Pattern Recognition Letters 31 (2010) 1728–1739

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Automatic detection and recognition of Korean text in outdoor signboard images

Jonghyun Park [a,*], Gueesang Lee [a], Euichul Kim [a], Junsik Lim [a], Soohyung Kim [a], Hyungjeong Yang [a], Myunghun Lee [a], Seongtaek Hwang [b]

[a] School of Electronics and Computer Engineering, Chonnam National University, Republic of Korea
[b] Multimedia Lab., Telecommunication R&D Center, Telecommunication Network Business, Samsung Electronics Co. Ltd., Republic of Korea

ARTICLE INFO

Article history:
Received 27 November 2009
Available online 1 June 2010
Communicated by C.L. Tan

Keywords:
Text detection
Text extraction
Text recognition
Text translation
Signboard image

ABSTRACT

In this paper, an automatic translation system for Korean signboard images is described. The system includes detection and extraction of text for the recognition and translation of shop names into English. It deals with impediments caused by different font styles and font sizes, as well as illumination changes and noise effects. Firstly, the text region is extracted by an edge-histogram, and the text is binarized by clustering. Secondly, the extracted text is divided into individual characters, which are recognized by using a minimum distance classifier. A shape-based statistical feature is adopted, which is adequate for Korean character recognition, and candidates of the recognition results are generated for each character. The final translation step incorporates the database of shop names, to obtain the most probable result from the list of candidates. The system has been implemented in a mobile phone and is demonstrated to show acceptable performance.

1. Introduction

Text contained in an image usually provides semantic information, and is often the crucial key to understanding the image content. These days, digital image capturing devices, including multi-functional mobile devices, are prevalent, and interest in the extraction of text information from natural scene images is increasing (Obinata and Dutta, 2007; Jung et al., 2004). Although text in a regular font of fixed size, with uniform background, can be successfully recognized by existing methods, it is difficult to recognize variable text as used in the real world. Compared with other text detection algorithms developed for computer vision applications, difficulties are compounded in natural scene images due to numerous reasons. The primary problems lie in the variety of the text appearance as it can vary in font style, size, orientation and position. Also, the text images are often corrupted by shadows, reflection of lights, or uneven illumination. They can also be distorted by slants or tilts caused by the position of the camera. Although many commercial OCR systems have good recognition capabilities on high quality scanned documents under well controlled environments, much higher error rates are common for character recognition in real world situations, when the input images do not satisfy the enforced constraints.

This paper presents an automatic translation system for Korean signboard images, where the text usually represents a shop name. The system was originally designed for foreigner tourists with little, or no, knowledge of the Korean language. The system consists of text detection, binarization, recognition and translation. First, local clustering is used to effectively handle the luminance variations of the captured images. The bounding boxes of the individual characters are obtained by connected component analysis. Secondly, the text is recognized using direction features extracted from each character region using a non-linear mesh. Finally, a database of shop names has been used to generate a translation result from the list of recognition candidates. The novel idea of this system is in the translation scheme, in which a database of shop names has been incorporated to compensate for the possible incorrectness of the recognition result, and the most probable interpretation of the word is generated by referencing the database. The block diagram of the proposed system is shown in Fig. 1.

The rest of this paper is organized as follows: In Section 2, problems involved in automatic detection and recognition of text in natural scene images are introduced, with related works in this area. In Section 3, the new proposed method for text detection and binarization is described. In Section 4, a shape-based statistical feature is adopted, which is adequate for Korean character recognition, and the final translation step is described in Section 5. In Section 6, experimental results are presented and Section 7 concludes the paper.

* Corresponding author. Tel.: +82 62 530 0147.
E-mail addresses: jhpark@chonnam.ac.kr, jhpark@jnu.ac.kr (J. Park), gslee@jnu.ac.kr (G. Lee), eckim@jnu.ac.kr (E. Kim), jslim@jnu.ac.kr (J. Lim), shkim@jnu.ac.kr (S. Kim), hjyang@jnu.ac.kr (H. Yang), shwang@samsung.com (S. Hwang).

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.05.024

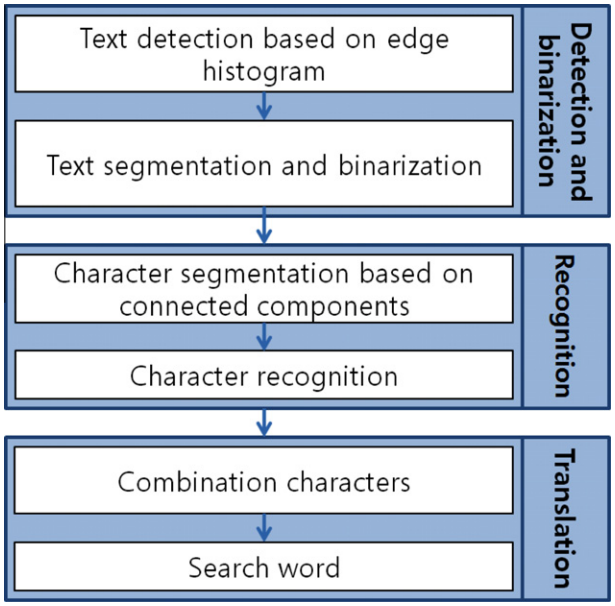

Fig. 1. The block diagram of the proposed system.

