

# Degraded document image binarization using structural symmetry of strokes



Fuxi Jia, Cunzhao Shi\*, Kun He, Chunheng Wang, Baihua Xiao

The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences, No.95 Zhongguancun East Road, Beijing 100190, PR China

## ARTICLE INFO

### Article history:

Received 19 December 2016

Revised 8 June 2017

Accepted 19 September 2017

Available online 23 September 2017

### Keywords:

Document image binarization

Structural symmetry of strokes

Local threshold

Stroke width estimation

## ABSTRACT

This paper presents an effective approach for the local threshold binarization of degraded document images. We utilize the structural symmetric pixels (SSPs) to calculate the local threshold in neighborhood and the voting result of multiple thresholds will determine whether one pixel belongs to the foreground or not. The SSPs are defined as the pixels around strokes whose gradient magnitudes are large enough and orientations are symmetric opposite. The compensated gradient map is used to extract the SSP so as to weaken the influence of document degradations. To extract SSP candidates with large magnitudes and distinguish the faint characters and bleed-through background, we propose an adaptive global threshold selection algorithm. To further extract pixels with opposite orientations, an iterative stroke width estimation algorithm is applied to ensure the proper size of neighborhood used in orientation judgement. At last, we present a multiple threshold vote based framework to deal with some inaccurate detections of SSP. The experimental results on seven public document image binarization datasets show that our method is accurate and robust compared with many traditional and state-of-the-art document binarization approaches based on multiple evaluation measures.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

Document image binarization is a fundamental step in most document analysis systems, which aims to extract text objects from the background [1,2]. The performance of subsequent steps is highly dependent on the success of binarization. But there are still many challenges when the document images contain various degradations, such as faint characters, bleed-through background, ink stains and so on. Therefore, the study on binarization for document images, in particular degraded images, is very essential.

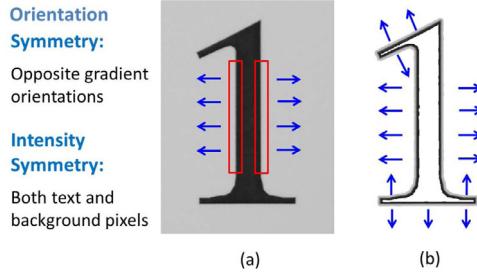
A successful binary result preserves meaningful information while discarding noisy information [1]. It is difficult to get such a successful result when we apply the traditional local thresholding methods [3,4] to binarize the degraded document images. The reason might lie in the fact that they compute a unique threshold using all the pixels in neighborhood including the possible random noise and background disturbance. In this paper, we use multiple threshold values computed by the SSPs of the region to find out whether one pixel belongs to the foreground or not. As shown in Fig. 1, the SSPs denote the stroke edges which contain both text

and background pixels. The intensity statistic of these pixels is a good approximation of the local threshold used to distinguish text from background.

The concept of SSP has first been proposed in our previous work [5]. Its effectiveness has already shown in that paper. However, since the original SSP extraction method is primitive and parameter-dependent, the binarization result is not always satisfied. In this paper, we present a modified extraction algorithm with minimum parameter tuning to extract the SSP more precisely. Specifically, the improvements are threefold. First, we use a more effective background removal process to obtain the compensated image. The gradient map is produced by the compensated one so as to deal with many types of degradations. Then we propose an adaptive threshold selection algorithm to compute a global threshold adaptively. We use this threshold to binarize the gradient magnitude map in order to extract the SSP candidates with large magnitude precisely. Finally, we adopt an iterative algorithm to estimate the text stroke widths and remove noise at the same time. By applying the orientation symmetry judgement based on the estimated stroke widths on the SSP candidates, we extract the real SSPs satisfying opposite orientation constraint in neighborhood. The other contribution of this paper is the voting framework in which we use multiple threshold values to decide whether each pixel belongs to text or not. In this way, some inaccurate detection

\* Corresponding author.

E-mail addresses: [cunzhao.shi@ia.ac.cn](mailto:cunzhao.shi@ia.ac.cn) (C. Shi), [chunheng.wang@ia.ac.cn](mailto:chunheng.wang@ia.ac.cn) (C. Wang).



