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.

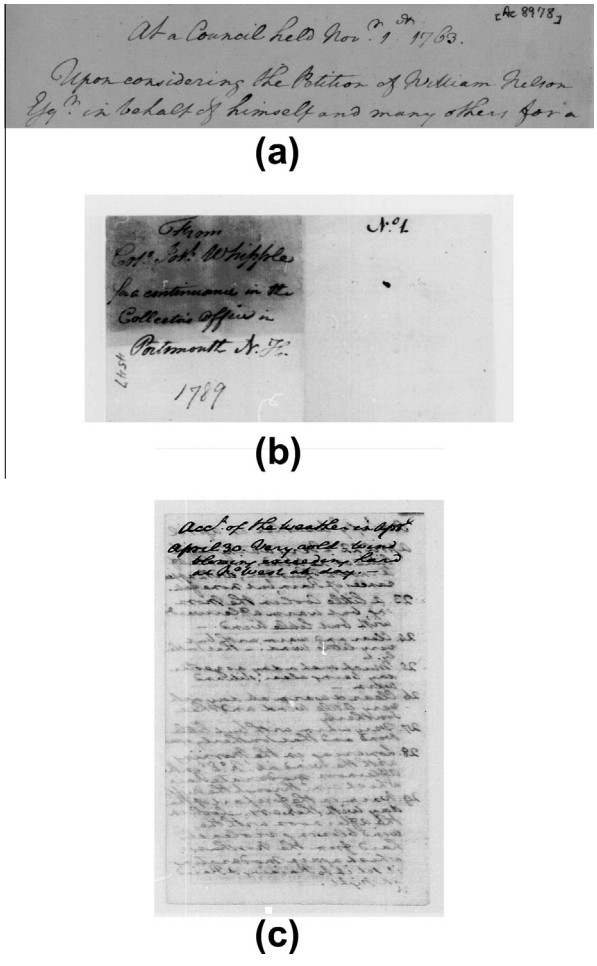

Fig. 1. Example degraded handwritten document images containing: (a) faint characters; (b) uneven background; (c) bleed-through.

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.,

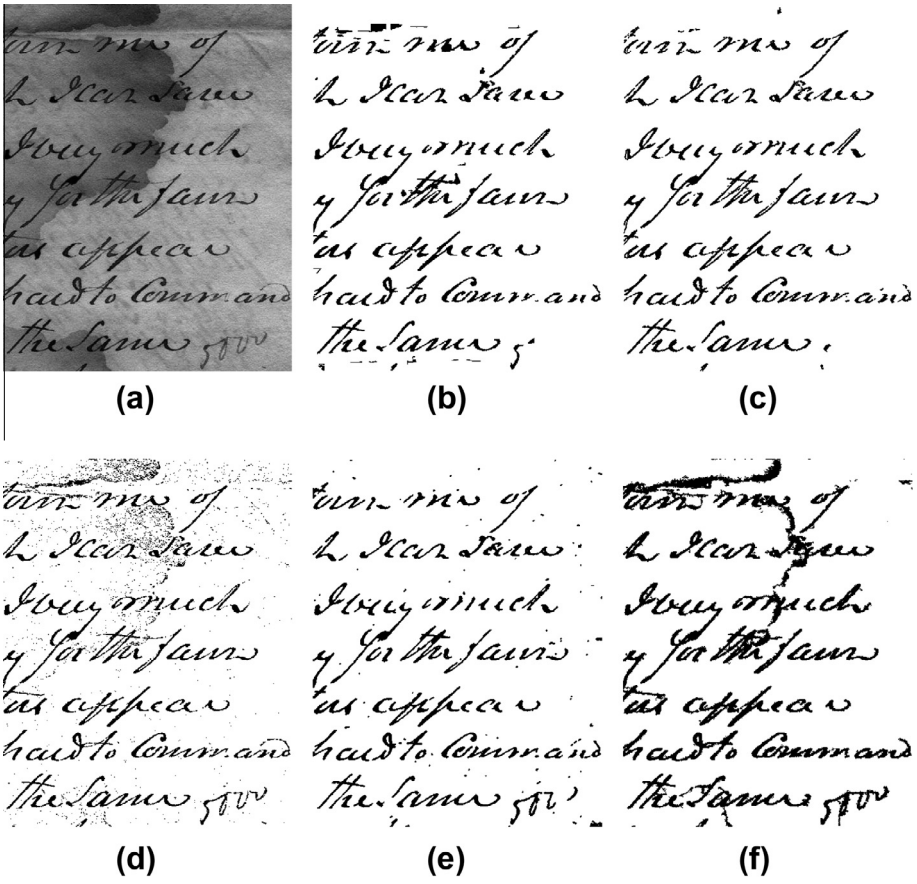

Fig. 2. (a) Representative image from DIBCO'11 dataset; corresponding binarization result produced by the method (b) Messaoud et al. (2011); (c) Su et al. (2010); (d) Ntirogiannis et al. (2009); (e) Howe (2011); (f) Gatos et al. (2008).

2011), the authors presented good results on the whole dataset (both printed and handwritten) of the DIBCO'09 and did not report any results on H-DIBCO'10 contest. Furthermore, the proposed contrast computation augments the contrast in the faint characters but it also augments the contrast in background noisy components and bleed-through. Another framework for combination of binarization methods was proposed by Gatos et al. (2008), in which foreground pixels from different binarization methods were selected using a majority vote. Canny edges were also incorporated to enhance the combination result by filling white runs between the Canny edges that correspond to foreground. This method improves the results of the authors previous work (Gatos et al., 2006) but it depends on the Canny edges by which faint characters can be undetected while background noise including bleed-through can be detected (Fig. 2(f)).

Motivated by the challenge to handle both faint characters and bleed-through which are frequently appearing in handwritten document images, we developed a new combined binarization method. The proposed method contains distinct steps including inpainting, background estimation and image normalization, contrast calculation and combination of global and local binarization results. For the background estimation, we follow an inpainting procedure which improves the combined binarization result. Furthermore, we calculate the contrast in a global manner and perform combination at connected component level to fully detect the faint characters when possible. Additionally, we perform post-processing at an intermediate step to remove small noisy components that the background estimation and image normalization cannot detect and remove, also avoiding in this way the combination with noisy background components.

In the following, the proposed method is detailed in Section 2, while in Section 3 the experimental results are demonstrated. Finally, conclusions are drawn in Section 4.

2. The proposed method

The proposed combined binarization approach comprises several steps. In the first step, background estimation is performed using an inpainting procedure initialized by the Niblack binarization output. In the sequel, image normalization is applied to correct large background variations. The normalized image is used in the global and local binarization steps that follow.

The global Otsu binarization followed by post-processing to discard very small components leads to the removal of most of the background noise, however, faint character could be also removed. In this respect, an appropriate binarization method that detects the faint characters is required. For this purpose, Niblack has been considered as the best candidate. For efficient estimation of Niblack's parameters, the window size w is set according to the character stroke width SW and the k parameter according to the image contrast C. For efficient stroke width and image contrast estimation a skeletonization step is performed.