2. Related works

The interest of this paper is in the automatic detection and recognition/translation of texts in outdoor signboard images for mobile applications. The application scenario is as follows: a camera is used to capture an image, or a sequence of images containing the text of interest.

There have been techniques previously developed for the detection and recognition of written text, targeting applications such as document image processing (Nagy, 2000; Peng et al., 2003), content-based image/video indexing (Li et al., 2000; Xi et al., 2001), assistance for visually impaired persons (Ezaki et al., 2005), text restoration (Ye et al., 2007; Lim et al., 2007), and text recognition or sign translation in natural scene images (Zhang et al., 2002; Yang et al., 2001).

There are many relevant works, particularly, for the detection of a text region. Most of the text detection methods can be classified as either edge-based methods, connected component based methods or texture-based methods. These methods have their own advantages and disadvantages in terms of reliability, accuracy and computation complexity.

The edge-based methods focus on the high contrast between the text and the background (Gllavata et al., 2003; Wu et al., 2005). Once the edges of the text boundary are identified, heuristics are used to filter out the non-text regions. Generally, edges are very useful for image analysis, and they can be used to find the text area, or the bounding area around the text. Text is mainly composed of strokes in a certain direction, so the region with higher edge strengths in a specific direction has a high probability of containing text.

Text detection methods based on connected components use the spatial structure of the connected components, and these work well on text on book covers, news titles or video captions (Jain and Yu, 1998). The connected component based method is a bottom-up approach in the sense that small components are grouped into successively larger components until all relevant regions in the image have been identified. The geometric properties of the connected components are considered in order to filter out the non-textual components, and to mark the boundaries of the region containing the text of interest.

The texture-based method assumes that the text in an image has a distinct textural property which can be used to discriminate the text from the background, or from other non-textual regions (Fujii and Hoefer, 2001). The texture-based algorithms are more robust in dealing with complex backgrounds than the connected component based methods. Tang et al. (2002) has proposed a text detection algorithm that uses texture properties, but the algorithm fails if the texture information cannot be extracted, e.g., in the case of small sized text. Fujii and Hoefer (2001) has proposed a method using wavelet features obtained from the fixed sized blocks of the pixels, and the feature vectors are classified into textual or non-textual groups using neural networks. However, neural networks are not generally efficient in terms of their large computation cost in highly complex images.

For the extraction of a text area, there have been many approaches developed to deal with the possible variations in text orientation, text size, the language used, low image quality, and so on. Most of the existing approaches assume that the text strokes are either horizontal or vertical, and that the text font size is fixed, hence they are very restrictive in the text types they can process.

In order to obtain information in an arbitrary image or a scene captured using a camera, the recognition system automatically recognizes characters of various conditions in the scene, and then provides information about the location of text within the captured image. The text regions, detected in the previous step, can be fed into the recognition step for the classification. This work is related to existing research in the recognition of text on special objects such as car license plates (Mullot et al., 1991). While the early methods required manual selection of the text area (Watanabe et al., 1998; Yang et al., 1999), recent attempts have moved toward automatic detection and recognition of text in natural scenes, for mobile system applications (Zhang et al., 2002; Yang et al., 1999).

Aside from the challenges for the recognition of text in natural scenes, the restrictions on execution time or the overall complexity of the algorithm also need to be enforced in real-time environments, as in mobile systems or personal digital assistants (PDAs). In order to overcome these problems, more robust features are required for the recognition and translation. Efficient methods for text detection, recognition and translation, with their robustness features, are given in the following sections.

3. Text detection

Texts in outdoor signboard images are affected by changes of lights, orientation and the actual location of the signboards, where the orientation is decided by the viewing angle of the camera. In this paper, it is assumed that the text region is located around the center line. In this section, a hierarchical detection framework is presented, including the computation of the edge histogram and the text extraction using fuzzy clustering and connected component analysis.

3.1. Detection of the candidate text region

Although the intensity of pixels in a digital image are important features for text detection, it is not robust in dealing with the variations in the lighting. On the other hand, the edge component is less sensitive to light changes and therefore is more dependable for identifying the text detection. The canny edge detector is applied to the gray-scale image to obtain the edges of the input image. In order to detect the candidates for the region containing the text, the horizontal profile of the edges are computed.

In this paper, it is assumed that texts are aligned horizontally, but the proposed approach can also be extended to vertically aligned texts. Fig. 2 shows edge profiles in the horizontal and