**Fig. 1.** The illustration of structural symmetric pixels (SSPs). (a) The motivation of SSP : pixels around strokes contain both text and background candidates. (b) SSP (white pixels represent Non-SSP, black and gray pixels denote text and background candidates respectively). The blue arrows denote the gradient orientations of stroke edges. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of SSP can be compensated. The executable code of our algorithm can be download from the following url<sup>1</sup>.

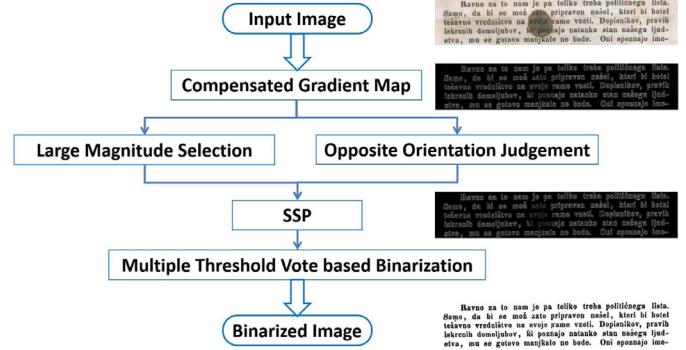
To demonstrate the effectiveness of the proposed method, we conduct comprehensive experiments on different datasets with various types of degradations. The experimental results show that our method achieves promising performance compared with many traditional and state-of-the-art document binarization algorithms tested on the datasets of DIBCO'09 [6], H-DIBCO'10 [7], DIBCO'11 [8], H-DIBCO'12 [9], DIBCO'13 [10], H-DIBCO'14 [11] and H-DIBCO'16 [12], based on various evaluation measures, including the F-measure, pseudo F-Measure, NRM, PSNR, DRD, and MPM.

The remainder of this paper is organized as follows. Section 2 briefly reviews the related work. In Section 3, we introduce the proposed binarization method using the structural symmetry of strokes in detail. The experimental results are presented in Section 4, and dummyTXdummy- conclusions are drawn in Section 5.

## 2. Related work

Document image binarization is to convert a color or gray image into a binary image, where the text and background pixels are marked in black and white respectively. Basically, the thresholding technique, which is one of the most useful binarization methods, are of three main types: global, local and hybrid [13,14]. In the global techniques, a single threshold value is calculated from whole image. The local techniques compute a local threshold based on the statistics in the neighborhood of each pixel. Hybrid method is the combination of the local and global one. For document images of good quality, global methods, like Otsu [15] and Kittler [16], is capable of extracting the text efficiently. However, for document images suffering from different types of degradations, local methods, such as Bernsen [17], Niblack [3] and Sauvola [4], usually produce better binarization results [18]. But they tend to introduce some background noise. The reason may lie in the fact that the local threshold is calculated by all pixels in neighborhood including the background disturbance.

To obtain more satisfactory binary results, Lu et al. [19] compute the local threshold only based on the stroke edge pixels and the performance is improved to some extent. To extract the stroke edges, they first estimate the background surface through an one-dimensional iterative polynomial smoothing procedure [20] and then use Otsu's method to binarize the compensated gradient map. While the gradient map is replaced with a local contrast image built with local maximum and minimum in Su's paper [21]. In [22], a more robust feature map is produced by combining the gradient



**Fig. 2.** The flowchart of the proposed binarization method.

map with the local contrast image. This method performs better than those in [19] and [21]. But these methods have a certain limitation as stated in [19]: the result may remain some bleed-through noise or ignore some faint characters since the final threshold is based on the local contrast. Lelore et al. [23] introduce the FAIR binarization algorithm based on a double-threshold edge detection approach and the detection strategy makes it possible to catch small details while remaining robust against noise.

The hybrid techniques combine the global and local thresholding techniques. It takes the advantages of both the techniques. Chou et al. [24] divide an input image into blocks and choose different binarization methods for each block. In [25], several methods are combined based on a vote on their outputs. In [26], the background is estimated first by performing inpainting, then a combination of the global and local adaptive binarization method at connected component level is proposed to binarize images. The performance of this method is extremely well but it is limited to binarize handwritten document images only.