In the final step, the aforementioned Otsu and Niblack binarization outputs are combined at connected component (CC) level to produce the final binarization result. It is worth mentioning that the proposed method, developed for handwritten documents, assumes that text is darker than the background at each step; thus if applied to machine-printed documents with areas of inverted text, those areas should be inverted back to normal (darker

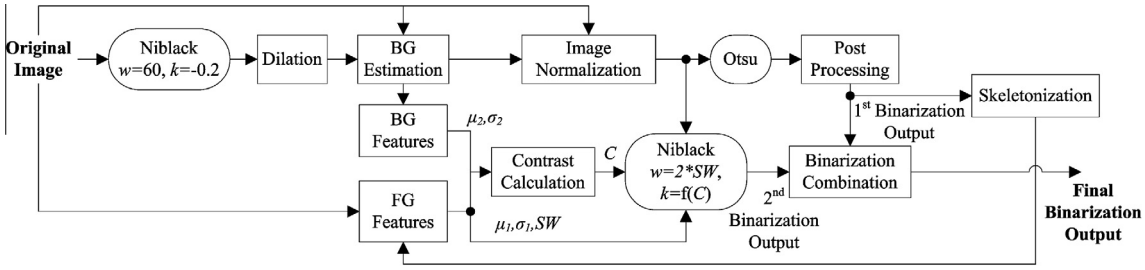

Fig. 3. The stages of the proposed technique.

characters on brighter background). A flow diagram of the distinct steps of the proposed method is shown in Fig. 3. These steps are detailed in the following sections.

2.1. Inpainting for background estimation

Concerning background estimation, some techniques use a smoothing procedure (e.g. Lu et al., 2010; Messaoud et al., 2011), while Gatos et al. (2006) use the Sauvola binarization result as a mask for interpolation calculating the average value of the Sauvola's background pixels within a large window. Another way for approximating the background is the use of inpainting. Zhang et al. (2009) used a quite complex (and relatively time-consuming) iterative inpainting procedure based on Chan and Shen (2002), mainly for shape reconstruction and background restoration purposes. They used the Canny edges and performed morphological dilation followed by closing operation to get the final inpainting mask.

In the proposed method, the underlying idea for the background estimation is to perform inpainting using Niblack's binarization output as the inpainting mask, since Niblack detects almost all the textual content achieving very high Recall rate as demonstrated in (Ntirogiannis et al., 2008, 2009). Particularly, the estimated background image inherits the intensity of the original image for pixels that correspond to background pixels of Niblack, while the foreground area produced by Niblack is filled in during inpainting with surrounding background information of the original image. In the proposed method, Niblack's foreground result is initially dilated using a 3x3 mask (Fig. 4(b)) in order to exclude background pixels near the character edges which may have intensity closer to the foreground intensity. At this stage, we use fixed parameter settings for Niblack that can handle effectively most cases, i.e. window size $w = 60$ and $k = -0.2$.

For the proposed inpainting, we use a simple, fast and effective procedure that requires five passes of the image. The first four image passes are performed following the LRTB, LRBT, RLTB and RLBT directions, where L, R, T and B refer to Left, Right, Top and Bottom, respectively. In each pass i, we use as input the original image and get as output an inpainting image $P_i(x,y)$. In particular, for each "mask" pixel, we calculate the average of the "non-mask" pixels in the 4-connected neighbourhood (cross-type) and consider that pixel as a "non-mask pixel" for consecutive computations of that pass. At the final (fifth) image pass, for any direction chosen, we take into account all four images of the four previous inpainting passes ($P_i(x,y)$, $i = 1,\ldots,4$) and for each pixel we keep the minimum intensity. Example inpainting passes are shown in Fig. 4(c) and (d). In mathematical terms, let $I(x,y)$ be the original grayscale image with range [0,255] with 0 corresponding to foreground and 255 to background and $IM(x,y)$ be the bi-level inpainting mask with 0 corresponding to foreground and 1 to background. Then, the final inpainting result $BG(x,y)$ that corresponds to the estimated background image is produced by the Algorithm 1 for which the corresponding pseudocode is given. Example background estimation results are presented in Fig. 4(e) and (h).

Algorithm 1. The proposed inpainting procedure

I(x,y): Original Image, IM(x,y): Inpainting Mask
$P_i(x,y)$: Inpainting result of pass i, BG(x,y): Estimated Background
Ix, Iy: Image Width and Height
$x_{start}[4] = 0,0,Ix,Ix$, $x_{end}[4]=Ix,Ix,0,0$
$y_{start}[4] = 0,Iy,0,Iy$, $y_{end}[4]=Iy,0,Iy,0$
for $i = 1 \rightarrow 4$ **do**
  $M = IM$
  for $y = y_{start}[i] \rightarrow y_{end}[i]$ **do**
    for $x = x_{start}[i] \rightarrow x_{end}[i]$ **do**
      if M(x,y) = 0 **then**
        $P_i(x,y)$ = Average (I(x − 1,y)·M(x − 1,y),I(x,y − 1)·M(x,y − 1),
        I(x + 1,y)·M(x + 1,y),I(x,y + 1)·M(x,y + 1))
        I(x,y) = $P_i(x,y)$
        M(x,y) = 1
      end if
    end for
  end for
end for
for $y = y_{start}[1] \rightarrow y_{end}[1]$ **do**
  for $x = x_{start}[1] \rightarrow x_{end}[1]$ **do**
    BG(x,y) = min($P_i(x,y)$), $i = 1,\ldots,4$
  end for
end for

It is worth mentioning, that the proposed inpainting procedure was developed aiming in improved performance on subsequent steps. In particular, by keeping the minimum value among the first four inpainting passes P_i, more background noise is erased when global Otsu is applied on the normalized image, while textual content is better separated from interfering noise from borders of large stains (or large background variations in general) when Niblack is applied. In Section 3, experiments demonstrate the effectiveness of the proposed inpainting procedure.

2.2. Image Normalization

Background estimation followed by the image normalization procedure are often used to balance the illumination of a picture that is taken under uneven lighting conditions. This method modifies the image towards a bi-modal distribution that can be globally binarized with improved success. The main principle of the aforementioned method is that the observed image $I(x,y)$ equals to the product of the illumination (scene lighting conditions) $S(x,y)$ and the reflectance $R(x,y)$ of the picture objects. If the lighting conditions can be reproduced without any scene objects $I'(x,y)$, then the normalized image equals to $I(x,y)/I'(x,y)$.

