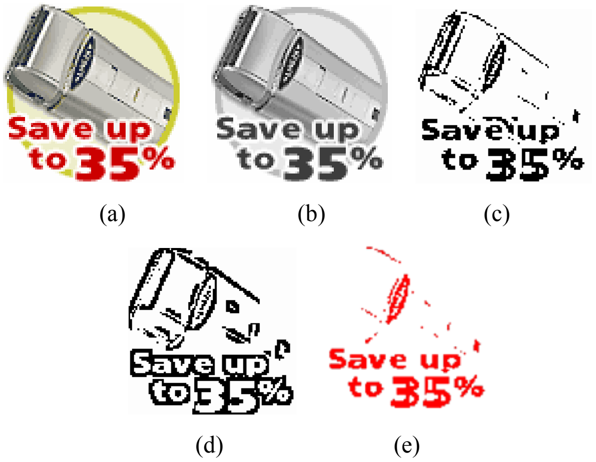

Web image I
Gray scale image Ig
Edge image e
Invert edge image e^{-1}
n = 2
m = 2
Image objects O_n
Image objects O_m^{-1}
Repeat until n<Maximum text thickness
n>2 AND Connected Component cc of O_n : $C(cc)$ = Text
Y
Y
m>2 AND Connected Component cc^{-1} of O_m^{-1} : $C(cc^{-1})$ = Text
Repeat until m<Maximum text thickness
N
cc
$[cc^{-1}]^{-1}$
N
n = n+1
Final B/W image
m = m+1

Fig. 4. Text area identification algorithm flowchart.

3 Experimental results

The proposed technique for text area identification in Web images has been implemented and tested with a large Web image corpus developed within the EU funded IST project CROSSMARC. Experiments were conducted with a variety of Web images containing text, inverse text and graphics. The corpus for the evaluation of the proposed technique was prepared by selecting more than 650 images from English, French, Greek and Italian Web pages. We compared the results obtained by the famous OCR engine FineReader 5 [11] with and without applying our text area location technique. For better OCR results, we artificially increased all Web image resolutions. In almost all cases, the recognition results were extremely improved after applying our text area identification technique. An example of the application of the proposed technique is demonstrated in Fig. 5. Text location results in terms of detection rate and recognition accuracy on the entire corpus of 650 images are shown in Table 1. Typical OCR results are shown in Table 2.



(a) (b) (c)

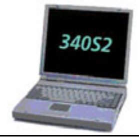
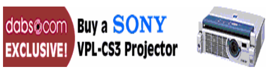
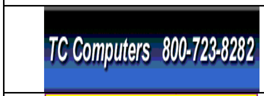
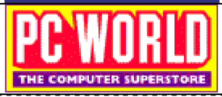
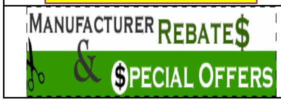
(d) (e)

Fig. 5. Text area identification algorithm example: (a) Original image, (b) gray scale image, (c) O_2 image, (d) O_2^{-1} image (e) Area identified as text.

Table 1. Text location evaluation results

	Detection Rate	Recognition Accuracy
English web image corpus	85,08	61,53
French web image corpus	84,32	61,61
Greek web image corpus	80,93	61,73
Italian web image corpus	78,41	61,50
TOTAL	**83,58**	**61,58**

Table 2. OCR results

	FineReader5	**Text extraction + FineReader5**
	-	*340S2*
	SONY *l*VPL-CS3 Projector	da*sOc*m exclusive! Buy a SONY VPL-CS3 Projector
	-	7C *Computers* *800-723-8282*
	ï ι π ´*-):^∑	PC WORLD THE COMPUTER SUPER-STORE
	manufac turer rebate^	manufacturer REBATE "∑ $ peciau offers

4 Conclusions and further work

This paper proposes a novel Web image processing algorithm for text area identification that helps commercial OCR engines to improve their Web image recognition efficiency. It is based on information on the transitions of brightness that are perceived by the human eye and includes edge detection and conditional dilation processes. The experimental results obtained are very promising since the recognition results were extremely improved after applying the proposed framework.

Further work includes investigation for parameter fine tuning, taking into account character geometric features, creating an automatic evaluation tool to record improvement of the OCR engine performance after applying our text extraction method.

References

[1] A. Antonacopoulos, D. Karatzas and J. Ortiz Lopez, "Accessing Textual Information Embedded in Internet Images", Proc. Of SPIE Internet Imaging II, pp. 198-205, San Jose, USA, January 24-26, 2001.

[2] D. Lopresti and J. Zhou, "Document Analysis and the World Wide Web", Proc. Workshop on Document Analysis Systems, pp. 417-424, Marven, Pennsylvania, October 1996.

[3] A.K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames", Pattern Recognition, Vol. 31, No. 12, pp. 2055-2076, 1998.

[4] Q. Huang, B. Dom, D. Steele, J. Ashley and W. Niblack, "Foreground/background segmentation of color images by integration of multiple cues", Proc. Of the Computer Vision and Pattern Recognition, pp. 246-249, 1995.

[5] H. Li, O. Kia and D. Doermann, "Text enhancement in digital video", Doc. Recognition & Retrieval VI (IS&SPIE Electronic Imaging"99), San Jose, Vol. 3651, pp.2-9, 1999.

[6] C. Strouthopoulos and N. Papamarkos, "Text identification for document image analysis using a neural network", Image and Vision Computing, Vol. 16, pp. 879-896, 1998.

[7] A. Antonacopoulos and D. Karatzas, "Text Extraction from Web Images Based on Human Perception and Fuzzy Inference", 1st Intíl Workshop on Web Document Analysis (WDA2001), pp. 35-38, Seattle, USA, September 2001.

[8] A. Antonacopoulos and D. Karatzas, "An Anthropocentric Approach to Text Extraction from WWW Images", Proc. Of the 4th IAPR Workshop on Document Analysis Systems (DAS2000), Rio de Janeiro, pp. 515-526, December 2000.

[9] A. Antonacopoulos and F. Delporte, "Automated Interpretation of Visual representations: Extracting textual Information from WWW Images", Visual Representations and Interpretations, R. Paton and I. Neilson (eds.), Springer, London, 1999.

[10] D. Lopresti and J. Zhou, "Locating and Recognizing Text in WWW Images", Information Retrieval, Vol. 2 (2/3), pp. 177-206, May 2000.

[11] www.finereader.com