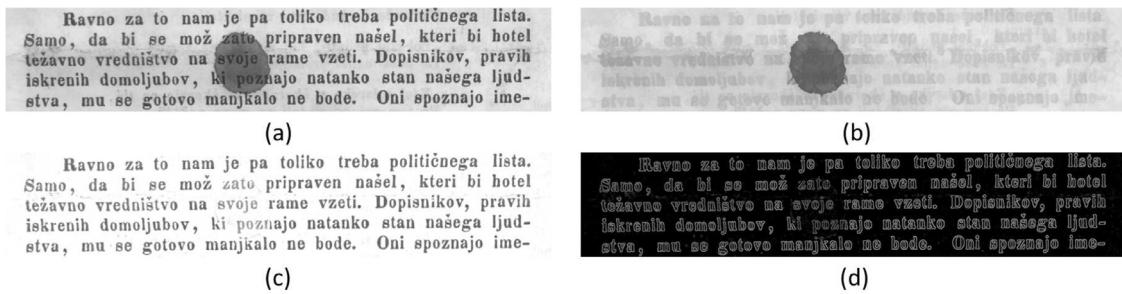
Other non-threshold approaches have been reported and the results are promising. Howe [27] proposes a method based on the Laplacian energy of the image intensity. The energy function is minimized via a graph-cut computation. The method is efficient but parameter-dependent. In [28], Howe improves this method by tuning two key parameters adaptively and the performance is improved. Mishra et al. [29] define a energy function so that the quality of the binarization is inversely related to the energy value. They minimize this energy function to find the optimal binarization using an iterative graph cut scheme. In some literatures [30–33], the input images are divided into three classes: foreground, background, and uncertain. Then, they classify those uncertain pixels by applying the MRF model or other strategy on the other two categories of pixels. In a study published recently [34], some reasonable pre and post processes are used to deal with the broken and degraded document images. These methods which combine different types of image information and domain knowledge usually have high computational complexity.

## 3. Proposed method

This section details the proposed document image binarization method.

The flowchart is shown in Fig. 2. Given a document image, firstly, we compute the gradient map from the compensated image which is obtained through a background removal process. Then the SSP in the gradient map will be extracted from two different aspects: large magnitude selection using adaptive gradient binarization method and opposite orientation judgement based on the stroke width estimation. Finally, we use the multiple threshold vote based framework to decide whether each pixel belongs to text or not. The concrete procedures will be provided below.

<sup>1</sup> <https://github.com/Fuxijia/DocumentBinarizationSSP>.



**Fig. 3.** An example of compensated gradient map production: (a) Original image: PR5 from DIBCO'13. (b) Estimated background. (c) Normalized image. (d) Compensated gradient map.

### 3.1. Compensated gradient map

We produce the compensated gradient map in order to deal with the different types of document degradations such as faint characters, bleed-through background and ink stains. The gradient map is our basis to extract the SSPs which are defined as pixels with large gradient magnitudes and opposite gradient orientations. But the gradient map produced by the original image may introduce some unnecessary noise for the degraded document images. Therefore, we estimate the background surface and compensate the variation of degradation to obtain an appropriate gradient map. Particularly, three procedures will be implemented as follows.

Firstly, the background surface is estimated by the background estimation algorithm of Ntirogiannis [26]. The underlying idea is to perform inpainting using Niblack's [3] binarization result as the inpainting mask. Fig. 3(b) shows an example of the estimated background surface.

The document degradation compensation is performed as follows:

$$I_{\text{norm}}(x, y) = \begin{cases} 255 \times \frac{I(x, y)}{B(x, y)} & I(x, y) < B(x, y), B(x, y) > 0 \\ 255 & \text{else} \end{cases} \quad (1)$$

where  $I$  and  $B$  denote the original image and the estimated background image respectively. Fig. 3(c) shows the normalized image corresponding to the image Fig. 3(a). As we can see, the low contrast stroke edge pixels caused by the ink stain of the original image are restored in the normalized image thanks to the background compensation process.