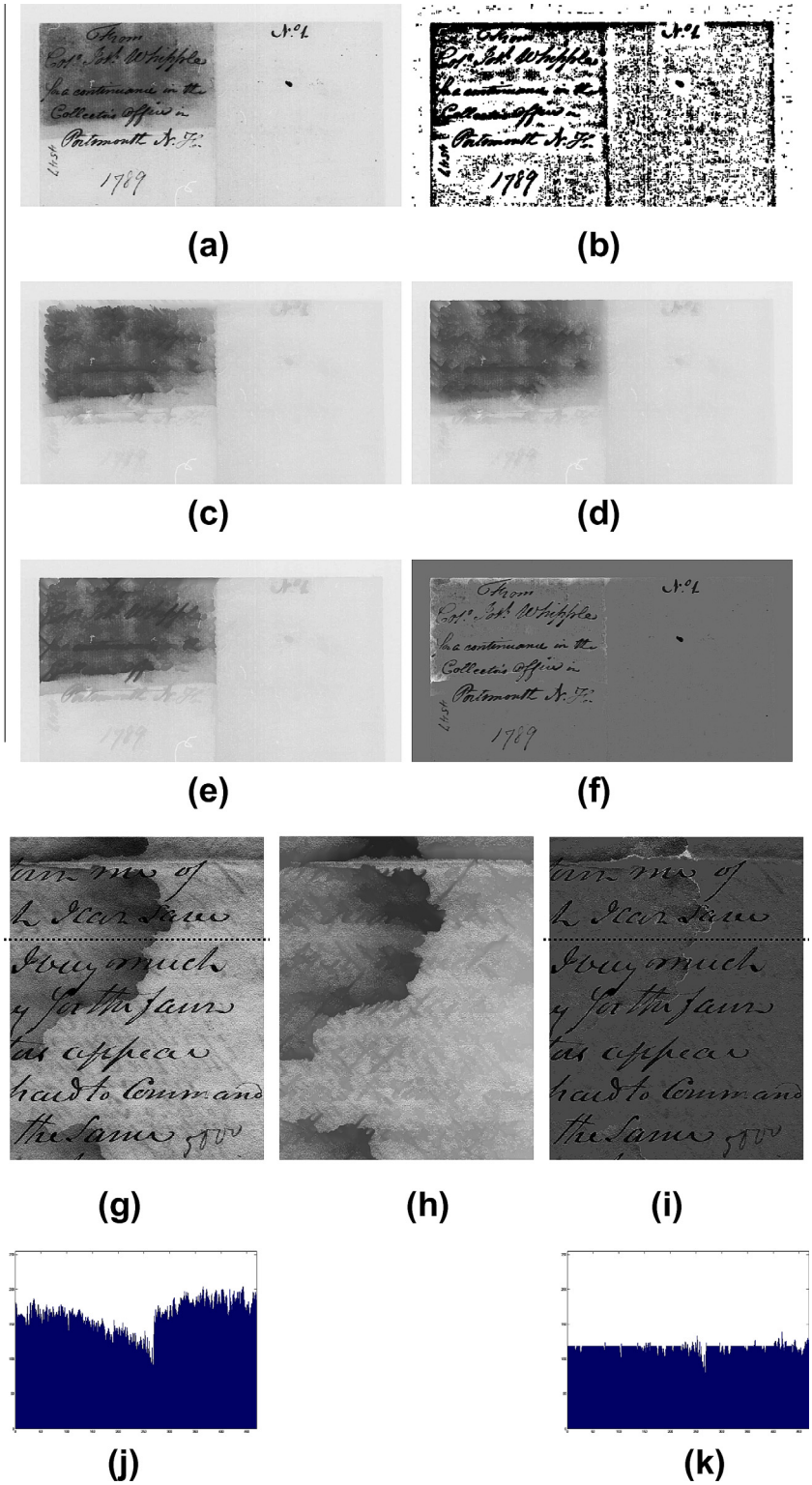

Fig. 4. Examples of Background Estimation and Image Normalization: (a) original image; (b) dilated Niblack, inpainting mask; (c) and (d) the LRTB and RLBT inpainting passes, respectively; (e) estimated background; (f) normalized image; (g) original image; (h) estimated background; (i) normalized image; (j) and (k) the intensity profile along the horizontal dashed black line of (g) and (i), respectively.

In the case of document images, we consider that the image that corresponds to the lighting conditions without any scene objects I'(x,y) is represented by the estimated background image BG(x,y). Then, according to the aforementioned discussion, the normalized image F(x,y) would be equal to I(x,y)/BG(x,y). The F(x,y) image is also normalized in the range between the minimum and maximum value of the original image Eq. (1), otherwise we could stretch the F(x,y) image to [0,255] but this leads to potential loss of faint characters and some noise remains.

$$N(x,y) = \left[(I_{max} - I_{min}) \cdot \frac{F(x,y) - F_{min}}{F_{max} - F_{min}} + I_{min} \right] \quad (1)$$

where: $F(x,y) = \frac{I(x,y)+1}{BG(x,y)+1}$, in order to be theoretically correct and eliminate any possibility of division by zero.

I_{min}, I_{max} and F_{min}, F_{max} denote the minimum and maximum values of I(x,y) and F(x,y) images, respectively.

Example normalization results are presented in Fig. 4(f) and (i), while in Fig. 4(j)–(k) the background flattening produced by the image normalization is shown.

2.3. Global binarization of the normalized image and post-processing

Given that the original image has been normalized using background estimation, we can efficiently apply the global binarization method of Otsu on the N(x,y) image. In this way, we detect the main text of the image, from which we can calculate document image features that are required in consecutive steps (e.g. the character stroke width and the image contrast). However, it is important to remove small noisy components to minimize the noise contribution in our calculations and to decrease the false combination when Otsu is combined at connected component level with the Niblack binarization result.

We consider that the aforementioned noise consist of several components corresponding to a small part of the detected foreground. Hence, the ratio between the total number of pixels that correspond to components of a small height and the number of those components, is expected to be very low. In particular, we consider O(x,y) to be the Otsu binarization result with "1" corresponding to foreground and "0" to background. Furthermore, let O_i(x,y) be the i-th connected component of O(x,y) with $i = 1,\ldots, NC_O$, where NC_O is the number of the connected components of O(x,y) and O_i(x,y) equals to "1" only at foreground pixels of the i-th connected component and "0" elsewhere. Then, we discard all connected components smaller than height h, as determined by the criterion specified by Eq. (2).

$$\sum_{j=1}^{h} \frac{RP^j}{RC^j} > 1 \tag{2}$$

where:

h denotes the minimum connected component height for which the above relation Eq. (2) is satisfied.

$RP^j = \frac{\sum_{x=1}^{lx}\sum_{y=1}^{ly} O^j(x,y)}{\sum_{x=1}^{lx}\sum_{y=1}^{ly} O(x,y)}$, denotes the ratio between the number of foreground pixels of Otsu connected components of height j, (i.e. $O^j(x,y)$) and the total number of foreground pixels of Otsu.

$RC^j = \frac{NC^j}{NC_O}$, denotes the ratio between the number NC^j of the connected components of height j and the total number NC_O of connected components of Otsu.

Based upon the aforementioned post-processing criterion, the Otsu image OP(x,y) after post-processing can be defined by Eq. (3) as follows.

$$OP(x,y) = \bigcup_{i=1}^{NC_O} O_i(x,y), \forall i : H(i) \geqslant h \tag{3}$$

where h is given by Eq. (2) and $H(i)$ denotes the height of the i-th connected component of Otsu O_i(x,y).

Example result of Otsu accompanied by the aforementioned post-processing OP(x,y) is presented in Fig. 5 in which the histogram (original values are multiplied by 100 for better visualization) of the CC height versus the post-processing criterion is demonstrated as well.

2.4. Local binarization of the normalized image

As mentioned in the previous section, the post-processing result of Otsu OP(x,y) is a good estimation of the text and can be used to detect the stroke width of the characters as well as the average value and the standard deviation of the text. Firstly, the post-processing result of Otsu is skeletonized S(x,y) using the technique of Lee and Chen (1992). The stroke width is adaptively detected based on the method described in (Ntirogiannis et al., 2009) which uses the skeleton and the corresponding contour points. In more details, for each skeleton point S(x,y), the smallest distance D from the contour points is computed and the corresponding stroke width that equals to 2D + 1 is assigned to that skeleton point ($S^{sw}(x,y)$). Afterwards, for each connected component j of the skeleton we consider the maximum of the stroke width values of the corresponding skeleton points ($S_j^{sw}(x,y)$) and the final stroke width of the image SW gets the average stroke width of all connected components of the skeleton Eq. (4).

$$SW = Average(SW_j), \quad j = 1,2,\ldots,\#\ skeleton\ CCs \tag{4}$$

where $SW_j = Max(S_j^{sw}(x,y))$.

In the proposed method, the skeleton S(x,y) is also used for the calculation of the average value and the standard deviation of the foreground that are required for the image contrast computation. The main reason for using the skeleton instead of the whole binarization result is that pixels near the contour of the binarization result may belong to transition pixels or even to background pixels and hence bias our calculations. However, skeleton points have much greater probability of belonging to foreground pixels. Hence, we calculate the average foreground intensity $FG_{average}$ and the corresponding standard deviation FG_{std} using the pixels of the original image that correspond to the skeleton pixels S(x,y), while the corresponding values (average and standard deviation) of the background are calculated using a modified estimated background image BG'(x,y). The estimated background image BG(x,y) was calculated according to the minimum of the 4 directional inpainting passes to achieve better normalization and consequently better binarization results (see Section 2.1). At this stage the average value of the four directional inpainting passes is considered ($BG'(x,y) = Average(P_i(x,y)), i = 1,..,4$) for better statistic results.

