Pattern Recognition Letters 35 (2014) 3–15

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

A combined approach for the binarization of handwritten document images

K. Ntirogiannis [a,b,*], B. Gatos [b], I. Pratikakis [c]

[a] Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimioupoli, Ilissia, GR-15784 Athens, Greece
[b] Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", GR-15310 Agia Paraskevi, Athens, Greece
[c] Department of Electrical and Computer Engineering, Democritus University of Thrace, GR-67100 Xanthi, Greece

ARTICLE INFO

Article history:
Available online 12 October 2012

Keywords:
Document image binarization
Document image pre-processing
Background estimation
Inpainting

ABSTRACT

There are many challenges addressed in handwritten document image binarization, such as faint characters, bleed-through and large background ink stains. Usually, binarization methods cannot deal with all the degradation types effectively. Motivated by the low detection rate of faint characters in binarization of handwritten document images, a combination of a global and a local adaptive binarization method at connected component level is proposed that aims in an improved overall performance. Initially, background estimation is applied along with image normalization based on background compensation. Afterwards, global binarization is performed on the normalized image. In the binarized image very small components are discarded and representative characteristics of a document image such as the stroke width and the contrast are computed. Furthermore, local adaptive binarization is performed on the normalized image taking into account the aforementioned characteristics. Finally, the two binarization outputs are combined at connected component level. Our method achieves top performance after extensive testing on the DIBCO (Document Image Binarization Contest) series datasets which include a variety of degraded handwritten document images.

1. Introduction

Document image binarization is the process that segments the document image into text and background by removing any existing degradations. It is an important pre-processing step of the document image processing and analysis pipeline that affects the segmentation stage as well as the final OCR performance. There are many challenges addressed in handwritten document image processing especially in the historical documents which are usually degraded. Handwritten documents are more difficult to be processed than the machine-printed documents because they lack a specific structure. For instance, in handwritten documents the characters may be connected within a word and words from different text lines may be connected due to the calligraphic writing style. Additionally, the use of pen quills, which was mostly used in historical handwritten documents, is responsible for several degradations including faint characters (Fig. 1(a)), bleed-through (Fig. 1(c)) and large stains (Fig. 1(b)).

Many document image binarization methods have been proposed which are usually classified in two main categories, namely global and local. Global thresholding methods use a single threshold for all the image, while local methods find a local threshold based on local statistics and characteristics within a window (e.g. mean μ, standard deviation σ, edge contrast). Reference points in binarization are considered the global thresholding method of Otsu (1979, hereafter Otsu), and the local methods of Niblack (1986, hereafter Niblack) and Sauvola and Pietikainen (2000, hereafter Sauvola) which are widely incorporated in binarization methods that followed (e.g. Kim et al., 2002; Gatos et al., 2006; Lu et al., 2010). In the case of document images with bimodal histogram, Otsu yields satisfactory results but cannot effectively handle documents with degradations such as faint characters, bleed-through or uneven background. Niblack calculates for each pixel local statistics (μ, σ) within a window and adapts the local threshold $(T = \mu + k * \sigma)$ according to those local statistics. Niblack has the advantage to detect the text but it introduces a lot of background noise. Sauvola modified the Niblack threshold to decrease the background noise but the text detection rate is also decreased while bleed-through still remains in most cases.

Certain binarization methods have incorporated background estimation and normalization steps (e.g. Gatos et al., 2006; Lu et al., 2010; Messaoud et al., 2011), as well as local contrast computations to provide improved binarization results (e.g. Su et al., 2010; Valizadeh and Kabir, 2012; Howe, 2011; Hedjam et al., 2011). Other binarization methods, aiming in an increased binarization performance, proposed combination methodologies of bina-

* Corresponding author at: Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimioupoli, Ilissia, GR-15784, Athens, Greece.
E-mail addresses: kntir@iit.demokritos.gr, kntir@di.uoa.gr (K. Ntirogiannis), bgat@iit.demokritos.gr (B. Gatos), ipratika@ee.duth.gr (I. Pratikakis).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.09.026

rization methods (e.g. Gatos et al., 2008; Su et al., 2011), while water-flow based methods (e.g. Kim et al., 2002), consider the image as 3D terrain with valleys and mountains corresponding to text and background regions, respectively. Representative binarization methods of the aforementioned categories are described in the following.

In (Kim et al., 2002) method, the original image was considered as a 3D terrain on which water was poured to fill the valleys that represented the textual components. The final binarization result was produced by applying Otsu to the compensated image, i.e. the difference between the original image and the water-filled image. Gatos et al. (2006), used wiener filter for pre-processing and estimated the background taking into account Sauvola's binarization output. The final threshold was based on the difference between the estimated background and the preprocessed image while post-processing enhanced the final result. Although this method achieves better OCR performance than Sauvola, it inherits some drawbacks of Sauvola. For instance, faint characters cannot be successfully restored while bleed-through remains. In the recent work of Lu et al. (2010), the background was estimated using polynomial smoothing for each row and column of the original image. Then, the original image was normalized and Otsu was performed to detect the text stroke edges. Furthermore, their local threshold formula was based on the local number of the detected text stroke edges and their mean intensity. Although this method provides high overall results achieving the first position in DIBCO'09 contest (Gatos et al., 2011), it has certain limitation as stated by the authors. It is based on the local contrast for the final thresholding and hence some bleed-though noise remains or noisy background components of high contrast. Messaoud et al. (2011) proposed a combination between a preprocessing step and a localization step. In the preprocessing step, wiener filtering or shading correction was performed after estimating the background using median filtering. In the localization step, Canny (1986, hereafter Canny) edges were used along with their bounding boxes. The final binarization was performed within the bounding boxes using Otsu, Sauvola or Lu et al. (2010) threshold. Although Canny edges may miss some information or detect noise, this method provides relatively good results and it is ranked at the 4th position of the DIBCO'11 contest concerning all images (both printed and handwritten). In Fig. 2(b), the noise introduced by the use of bounding boxes can be seen.