In order to deal with the random noise, we compute the gradient of the normalized image  $I_{\text{norm}}$  through a deformation of Sobel operator. In practical experiments, it is better than the Sobel operator in our binarization framework. The operator combines the Gauss property with the original Sobel characteristic. Eq. (2) shows its concrete forms in the horizontal and vertical directions. Then the gradient magnitude map  $G$  is computed approximately by adding the L1-norm image gradient in the two directions for reducing the computational complexity. We use the coefficient  $c$  in Eq. (3) to stretch the intensity range of the gradient magnitude map  $G$  to  $[0, 255]$ .

$$G_x = \begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix} \quad G_y = \begin{bmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{bmatrix} \quad (2)$$

$$G(x, y) = (|G_x * I_{\text{norm}}(x, y)| + |G_y * I_{\text{norm}}(x, y)|) \times c \quad (3)$$

### 3.2. SSP extraction

For the degraded document images, the intensity statistic of SSPs is a good approximation of the local threshold. As shown in Fig. 1, the SSPs are defined as the pixels around strokes which contain both text and background pixels and these pixels have large

gradient magnitudes and opposite gradient orientations. Therefore, we will extract the SSP in the gradient map from two different aspects: 1) large magnitude: we propose an adaptive gradient map binarization algorithm to compute a global threshold. The pixel whose gradient magnitude is larger than that threshold value will be regarded as a SSP candidate. 2) opposite orientation: we conduct orientation symmetry test on all the candidates. Only the candidates that satisfy the symmetry constraint are preserved as the SSPs. Since the gradient orientations can not be symmetric opposite when the size of neighborhood is smaller than the stroke width, we propose an iterative stroke width estimation algorithm to estimate the stroke width adaptively.

#### 3.2.1. SSP candidates extraction using adaptive gradient binarization

Background estimation followed by the image normalization procedure compensates many variations of image degradation. Therefore the gradient magnitude map can be binarized by a global threshold value. We adopt the Otsu threshold [15] in our previous work [5]. But this threshold is inappropriate when the gradient map does not fit bimodal distribution. Besides, we can hardly distinguish the edge and non-edge pixels in intensity space when it comes to faint characters and bleed-through background. As shown in Fig. 4 (e) and (h), these two images have similar intensity characteristics. A good binarization of the gradient map must be able to maintain the completeness of the character edges which should be neither broken nor fused. To this end, we propose an adaptive global threshold selection approach by finding the maximum of the average edge pixels in gradient map. Pixels whose magnitudes are larger than that threshold value will be regarded as SSP candidates.

Based on the observation that a good binarized result should contain complete stroke edges as many as possible and the less isolate noise the better, we introduce an evaluation function  $E(t)$  on the global threshold  $t$  as follows:

$$E(t) = \frac{\sum_{x=1}^M \sum_{y=1}^N f_{\text{edge}}(G_t(x, y))}{N_{CC}}, \quad t = 0, 1, \dots, 255 \quad (4)$$

$$G_t(x, y) = \begin{cases} 1 & G(x, y) > t \\ 0 & \text{else} \end{cases} \quad (5)$$

$$f_{\text{edge}}(G_t(x, y)) = \begin{cases} 1 & G_t(x, y) = 1 \quad G_t(p, q) = 0, \exists (p, q) \in N_{(x,y)} \\ 0 & \text{else} \end{cases} \quad (6)$$

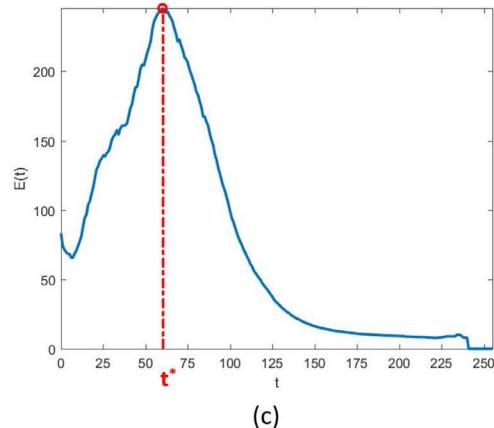
where  $M$  and  $N$  denote the height and width of the input image respectively.  $G_t$  denotes the bi-level image binarized by the threshold  $t$  with 0 corresponding to non-edge and 1 to edge.  $N_{CC}$  denotes the total number of connected components in the binarized image  $G_t$ . The symbol  $f_{\text{edge}}(\cdot)$  denotes a edge discriminant function whose value is 1 when the pixel belongs to edge and there is at least one