Image contrast is determined by the difference among the image foreground and background intensity. In (Su et al., 2010), the local contrast was computed taking into account the minimum and maximum intensity within a neighbourhood following the contrast formula of Michelson (1927). In (Su et al., 2011), the same authors used the maximum and the central intensity of the local neighbourhood to compute the local contrast following the "Weber" rule for contrast. Concerning the global contrast of the image, the contrast ratio is often used, i.e. the ratio between the foreground and the background intensity, and the logarithmic (log) of the contrast ratio has also been used (e.g. Spillmann and Levine, 1971) including document image applications (Roufs and Boschman, 1997). In the proposed method, we use the log contrast ratio Eq. (5) modified for degraded document images. As numerator, we use the average foreground intensity $FG_{average}$ increased by the foreground standard deviation FG_{std} in order to be closer to the faint character's intensity and as denominator we use the average background intensity $BG'_{average}$ decreased by the background standard deviation BG'_{std} in order to be closer to the background variations such as shadows and stains. In Eq. (5), the constant "-50" is used to limit the values between 0–100, approximately. Specifically, excluding the case in which all foreground pixels have 0 intensity and the all background pixels have 255 intensity, then in the extreme case of having numerator $FG_{average} + FG_{std} = 2.5$ and denominator $BG'_{average} - BG'_{std} = 252.5$, the log ratio equals to -2 and hence contrast C equals to 100 in this case. From our experiments that were conducted using historical images with various degradations, the values of contrast C were between 3–40 approximately. For instance, the contrast C of Figs. 1(c), 4(g) and 5(a) are 38.9, 14 and 3.4, respectively; while the corresponding SW values are 7, 4 and 8, respectively.

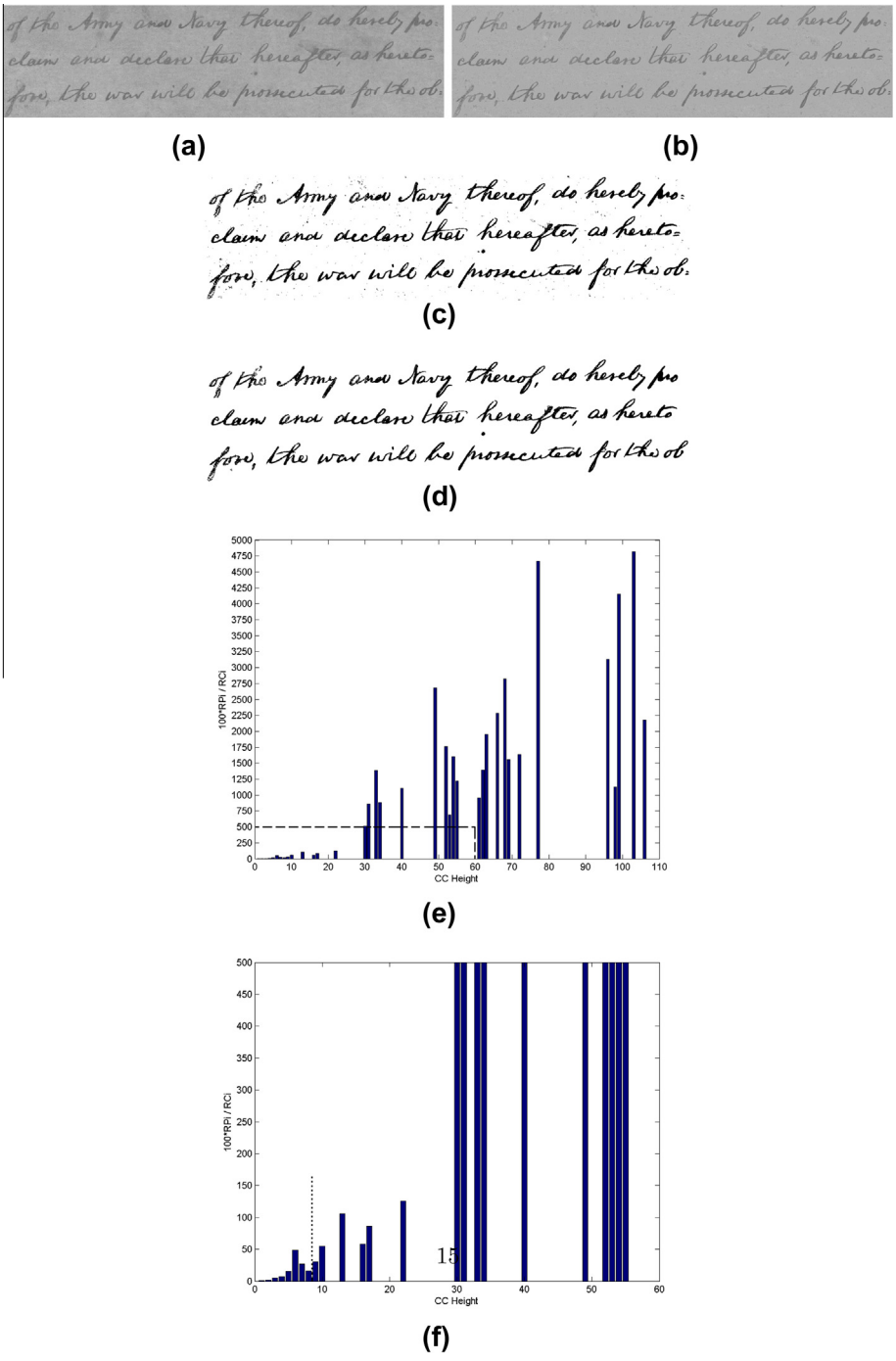

Fig. 5. Otsu binarization after normalization and post-processing: (a) Original image; (b) normalized image; (c) Otsu binarization; (d) post-processing; (e) histogram of 5(c); (f) focus on the dashed box of 5(e), the dashed line indicates the CC Height for which components were excluded.

$$C = -50 \cdot log10\left(\frac{FG_{average} + FG_{std}}{BG'_{average} - BG'_{std}}\right) \quad (5)$$

$$k = -0.2 - 0.1 \cdot \left\lfloor \left(\frac{C}{10}\right) \right\rfloor \quad (6)$$

The Niblack parameters which are the window size w and the "k" are set according to the Stroke Width of the characters SW and the image contrast C. Specifically, w equals 2SW and k (which is set to -0.2 by default) is proportional to the image contrast C but quantized according to Eq. (6) for better faint character detection. The constant window size of Niblack is suitable for handwritten documents which contain small variations in stroke width sizes and for fast algorithmic implementation.

2.5. Combination of the global and local binarization results

The result of the Otsu binarization on the proposed normalized image followed by the post-processing step contains low background noise but fails to retain the faint parts of the characters. On the contrary, the Niblack result contains much background noise but detects the faint characters efficiently enough. In the

