

Text detection and restoration in natural scene images

Qixiang Ye *, Jianbin Jiao, Jun Huang, Hua Yu

College of Engineering, Graduate University of Chinese Academy of Sciences, No. 19 Yu Quan Road, Shi Jing Shan District, Beijing, PR China

Received 12 May 2006; accepted 10 July 2007

Available online 28 July 2007

Abstract

A new method for text detection and recognition in natural scene images is presented in this paper. In the detection process, color, texture, and OCR statistic features are combined in a coarse-to-fine framework to discriminate texts from non-text patterns. In this approach, color feature is used to group text pixels into candidate text lines. Texture feature is used to capture the “dense intensity variance” property of text pattern. Statistic features from OCR (Optical Character Reader) results are employed to further reduce detection false alarms empirically. After the detection process, a restoration process is used. This process is based on plane-to-plane homography. It is carried out to refine the background plane of text when an affine transformation is detected on a located text and independent of camera parameters. Experimental results tested from a large dataset have demonstrated that the proposed method is effective and practical.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Text detection; Text recognition; Text restoration

1. Introduction

Texts are important objects embedded in natural scenes. They often carry useful information such as traffic signals, advertisement billboards, dangerous warnings, etc. [1]. Automatic text recognition induces a lot of potential applications such as:

- (1) Helping a foreigner to understand the contents on an information board (by translating the recognized text into his native language);
- (2) Helping blind people to walk freely in a street; and
- (3) Drawing attention of a driver to traffic signs.

With the gaining momentum of powerful portable digital devices, automatically extracting useful information from natural scene images from them is becoming a more of a practical application. As a result, research on text recognition from scene images is becoming a hot topic in recent years [1–7].

Text detection is the very first step for the correct text recognition from an image. In a natural scene, there are a lot of text-like patterns such as building windows and fences. They are easily to be mis-identified as texts. Different from patterns like human faces and single character, text pattern is not a square image “block”. Its size varies with the change of character numbers in it (as shown in Fig. 2a). Furthermore, text appearance varies dramatically with the change of characters and character positions, making its structure uncertain. An intuitive comparison between text and human face pattern is shown in Fig. 2b. From this figure, we can see that, a “average face”, obtained by summing up the gray values of the images and average the results with the image number, keeps basic structure while “average text” contains no information. Fig. 2c shows that, by applying PCA (principle component analysis) on texts, there are far more non-zero eigen-values than that of human faces, which implies that there are many challenges in building a model to represent text pattern.

The success of OCR technique has made it an easier task to read frontal text lines. However, in some case, a camera is not perpendicular to the text plane, the text in the image

* Corresponding author. Fax: +86 10 8825 6278.

E-mail address: qxyc@gucas.ac.cn (Q. Ye).

plane may deform (as shown in Fig. 1b). This will dramatically affect the recognition result. Therefore, before feeding the located text into an OCR system, it is necessary to refine the text with affine deformation into frontal text, a process we call text restoration in this paper.

There are quite some relevant works on text detection reported in the literature, including overlay text location in both images and videos. Edge (gradient) and edge layout are often used [1,9,10]. In videos produced by video editors, overlay texts often have marked contrasts with its background making them easier to be located by edge (gradient) features. However, in natural scenes, pure edge (gradient) feature is not very effective for discriminating text from its background, this is because text itself is embedded in scene and surrounded by text-like objects such as tree leaves and window curtains and they have gradients as strong as that of text. In [11], Jain et al. proposed a classic method called connected color component analysis (CCA) for text location. CCA uses spatial structure analysis of the color connected components and can work well on most kinds of texts such as the characters on book covers, news titles, and video captions. The authors obtained high recall rate at the cost of high false alarm rate through CCA. In [12], Gao et al. combined the edge information with color layout analysis to detect scene text, the principal idea of which was to depend on the “dense edge” property to locate text region. However, their method may fail when texts are surrounded by or near to objects of dense edges. In [8,13–15], a texture discrimination method was used to distinguish text from non-text. Unfortunately, we have found out that using the experiments reported in [15], the variable content of the text makes the text a weak and irregular texture property. It is difficult to discriminate text, especially scene text, from general textures by pure texture features.

For text restoration, Chen et al. [1,16] proposed a method for affine deformation restoration using camera parameters and vanish points of parallel lines of the text bounding box. Clark et al. [17] also used a similar method for text line refinement from scanned documents. In both methods, precise outline rectangles of text lines are needed, with shape and texture change of the characters, which will affect the restoration result dramatically. Furthermore, these two methods rely on precise camera parameters which will definitely limit their usefulness.

In this paper, an innovative and robust method for text detection and restoration is proposed for text recognition in natural scene image. In text location, color, texture, and OCR feedback are combined in different stages to segment the text area. We use a color quantization method to separate text from its background and then use region spatial layout analysis to locate candidate text lines. Since text is made up of vertical/horizontal/skew strokes, histogram features of wavelet coefficients and color variance features are extracted to capture the texture properties of texts. However, although these texture features can be used to eliminate most of the non-text part in the images, they are not effective for discriminating text with image blocks of large gradients, especially for general textures. Statistic features from OCR are used to further reduce the false alarms after the texture-based text/non-text classification. The OCR software based on shape feature of isolated character can finally discriminate candidate into text or non-text. For located text line, a procedure is carried out to judge whether the text plane is of affine transformation. If it is, a homography operation between image plane and text plane is proposed to refine the distorted text line. In the procedure, we use line intersections as the corresponding points for homography operation. Fig. 3 shows the flow chart of the proposed method.

Compared with existing approaches, the proposed algorithm has the following advantages:

(1) *Scale invariance*: We use color information and adaptive region layout analysis to locate candidates in the text detection method, which makes it unnecessary to consider the scale problem.

(2) *“Uncertain” pattern detection by feature combination*: It has been made clear that the text pattern is “uncertain” in both length and appearance. To detect this “uncertain” pattern, a feature combination method is proposed. Color, spatial layout, texture, and statistics on OCR result are used in the region location, candidate text location, text verification stages, respectively. This is a new feature combination method for text detection.

(3) *Affine restoration without using camera parameters*: By using simple but effective homography operation between two planes, we can restore the text plane of affine deformation. No camera parameters are needed in the process.

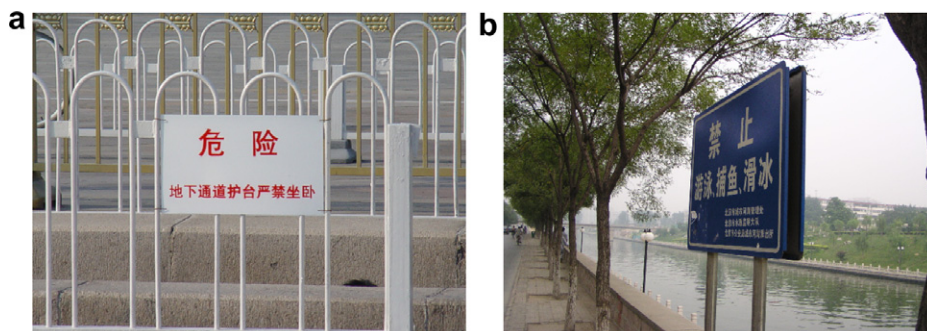


Fig. 1. Text examples. (a) Text with complex scene and (b) with affine deformation.

Download English Version:

<https://daneshyari.com/en/article/529575>

Download Persian Version:

<https://daneshyari.com/article/529575>

[Daneshyari.com](https://daneshyari.com)