Ravno za to nam je pa toliko treba političnega lista. Samo, da bi se mož zato pripraven našel, kteri bi hotel težavno vredništvo na svoje rame vzeti. Dopisnikov, pravih iškrenih domoljubov, ki poznajo natanko stan našega ljudstva, mu se gotovo manjkalo ne bode. Oni spoznajo ime-

(a)

Ravno za to nam je pa toliko treba političnega lista. Samo, da bi se mož zato pripraven našel, kteri bi hotel težavno vredništvo na svoje rame vzeti. Dopisnikov, pravih iškrenih domoljubov, ki poznajo natanko stan našega ljudstva, mu se gotovo manjkalo ne bode. Oni spoznajo ime-

(b)



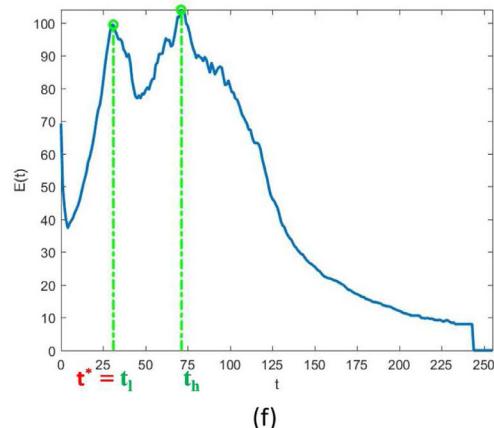
(c)

*Now if there be any meaning in words all  
as a Leader is preparing for a party for explaining  
this must signify that the monad conduct of a  
Loyalty party a Being as this.*

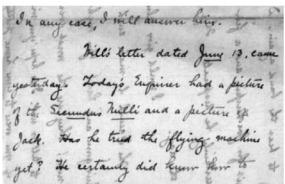
(d)

*Now if there be any meaning in words all  
as a Leader is preparing for a party for explaining  
this must signify that the monad conduct of a  
Loyalty party a Being as this.*

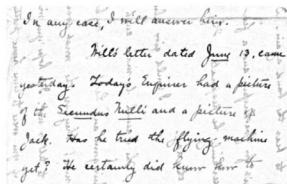
(e)



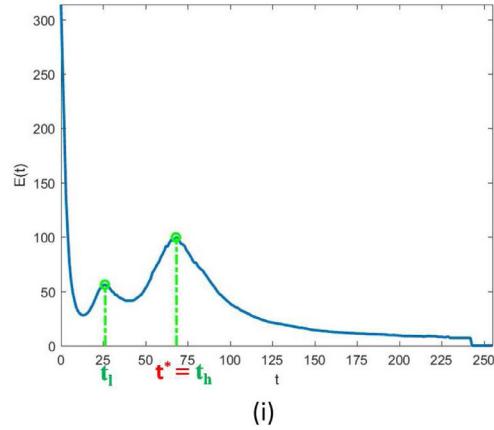
(f)



(g)



(h)



(i)

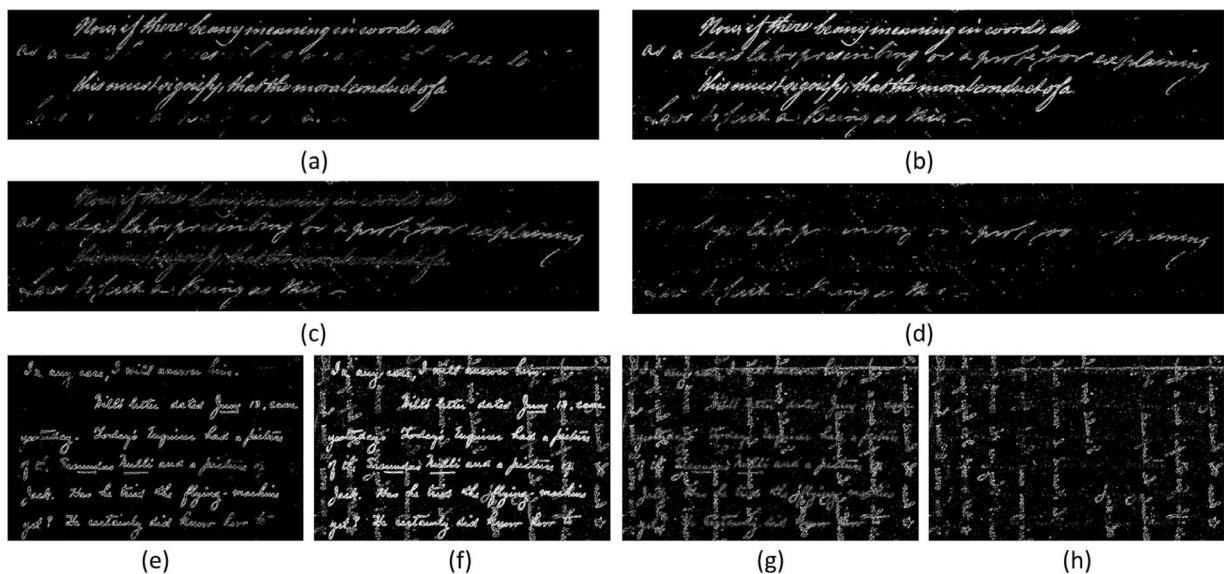
**Fig. 4.** Three examples of global threshold selection: (a)(d)(g) original images: PR5 and HW3 from DIBCO'13 and H06 from H-DIBCO'12. (b)(e)(h) Normalized images after background compensation. (c)(f)(i) The variation curve between  $E(t)$  and the global threshold  $t$ .