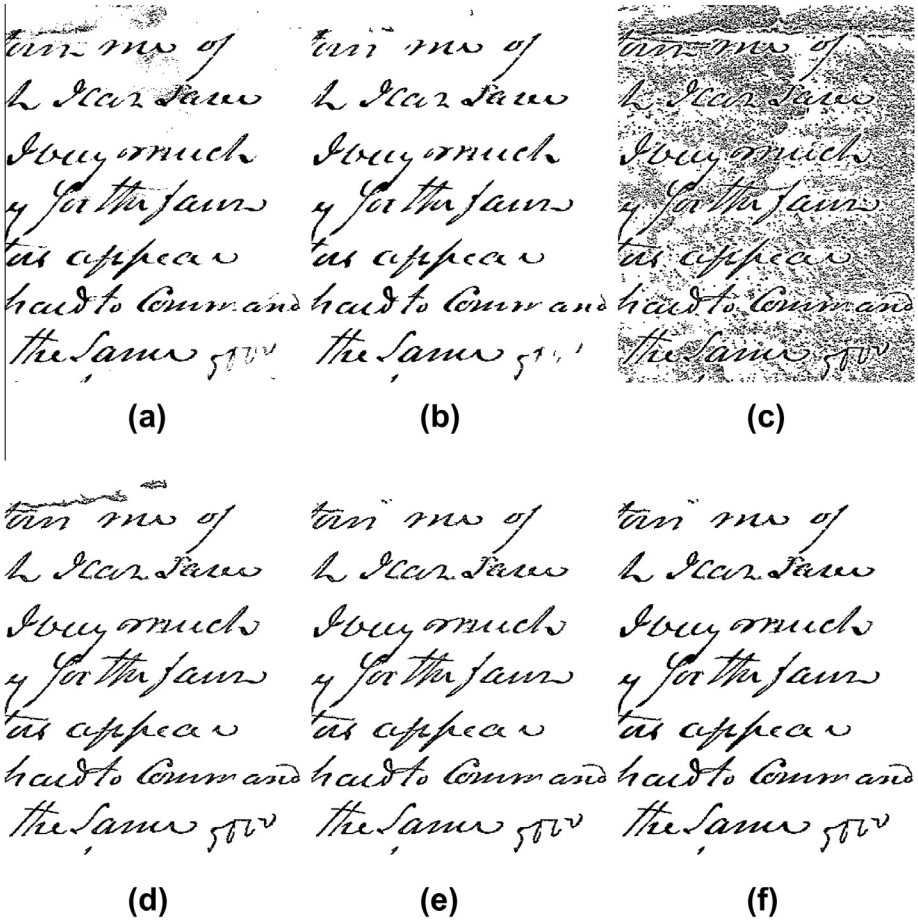

Fig. 6. Combination of global and local binarization results of 4(g): (a) Otsu of 4(i); (b) post-processing of (a); (c) Niblack of 4(i); (d) and (e) combination of (b) and (c) at connected component level without and with the proposed indirect post-processing, respectively; (f) final result after image enhancement using (a).

proposed method we combine the aforementioned Otsu result with the Niblack result, taking into account that Niblack achieves high text-to-background separation rate when applied on the proposed normalized image N(x,y) and under the correct w, k. In this respect, we aim in recovering the faint parts of the aforementioned Otsu result while retaining low levels of background noise.

In particular, we keep every connected component of Niblack that has common foreground pixels with the Otsu binarization result after post-processing OP(x,y). However, Niblack's components are excluded if they are covered only by a small percent of Otsu pixels. Formally, let $NB_j(x,y)$ be the j-th connected component of the Niblack result with "1" at the foreground pixels of the j-th

Table 1
Comparison with the top-3 binarization methods using the Handwritten images of the DIBCO'09, H-DIBCO'10 and DIBCO'11 contests, followed by the individual score per image for DIBCO'11.

	Method	Score	Measures						
			FM	PSNR	NRM	MPM	p-FM	DRD	
DIBCO'09	3rd	36	86.16	18.32	6.25	1.54	-	-	
	2nd	31	86.02	18.57	6.39	0.95	-	-	
	1st	10	88.65	19.42	5.11	0.34	-	-	
	Proposed	7	92.64	21.28	2.84	0.48	-	-	
HDIBCO'10	3rd	23	91.78	19.67	4.77	1.33	94.43	-	
	2nd	21	91.50	19.78	5.98	0.49	93.58	-	
	1st	19	89.70	19.15	8.18	0.29	95.15	-	
	Proposed	9	94.34	21.60	3.04	0.32	94.14	-	
DIBCO'11	3rd	233	87.18	17.76	-	2.94	-	4.14	
	2nd	166	88.74	18.76	-	5.47	-	4.01	
	1st	85	92.38	19.93	-	2.04	-	2.36	
	Proposed	73	94.05	21.65	-	4.93	-	2.60	
DIBCO'11 Details		*Images*							
		HW1	HW2	HW3	HW4	HW5	HW6	HW7	HW8
		Score per Image							
	3rd	20	45	32	32	16	14	39	35
	2nd	45	24	14	15	10	19	10	29
	1st	7	23	14	10	8	4	9	10
	Proposed	15	4	4	4	13	19	8	6

connected component and "0" elsewhere. Then, the combination result CO(x,y) contains all NB_j components that satisfy Eq. (7). In other words, indirect post-processing is performed since large Niblack components that correspond to just a few Otsu pixels are excluded from the combination. Moreover, the image contrast C is closely related to the level of background noise that remains after the global Otsu step, since for bleed-through images that usually have high contrast C, more background noise remains, even though the background estimation and normalization step. Example binarization combination result is presented in Fig. 6 with and without the proposed indirect post-processing.

$$CO(x,y) = \bigcup_{j=1}^{n} NB_j(x,y), \quad \forall j : d(j) = true \tag{7}$$

$$\text{where } d(j) = \begin{cases} true, \text{if} 100 \dfrac{\sum_{x=1}^{lx}\sum_{y=1}^{ly} OP(x,y)\cdot NB_j(x,y)}{\sum_{x=1}^{lx}\sum_{y=1}^{ly} NB_j(x,y)} \geqslant C \\ false, \text{otherwise} \end{cases}$$

It is worth mentioning that the components of Niblack that we kept may contain some artefacts due to the image normalization and the constant window size. Hence, we enhance the binarization result CO(x,y), by using the result of Otsu. In more details, in the final binarization result FB(x,y) Eq. (8) we include all pixels of Otsu O(x,y) for which a pixel of the combination result image CO(x,y) can be detected in the 8-connected neighbourhood (Fig. 6(f)).

$$FB(x,y) = CO(x,y) \cup (O(x,y)\cdot f(x,y)) \tag{8}$$

$$\text{where } f(x,y) = \begin{cases} 1, \text{if} \sum_{m=x-1}^{x+1}\sum_{n=y-1}^{y+1} CO(m,n) > 0 \\ 0, \text{otherwise} \end{cases}$$

3. Experimental results

For the experiments we used the handwritten images of the DIBCO series which are publicly available, Gatos et al. (2011); Pratikakis et al. (2010); Pratikakis et al. (2011a). All images of the aforementioned contests were taken from the Library Of Congress (http://memory.loc.gov/ammem), contain various degradations such as shadows, non-uniform illumination, stains, smudges, bleed-through, faint characters, etc. and have various specifications, such as grayscale (8-bit) or colour (24-bit), jpeg or tiff format and their resolution range from 200 to 400 dpi, with the majority being at 300 dpi. Additionally, the first contest (DIBCO'09) contains five (5) handwritten images with forty-three (43) competing techniques while the second (H-DIBCO'10) and the third (DIBCO'11)