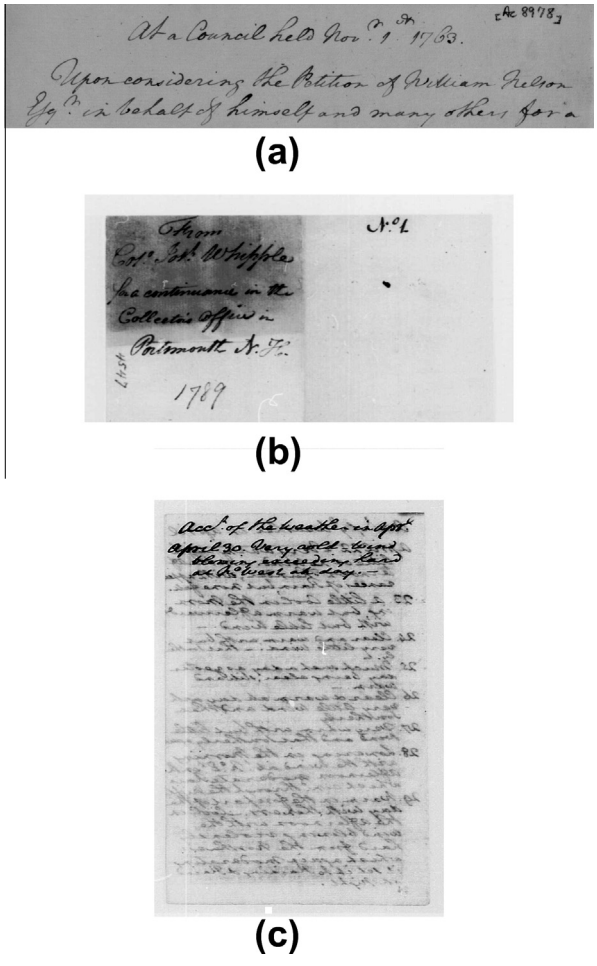

Fig. 1. Example degraded handwritten document images containing: (a) faint characters; (b) uneven background; (c) bleed-through.

In (Su et al., 2010), the authors calculated the image contrast (based on the local maximum and minimum intensity; van Herk (1992)) and binarized the contrast image using Otsu to detect the text edge pixels. They used a local thresholding formula similar to Lu et al. (2010) and presented better results than Lu et al. (2010). The authors also participated with a modified version of this method (Canny edges are additionally considered) in the binarization contests of H-DIBCO'10 and DIBCO'11 (Pratikakis et al., 2010; Pratikakis et al., 2011a), where they achieved 1st and 2nd rank, respectively. This method is capable of removing the majority of the background noise and bleed-through but it does not detect the faint characters effectively (Fig. 2(c), Pratikakis (2011b)). Howe (2011) proposed a method based on the Laplacian of the image intensity, in which an energy function was minimized via a graph-cut computation. It incorporates Canny edge information in the graph construction to encourage solutions where discontinuities align with detected edges. It is an efficient method that uses several parameters. The author tuned the parameters on the DIBCO'09 set and achieved 3rd position on H-DIBCO'10 and the same position on DIBCO'11 (including both printed and handwritten images). However, this method misses faint character parts and introduces background noise (Fig. 2(e)). Hedjam et al. (2011) used Gaussian models for the foreground and the background and performed inpainting (Bertalmio et al., 2000) using the foreground of Sauvola binarization as the inpainting mask. The inpainting mask and the inverse version of it were filled in using local statistics (e.g. mean, standard deviation) of the background and the foreground, respectively. Experimental results on the DIBCO'09 demonstrate a slight improvement of the Lu et al. (2010) method, however information about the presence of bleed-through was considered to be known. Ntirogiannis et al. (2009), modified the logical level technique of Yang and Yan (2000) which considers both the gray-level and the local contrast calculated from anti-diametric points. The authors used the Gatos et al. (2006) binarization result to adaptively define the local stroke width and modified the local threshold criterion to augment the contrast in favour of faint character detection. This method had similar performance to Gatos et al. (2006) and could not effectively remove the bleed-through (as being similar to faint characters) and noise in regions of local contrast changes (Fig. 2(d)).

As far as the combination methodologies are concerned, Su et al. (2011) proposed a framework for combination of binarization methods in which pixels are classified in 3 categories, foreground, background and uncertain. Specifically, a pixel is considered foreground if it is a foreground pixel in all the combined binarization outputs and the same holds for the background pixels. Uncertain pixels are classified (iteratively until all uncertain pixels are classified) according to their difference (in terms of intensity and local contrast), from the background and the foreground classified pixels within a neighbourhood. Local contrast is calculated according to the image intensity and the maximum intensity. In (Su et al.,