non-edge pixel in its neighborhood.  $N_{(x, y)}$  in Eq. (6) denotes the 8 connected neighborhoods centered on the pixel  $(x, y)$ .

The evaluation function measures the average number of contour pixels of stroke edge in one connected component. If the threshold  $t$  is too big, there will be a few foreground pixels in  $G_t$  which leads to a small numerator in Eq. (4). In contrast, if  $t$  is too small, the isolated noise will appear so that the denominator, in other words, the number of connected components becomes bigger. Therefore, the value of the fraction in Eq. (4) is small in the both situations and there should exist a certain threshold  $t^*$  to maximize the value of the evaluation function. Fig. 4 shows three function curves obtained from three images respectively. As we can see, the normalized image (b) with clear text and simple back-

ground only contain one peak which reflects the best threshold  $t^*$  undoubtedly. But both the other curves contain two peaks for very different reasons: there are some faint characters in the image (e) while large area bleed-through background appears in image (h). Thus image (e) should use the low threshold  $t_l$  to preserve the light text strokes and the high threshold  $t_h$  should be used to remove the dark background pixels in image (h). In practice, there may be more than two peaks when evaluated over a large amount of images with various characteristics. But we only process the two highest peaks. Next we will introduce a valid method to distinguish the two kinds of threshold candidates.

In light of our observation, the proportion of bleed-through background pixels is usually higher than that of the faint charac-



**Fig. 5.** The procedure to select the better threshold from  $t_l$  and  $t_h$ . (a)(e) Binarized image  $G_{t_h}$ . (b)(f) Binarized image  $G_{t_l}$ . (c)(g) The middle magnitude images  $G_{mid}$ . (d)(h) The weak stroke edge images  $G_{dif}$ .

ter pixels. Therefore, we compare the proportion of the middle-level intensity pixels with a ratio threshold  $R_{thre}$ . If the proportion is higher than  $R_{thre}$ , the optimal global threshold  $t^*$  will be the high threshold  $t_h$ , otherwise,  $t^*$  will be the low threshold  $t_l$ . In this paper, we use the value of 0.05 to make the right selection between  $t_h$  and  $t_l$ . In practical application, we can adjust its value to adapt to different situations.

To extract the stroke edges which would not appear with the high threshold but appear with the low threshold, we first produce a binary image  $G_{mid}$  with 1 corresponding to pixels whose intensity is between  $t_l$  and  $t_h$  and 0 to others:

$$G_{mid}(x, y) = \begin{cases} 1 & G(x, y) > t_l, \quad G(x, y) < t_h \\ 0 & \text{else} \end{cases} \quad (7)$$

Then we further extract the emerging stroke edges which appear only because of the low threshold:

$$G_{dif}(x, y) = \begin{cases} 1 & G_{mid}(x, y) = 1 \quad G_{t_h}(p, q) = 0, \quad \forall (p, q) \in N_{(x, y)} \\ 0 & \text{else} \end{cases} \quad (8)$$