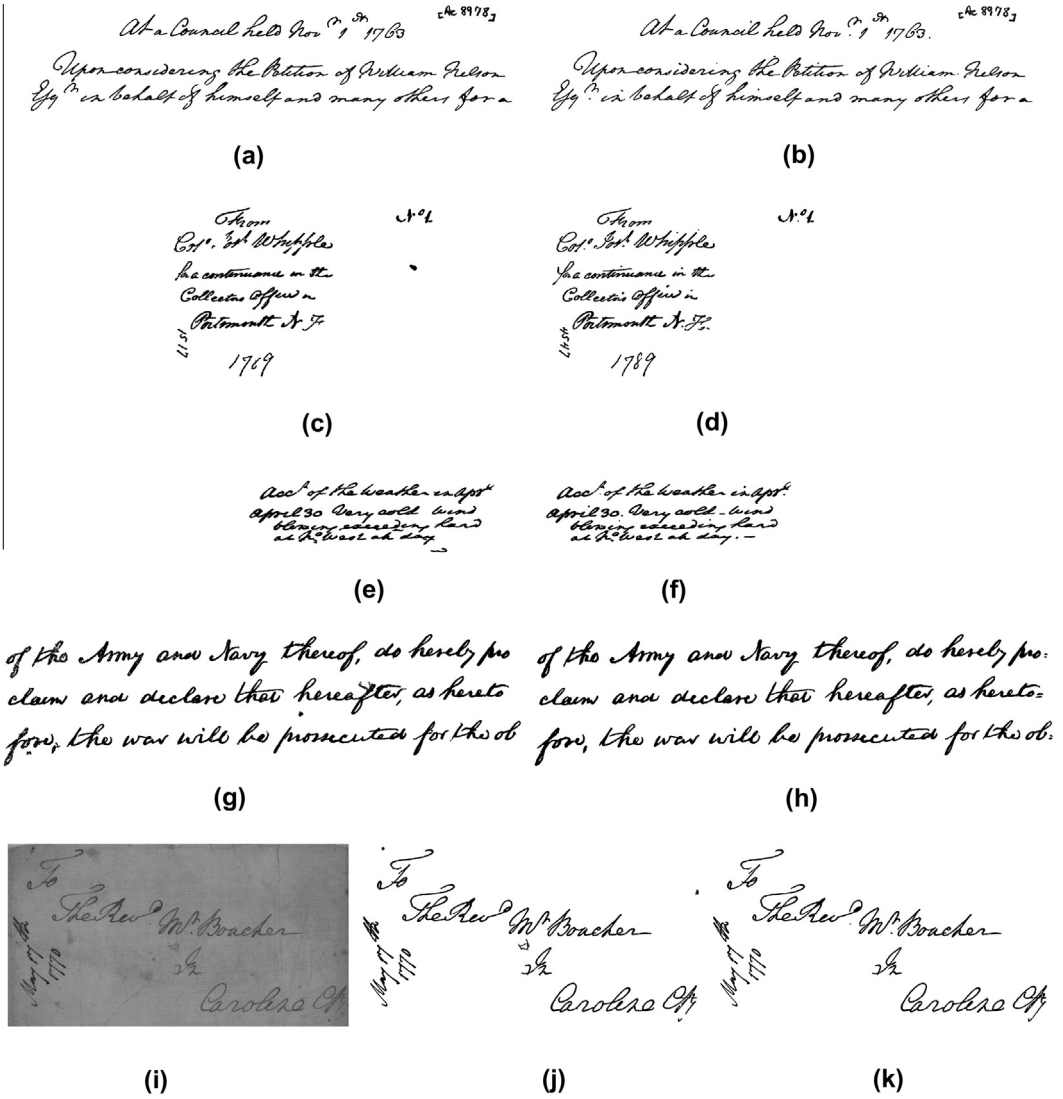

Fig. 7. Results of the proposed method on handwritten images of DIBCO'09 and HDIBCO'10 contests: (a); (c); (e); (g) proposed binarization of 1(a), 1(b), 1(c), 5(a), respectively; (b); (d); (f); (h) the corresponding ground truth; (i) original image; (j) proposed binarization; (k) ground truth.

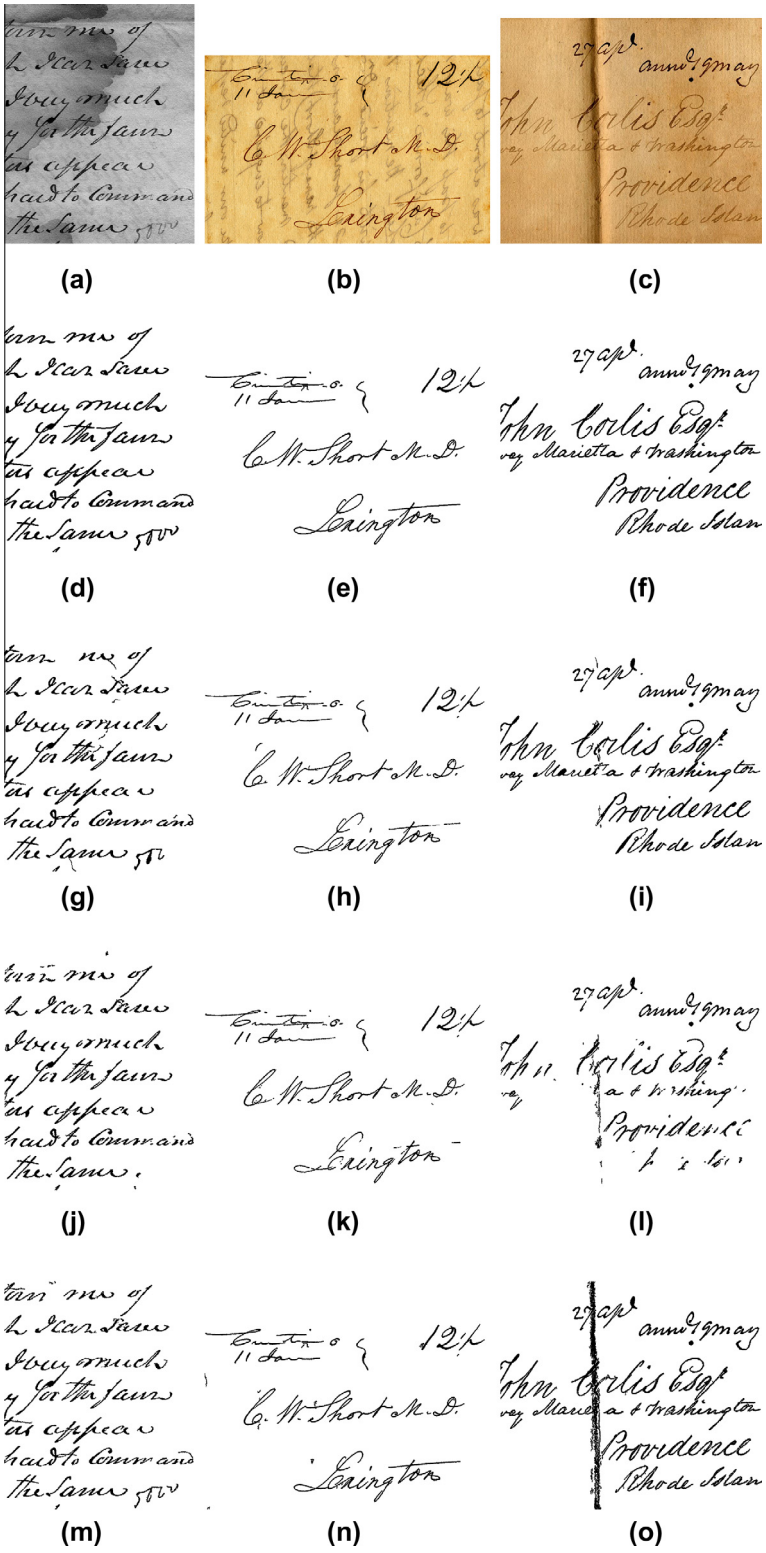

Fig. 8. Comparison with top methods of DIBCO'11 contest: (a)–(c) The original images of HW4, HW7 and HW6, respectively; (d)–(f) ground truth; (g)–(i) 1st (participant No. 10); (j)-(l) 2nd (participant No. 8); (m)–(o) proposed.

contests contain ten (10) and eight (8) handwritten images with seventeen (17) and eighteen (18) competing techniques, respectively. Furthermore, for the images used in the experiments, we measured the required processing time using a PC with dual processor at 2.2 GHz with 2 GB of RAM memory. The average processing time was 3.31 s and the average image size was 1231x562.

