Pattern Recognition 43 (2010) 1518–1530

Contents lists available at ScienceDirect

Pattern Recognition

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

A binarization method with learning-built rules for document images produced by cameras

Chien-Hsing Chou [a], Wen-Hsiung Lin [b], Fu Chang [b,*]

[a] Department of Electrical Engineering, Tamkang University, Taiwan
[b] Institute of Information Science, Academia Sinica, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 7 July 2006
Received in revised form
27 June 2008
Accepted 23 October 2009

Keywords:
Document image binarization
Global threshold
Image processing
Local threshold
Multi-label problem
Non-uniform brightness
Support vector machine

ABSTRACT

In this paper, we propose a novel binarization method for document images produced by cameras. Such images often have varying degrees of brightness and require more careful treatment than merely applying a statistical method to obtain a threshold value. To resolve the problem, the proposed method divides an image into several regions and decides how to binarize each region. The decision rules are derived from a learning process that takes training images as input. Tests on images produced under normal and inadequate illumination conditions show that our method yields better visual quality and better OCR performance than three global binarization methods and four locally adaptive binarization methods.

1. Introduction

Binarizing images of documents captured with camera-equipped electronic devices, such as PDAs or cellular phones, presents a new challenge. The captured content can be transformed in various ways. For example, French content can be translated into English, or information on a business card can be identified and stored in proper categories, such as name, address, telephone number, etc. The new challenge to binarization arises because such images are produced under illumination conditions that are inferior to those found in a scanning environment. As a result, there are varying degrees of brightness over the images. If we simply apply a global threshold, as we do with scanned images, the binarized results could be too bright in one area and too dark in another area. A more effective way of binarizing such images is therefore desired.

Before discussing the problem in detail, we briefly review binarization methods proposed in the literature. Following Sezgin and Sankur [1], we classify the methods into six categories.

Histogram-based methods: These methods determine the binarization threshold by analyzing the shape properties of the histogram, such as the peaks and valleys [2], or the concavities [3]. Pavlidis [4] constructs a histogram by using gray-image pixels with significant curvature, or second derivative, values and then selects a threshold based on the histogram.

Clustering-based methods: The threshold is selected by partitioning the image's pixels into two clusters at the level that maximizes the between-class variance [5], or minimizes the misclassification errors of the corresponding Gaussian density functions [6].

Entropy-based methods: These methods employ entropy information for binarization [7].

Object attribute-based methods: These methods select the threshold based on some attribute quality (e.g., edge matching of Hertz and Schafer [8]) or the similarity measure between the original image and the binarized image [9].

Spatial binarization methods: These methods binarize an image according to the higher-order probability or the correlation between pixels [10].

Locally adaptive methods: These methods compute a local threshold based on the information contained in the neighborhood of each pixel, or in the region of the image. In Bernsen's method [11], for example, the threshold is a function of the lowest and highest gray values; however, in Niblack's method [12], it is a function of the mean and the standard deviation of the gray scales. Taxt et al. [13] apply the EM algorithm to compute the local threshold, while Eikvil et al. [14] apply Otsu's method to

* Corresponding author.
E-mail addresses: ister@iis.sinica.edu.tw (C.-H. Chou), bowler@iis.sinica.edu.tw (W.-H. Lin), fchang@iis.sinica.edu.tw (F. Chang).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.10.016

compute it. Mardia and Hainsworth [15] compute the local threshold based on the estimation of two-point spatial covariance. Chow and Kaneko [16], and Nakagawa and Rosenfeld [17], compute the local threshold by analyzing the bimodality of gray values. The threshold can also be determined by comparing a pixel's gray value with the average gray value of the pixels in its neighborhood [18], or its local variance [19]. Sauvola and Pietikäinen [20] (also, Sauvola et al. [21]) first partition an image into windows and rapidly classify each window into background, picture and text. They then apply various binarization rules to the different types of window. Kim [22] modifies Sauvola and Pietikäinen's approach by introducing more than one window size for textual content. As an alternative to the above approaches, some special features extracted in the neighborhood of a pixel can be used to determine the local threshold. Examples are gradient information [23–25], and character stroke width [26–28].