The binarized image  $G_{dif}$  shows the stroke edge pixels of faint characters or bleed-through background. If the gradient magnitude of the pixel is between  $t_l$  and  $t_h$  and all its neighborhood pixels do not appear in the image  $G_{t_h}$ , we label the corresponding pixel in  $G_{dif}$  as 1. In this way we can remove the pixels adjacent to the strong edges from  $G_{mid}$ . Two examples of the various binarized images corresponding to the normalized images from Fig. 4 are shown in Fig. 5. Finally, we compute the ratio  $r$  between the remained pixels labeled as 1 in  $G_{dif}$  and the total number of pixels and compare  $r$  with the fixed ratio threshold  $R_{thre}$ . Therefore, for the original image in Fig. 4(d), the optimal global threshold is  $t_l$  and the corresponding binarized gradient map is shown in Fig. 5(b). In contrast, we choose the higher threshold  $t_h$  for the image Fig. 4(g) and then obtain the binarized image shown in Fig. 5(e).

In conclusion, the complete procedure to compute the global threshold is shown in [Algorithm 1](#). Then we get the text edge binarized image  $G_{t^*}$  with 1 representing the SSP candidates.

---

**Algorithm 1** Adaptive global threshold selection.

**Require:** The normalized image  $I_{norm}$  and ratio threshold  $R_{thre}$

**Ensure:** The best global threshold  $t^*$

```

1: procedure FINDGLOBALTHRESHOLD( $I_{norm}, R_{thre}$ )
2:   for  $t = 0 \rightarrow 255$  do
3:     Get the bi-level image  $G_t$ 
4:     Calculate the evaluation value  $E(t)$ 
5:   end for
6:   Detect one or two peaks of the function curve
7:   if (there is only one peak corresponding to  $T$ ) then
8:      $t^* = T$ 
9:   else
10:    Get the middle magnitude image mask  $G_{mid}$  based on  $t_l$ 
      and  $t_h$ 
11:    Get  $G_{dif}$  by removing the pixels adjacent to the strong
      edges in  $G_{mid}$ 
12:    Calculate the ratio  $r$  corresponding to  $G_{dif}$ 
13:    if  $r > R_{thre}$  then
14:       $t^* = t_h$ 
15:    else
16:       $t^* = t_l$ 
17:    end if
18:  end if
19: end procedure

```

### 3.2.2. SSP extraction using stroke width estimation

To further extract SSPs from the SSP candidates, we will remove those pixels with asymmetric gradient orientation from the foreground in  $G_{t^*}$ . In our previous work [5], we divide angle plane into eight overlapping intervals, which are evenly distributed throughout the plane of 360 degrees and each one of them is a range of 135 degrees. Every interval is called an orientation group. For all the SSP candidate-centered neighborhood window, we calculate the SSP candidate count for each orientation group. If the gradient orientations focus on only one group, the center SSP candidate is determined to be asymmetric. According to [5], we use the threshold value of 75 percent of the total count of SSP candidates in neighborhood to determine if the center pixel is asymmetric or

not. Since the neighborhood size is closely related to the width of text strokes, we will estimate the stroke width first.

Based on the observation of the document images in different datasets, we propose the following assumptions. First, the document consists of one or more horizontal text parts. Secondly, the stroke width is uniform through each part. Once the text strokes are detected and labeled as 1 in  $G_{t^*}$ , we can estimate the stroke width by using the method proposed in Lu's paper [19]. Specifically, they scan the stroke edge image row-by-row and record the distance between all adjacent stroke edge pixel pairs in each row. The most frequent distance will be the estimated stroke width. This method works well in the case of uniform text strokes and low noise. Considering that machine-printed document images usually contain characters with different font sizes and the noise may introduce some wrong adjacent distance values, we make two improvements: 1) multiple stroke width values are estimated using horizontal projection; 2) an iterative framework for noise removing is used to increase the precision of estimated stroke width. First we segment the original image into multiple sub-images by applying horizontal projection. Then we estimate stroke width for each sub-image. Once the estimated values are obtained, some connected components which are smaller than half of the stroke width will be removed. Then we repeat these steps until the terminating conditions are satisfied. The concrete procedure is shown in [Algorithm 2](#). In [Section 4](#), we will use the experimental results to demonstrate the advantages of the two improvements.