In the first document image binarization contest (DIBCO'09), the evaluation measures of FM (F-Measure), PSNR (Peak Signal to Noise Ratio), NRM (Negative Rate Metric) and MPM (Misclassification Penalty Metric) were used. According to Gatos et al. (2011), the average value for each evaluation measure (the average F-Measure was computed from the average Recall and average Precision) was used to compute the corresponding ranking and

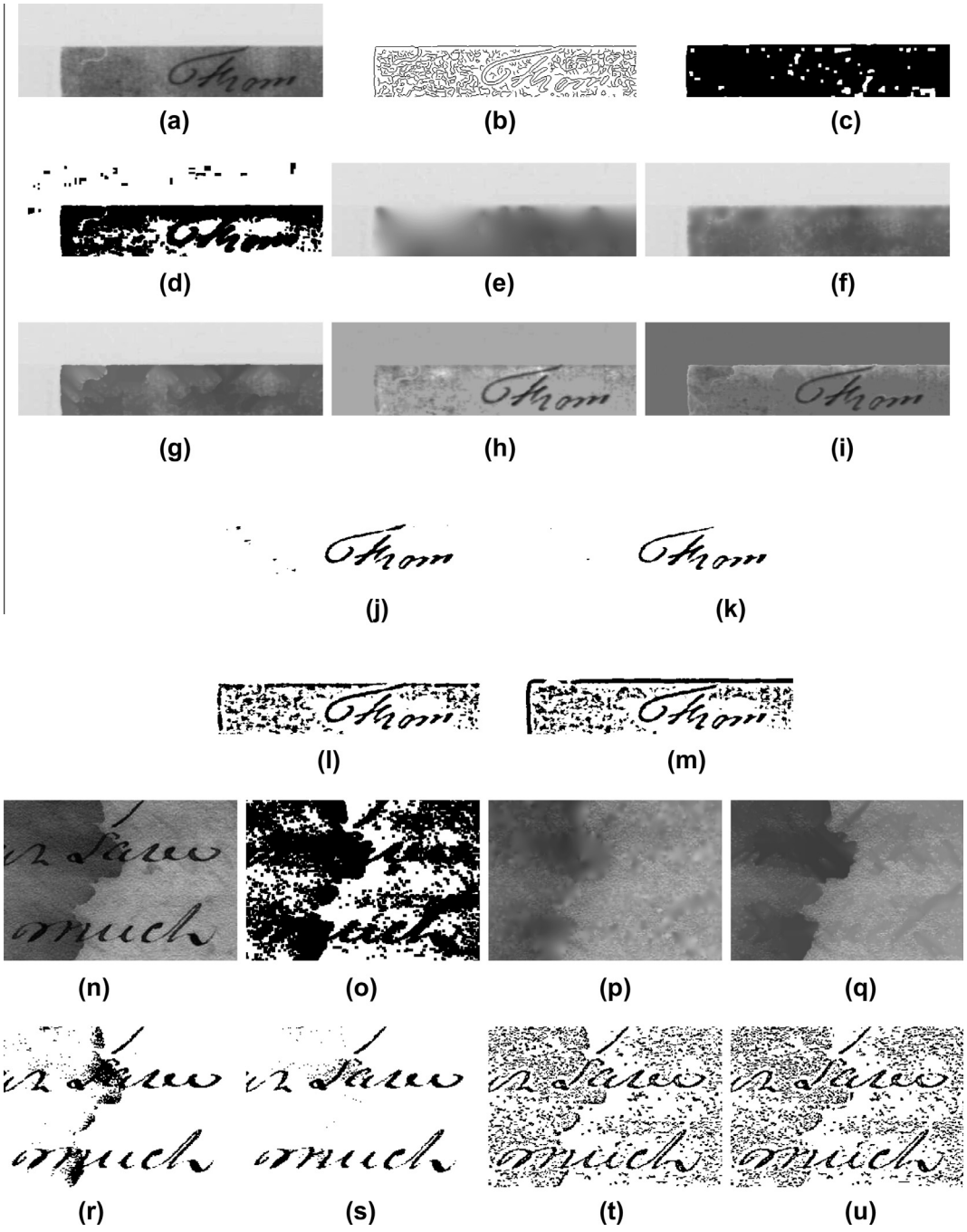

Fig. 9. Comparison with different inpainting approaches: (a) original image (focus on the upper-left corner of 4(a)); (b) Canny edges; (c) inpainting mask based on (b) (Zhang et al., 2009); (d) proposed inpainting mask; (e) and (f) D'Errico (2006) inpainting using (c) and (d), respectively; (g) proposed inpainting using (d); (h) and (i) image normalization of (f) and (g), respectively; (j) and (k) Otsu of (h) and (i), respectively; (l) and (m) Niblack ($w = 2SW = 16$, $k = -0.2$) of (h) and (i), respectively; (n) Original image (focus on 2nd-3rd lines of 4(g)); (o) proposed inpainting mask; (p) D'Errico (2006) inpainting using (o); (q) proposed inpainting using (o); (r) and (s) Otsu based on (p) and (q), respectively; (t) and (u) Niblack ($w = 8$, $k = -0.3$) based on (p) and (q), respectively.

the final Score was calculated using the summation of the ranking values Eq. (9). In the second contest (H-DIBCO'10), the performance measure of p-FM (p-FMeasure) was additionally used. In DIBCO'11 contest the evaluation measures of FM, PSNR, DRD (Distance Reciprocal dostortion) (Lu et al., 2004) and MPM were used. For the final Score (as described in (Pratikakis et al., 2011a)), the summation of the ranking values for every image and for all evaluation measures was used Eq. (10).

$$Score_{dibco09\&10} = \sum_{m=1}^{\#Measures} Rank_m \tag{9}$$

where $Rank_m$ denotes the ranking value based on the average value for all images when measure m is used.

$$Score_{dibco11} = \sum_{i=1}^{\#Images} \sum_{m=1}^{\#Measures} Rank_m^i \tag{10}$$

where $Rank_m^i$ denotes the ranking value concerning image i when measure m is used.

In Table 1, the detailed results of the top-3 methods along with the results of the proposed method are presented for all DIBCO series and representative examples are shown in Figs. 7 and 8. For DIBCO'11 contest concerning the handwritten images only, the top-3 methods are participants No. 10,8,7 (Pratikakis et al., 2011a). From Table 1, it is demonstrated that the proposed method achieves top performance among all participants in the DIBCO series datasets. There were further experimentations with the DIBCO series datasets reported in works that have not participated in DIBCO competition. For instance, high performance is reported in (Hedjam et al., 2011) with average FM of 89.27 concerning the handwritten document images of DIBCO'09 contest. However, the authors considered the presence of the bleed-through to be known. For the same dataset, Su et al. (2010) also reported high performance with FM, PSNR, NRM and MPM equal to 89.93, 19.94, 6,69 and 0.30, respectively. We also tested the combination methodology for binarization methods of Gatos et al. (2008) on the same dataset achieving performance of 80.19, 16.11, 2.01 and 4.85 concerning FM, PSNR, NRM and MPM, respectively. As far as the proposed method is concerned, the corresponding values are 92.63 (average FM calculated from mean F-Measure), 21.28, 2.84 and 0.48, that demonstrate higher performance on the majority of the evaluation measures.

A visual analysis of the results demonstrates that the proposed method can detect the majority of the faint characters and remove almost all the bleed-through (Fig. 7(e)-(f) appear cropped since the bleed-through was effectively removed from Fig. 7(e)). It is not affected by large background stains and it handles effectively documents of low contrast (Fig. 7(g)). Additionally, in the bottom part of Table 1 (DIBCO'11 Details) the individual score for each image of DIBCO'11 (HW1-HW8) is presented. Those images contain various degradations. Particularly, image HW1 contains border noise and is of high contrast, HW2 contain faint characters and low background texture, HW3 contain faint and slim characters and smudges, HW4 contains large stain, faint characters and low horizontal folding, HW5 contains smudges, HW6 contains faint characters and shadow from vertical folding, HW7 contains faint characters and bleed-through and HW8 contains faint characters. According the detailed results of the DIBCO'11 and example images of our method that contained faint characters (e.g. Fig. 7(a)), it is depicted that the proposed method achieves high performance in documents with faint characters which is a common degradation of handwritten documents.