If a binarization method computes a threshold for an entire image, it is called *a global method*. Trier and Taxt [29] evaluated four such methods [5–7,10] and concluded that Otsu's approach [5] outperforms the other three. On the other hand, if a method computes a threshold for the neighborhood around each pixel or for each designated block in the image, it is called a *local method*. Trier and Jain [30] evaluated some of these methods [11–16,18,23–25] and concluded that those proposed by Bernsen [11], Niblack [12] and Eikvil et al. [14] are the top-ranked local threshold methods in terms of the error rate and rejection rate for character recognition, and also for the visual criterion. More complete surveys of image thresholding techniques can be found in [1,29–35].

As noted earlier, using cameras to produce document images creates a new challenge for document image binarization. To address the problem, Park et al. [36] proposed block adaptive binarization of business card images produced by a PDA camera. This method is very similar to that of Eikvil et al. [14], which partitions an input image into blocks. For a given block, b, a larger concentric block, denoted as $L(b)$, is found and Otsu's method is applied to it. If the difference between the means of two classes, determined by Otsu's method, exceeds a certain threshold, block b is classified as a content block; otherwise, it is classified as a background block. Content blocks are binarized according to Otsu's thresholds, while background blocks are set directly to white or black based on the average of gray values found in them. The method in [36] differs from that of [14] in the way content blocks are differentiated from background blocks, and also in the way the sizes of b and $L(b)$ are set.

The proposed method also divides a document image into smaller areas, but differs from the methods proposed in [14,36] in a number of respects. For example, instead of dividing an image into fixed-size blocks, we divide it into $k \times k$ regions, using the value of k obtained in experiments. Dividing each image into the same number of regions ensures that the binarization effect is relatively invariant with respect to the resolution of the camera. Within each region r, one of the following four operations is applied: set the whole of r to black, set the whole of r to white, use Otsu's method to compute the threshold for r, or use the smallest Otsu threshold in the neighboring regions as the threshold for r. A learning process is used to establish the rules for deciding which of the above options should be adopted for each region. The rules are expressed as decision functions, which take a number of features extracted from r as input. The experiment results demonstrate that the above factors have a significant impact on the successful performance of our method.

The crucial step in our approach is establishing rules to decide which operation should be applied to each sub-divided region. To do this, we utilize a machine learning approach, namely, the support vector machine (SVM) [37,38], which represents a major development in pattern classification research. Two innovations of SVM are responsible for its success: (1) the ability to find a hyperplane that divides samples into two groups with the widest margin between them; and (2) the extension of the concept in (1) to a higher-dimensional setting using a kernel function to represent a similarity measure on that setting. Both innovations can be formulated in a quadratic programming framework whose optimal solution is obtained in a computation time of a polynomial order. This makes SVM a practical and effective solution for many

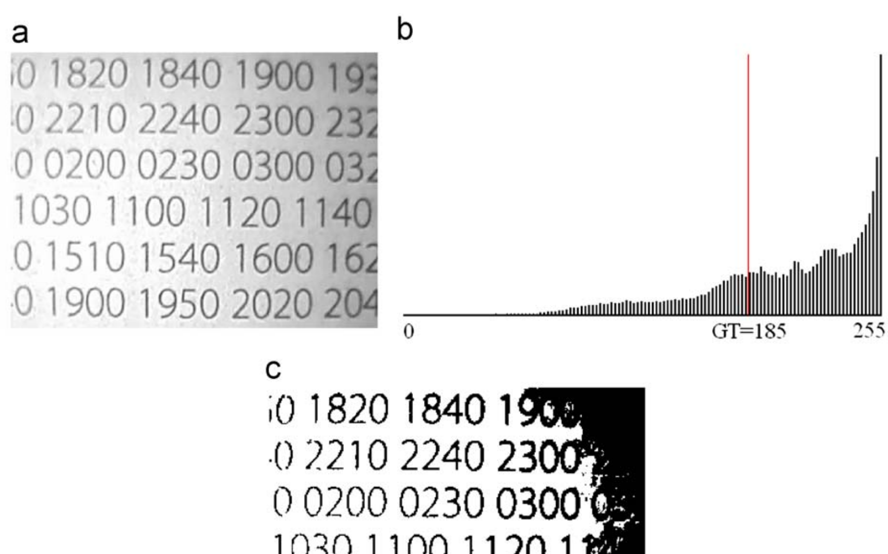

Fig. 1. (a) A document image obtained by a camera under the inadequate illumination condition. (b) The histogram of (a). (c) The document image binarized using Otsu's method.