In order to demonstrate the effectiveness of the proposed inpainting procedure (Section 2.1), we conducted experiments using the proposed inpainting mask along with a standard inpainting D'Errico (2006). The inpainting mask of Zhang et al. (2009) is based on Canny edges (Fig. 9(b)) followed by morphological dilation and closing operation (Fig. 9(c)). This mask yielded poor results, since in order to detect the majority of the edges that correspond to the faint characters, many background edges were also detected. In this respect, we used the proposed inpainting mask of the dilated Niblack that yielded better results (Fig. 9(e)–(f)). From Fig. 9(j)–(m) and 9(r)–(u), it is proved that the proposed inpainting procedure results in improved Otsu and Niblack binarization that consequently leads to more efficient combination, while other approaches fail to separate efficiently the textual content from any interfering background noise.

As demonstrated in this section, the proposed method achieves high performance in many cases, however very small components like dots and punctuation (e.g. the dot of the i and j) are discarded because of the intermediate post-processing step and large noisy components encountered on page splits (e.g. Fig. 8(o)) cannot be efficiently removed. Even though, in the case of HW6 (Fig. 8(o)), the proposed method has the same performance with the second best method concerning the overall performance on the handwritten documents (bottom part of Table 1). We have excluded large component manipulation from the final post-processing step because in handwritten document images, large components often correspond to textual information of several characters, words or even text-lines. Following a common post-processing step we could discard the large noisy components but with loss of textual information. Hence, a more sophisticated post-processing step is required to discard large noise from foldings or page splits without discarding valuable text information.

4. Conclusion

In this work we presented a binarization method based on several image processing steps that achieves high performance in a wide range of degraded handwritten documents. We proposed a goal-oriented inpainting procedure for background estimation and image normalization that assists the global and local binarization results of Otsu and Niblack to be combined effectively. Post-processing is performed both at an intermediate and the final step of our method. We did not include large component manipulation at the final post-processing step because handwritten document images may contain large components corresponding to text information of several characters, words or even text-lines. Although the proposed method could be improved to discard large noise from foldings or page splits without discarding valuable text information, it achieves top performance in the DIBCO series of 2009–2011.

References

Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C., 2000. Image inpainting. In Proceedings of the 27th annual conference on Computer graphics and interactive, techniques (SIGGRAPH'00), pp. 47–424.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Machine Intell. 8 (6), 679–698.

Chan, T., Shen, J., 2002. Mathematical models for local nontexture inpaintings. IEEE Trans. Pattern Anal. Machine Intell. 62 (3), 1019–1043.

D'Errico, J., 2006. Inpaintingnans. MATLAB Central File Excange. <http://www.mathworks.com/matlabcentral/fileexchange/4551>.

Gatos, B., Ntirogiannis, K., Pratikakis, I., 2011. DIBCO 2009: document image binarization contest. Internat. J. Document Anal. Recognition 14 (1), 35–44.

Gatos, B., Pratikakis, I., Perantonis, S.J., 2006. Adaptive degraded document image binarization. Pattern Recognition 39 (3), 317–327.

Gatos, B., Pratikakis, I., Perantonis, S.J., 2008. Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information. In Proc. 19th Internat. Conf. on, Pattern Recognition, pp. 1–4.

Hedjam, R., Moghaddam, R.F., Cheriet, M., 2011. A spatially adaptive statistical method for thebinarization of historical manuscripts and degraded document images. Pattern Recognition 44 (9), 2184–2196.

Howe, N.R., 2011. A laplacian energy for document binarization. In Proc. 11th Internat. Conf. Document Anal. Recognition, pp. 6–10.

Kim, I.K., Jung, D.W., Park, R.H., 2002. Document image binarization based on topographic analysis using a water flow model. Pattern Recognition 35 (1), 265–277.

Lee, H.J., Chen, B., 1992. Recognition of handwritten chinese characters via short line segments. Pattern Recognition 25 (5), 543–552.

Lu, H., Kot, A.C., Shi, Y.Q., 2004. Distance-reciprocal distortion measure for binary document images. IEEE Signal Process. Lett. 11 (2), 228–231.

Lu, S., Su, B., Tan, C.L., 2010. Document image binarization using background estimation and stroke edges. Internat. J. Document Anal. Recognition 13 (4), 303–314.

Messaoud, I.B., Amiri, H., Abed, H.E., Margner, V., 2011. New binarization approach based on text block extraction. In Proc. 11th Internat. Conf. on Document Anal. Recognition, pp. 1205–1209.

Michelson, A., 1927. Studies in Optics. University of Chicago Press.

Niblack, W., 1986. An Introduction to Digital Image Processing. Prentice-Hall, Englewood Cliffs, NJ, pp. 115–116.

Ntirogiannis, K., Gatos, B., Pratikakis, I., 2008. An objective evaluation methodology for document image binarization techniques. In Proc. 8th IAPR Internat. Workshop on Document Anal. Systems, pp. 217–224.

Ntirogiannis, K., Gatos, B., Pratikakis, I., 2009. A modified adaptive logical level binarization technique for historical document images. In Proc. 10th Internat. Conf. on Document Anal. Recognition, pp. 1171–1175.

Otsu, N., 1979. A thresholding selection method from gray-level histogram. IEEE Trans. Systems Man Cybernet. 9 (1), 62–66.
Pratikakis, I., 2011b. DIBCO 2011 binarization results, <http://utopia.duth.gr/~ipratika/DIBCO2011/dibco2011results.htm/>.
Pratikakis, I., Gatos, B., Ntirogiannis, K., 2010. H-DIBCO 2010 handwritten document image binarization competition. In Proc. 12th Internat. Conf. on Frontiers in Handwriting Recognition, pp. 727–732.
Pratikakis, I., Gatos, B., Ntirogiannis, K., 2011a. ICDAR 2011 document image binarization contest (DIBCO 2011). In Proc. 11th Internat. Conf. Document Anal. Recognition, pp. 1506–1510.
Roufs, J.A.J., Boschman, M.C., 1997. Text quality metrics for visual display units: I. Methodological aspects. Displays 18 (1), 37–43.
Sauvola, J., Pietikainen, M., 2000. Adaptive document image binarization. Pattern Recognition 33 (2), 225–236.
Spillmann, L., Levine, J., 1971. Contrast enhancement in a hermann grid with variable figure-ground ratio. Exp. Brain Res. 13 (5), 547–559.
Su, B., Lu, S., Tan, C.L., 2010. Binarization of historical document images using the local maximum and minumim. In Proc. 9th IAPR Internat. Workshop on Document Anal. Systems, pp. 303–314.
Su, B., Lu, S., Tan, C.L., 2011. Combination of document image binarization techniques. In Proc. 11th Internat. Conf. on Document Anal. Recognition, pp. 22–26.
Valizadeh, M., Kabir, E., 2012. Binarization of degraded document image based on feature space partitioning and classification. Internat. J. Document Anal. Recognition 15 (1), 57–69.
van Herk, M., 1992. A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels. Pattern Recognition Lett. 13 (7), 517–521.
Yang, Y., Yan, H., 2000. An adaptive logical method for binarization of degraded document images. Pattern Recognition 33 (5), 787–807.
Zhang, L., Yip, A.M., Brown, M.S., Tan, C.L., 2009. A unified framework for document restoration using inpainting and shape-from-shading. Pattern Recognition 42 (11), 2961–2978.