---

**Algorithm 2** Iterative stroke width estimation.
 

---

**Require:** The text edge binarization image  $G_{t^*}$  and the maximum number of iterations  $N_{\max}$

**Ensure:** The estimated stroke width array  $SW$  and a new text edge binarization image  $G_{SW}$

- 1: **procedure** ITERATEESTIMATESTROKewidth( $G_{t^*}, N_{\max}$ )
- 2:    $SW = \text{ESTIMATESTROKewidth}(G_{t^*})$
- 3:   Create a new empty array  $SW2$
- 4:   Initialize the iteration number  $i = 0$
- 5:   **while** ( $SW2 \neq SW, i < N_{\max}$ ) **do**
- 6:      $SW2 = SW$
- 7:     Get a new binarized image  $G_{SW}$  produced by removing the small connected components from  $G_{t^*}$  according to the estimated stroke width array  $SW$
- 8:      $SW = \text{ESTIMATESTROKewidth}(G_{SW})$
- 9:      $i \leftarrow i + 1$
- 10:   **end while**
- 11: **end procedure**
- 12:
- 13: **procedure** ESTIMATESTROKewidth( $G_{t^*}$ )
- 14:   Divide the image into  $P$  parts by using horizontal projection on  $G_{t^*}$
- 15:   **for**  $i = 1 \rightarrow P$  **do**
- 16:     Get the part bi-level image  $I_i$
- 17:     Calculate the most frequency adjacent distance  $d$  in  $I_i$
- 18:      $SW[i] = d$
- 19:   **end for**
- 20: **end procedure**

---

Two examples of SSP extraction procedure are shown in [Fig. 6](#). Through the stroke width estimation process, we filter the noise in SSP candidates to get a new text edge binarization image  $G_{SW}$ . Then we remove those pixels with asymmetric gradient orientation from the foreground in  $G_{SW}$  by using the method proposed in [5] in which the neighborhood size is set to be 6 times of the stroke width. As shown in [Fig. 6](#), the single white curve in the top of the image (g) disappears in image (h) through the judgement on ori-

entation symmetry. Finally, the precise SSPs are extracted through a series of the process described above.

### 3.3. Multiple threshold vote based binarization

In this paper, we utilize the SSPs to calculate the local threshold in neighborhood so as to suppress the non-text pixels and maintain the text ones as well. In order to deal with the inaccurate detection of SSPs, we further propose a multiple threshold vote based framework to determine whether one pixel belongs to the foreground or not.

First, we compute a local threshold value in each SSP-centered neighborhood as follows:

$$T(x, y) = m(x, y) + k \times \sigma(x, y), \quad \forall (x, y) \in \text{SSP} \quad (9)$$

where  $m(x, y)$  and  $\sigma(x, y)$  are the mean and standard deviation of the intensity of SSPs within a neighborhood window around the pixel  $(x, y)$  respectively. The parameter  $k$  controls the behavior of the method: the higher its value, the more text pixels occur. The neighborhood window should be at least larger than the stroke width to contain stroke edge pixels. So we assign the size of the neighborhood to be  $\alpha$  times the corresponding stroke width. Then  $\alpha$  becomes the second parameter in our method. Through the comparative experiments in [Section 4.2](#),  $k$  and  $\alpha$  are set to be 0.6 and 2 respectively.

In this way, pixels in the overlapped neighborhoods would have multiple threshold values while pixels which never appear in any neighborhood have no threshold. As shown in [Fig. 7](#), there are three overlapped neighborhood windows around three SSPs. We use [Eq. \(9\)](#) to calculate a local threshold for each window. Every pixel in the window is compared with the threshold. Then we create a map  $N$  to record the comparison results and initialize each value in it as 0. For every pixel  $(i, j)$  in the neighborhood centered on the SSP  $(x, y)$ , we compare the intensity  $I(i, j)$  with the local threshold  $T(x, y)$  as follows:

$$N(i, j) \leftarrow \begin{cases} N(i, j) + 1 & I(i, j) < T(x, y) \\ N(i, j) - 1 & \text{else} \end{cases} \quad (10)$$

Finally, the binarized image  $I_{bin}$  is calculated as follows:

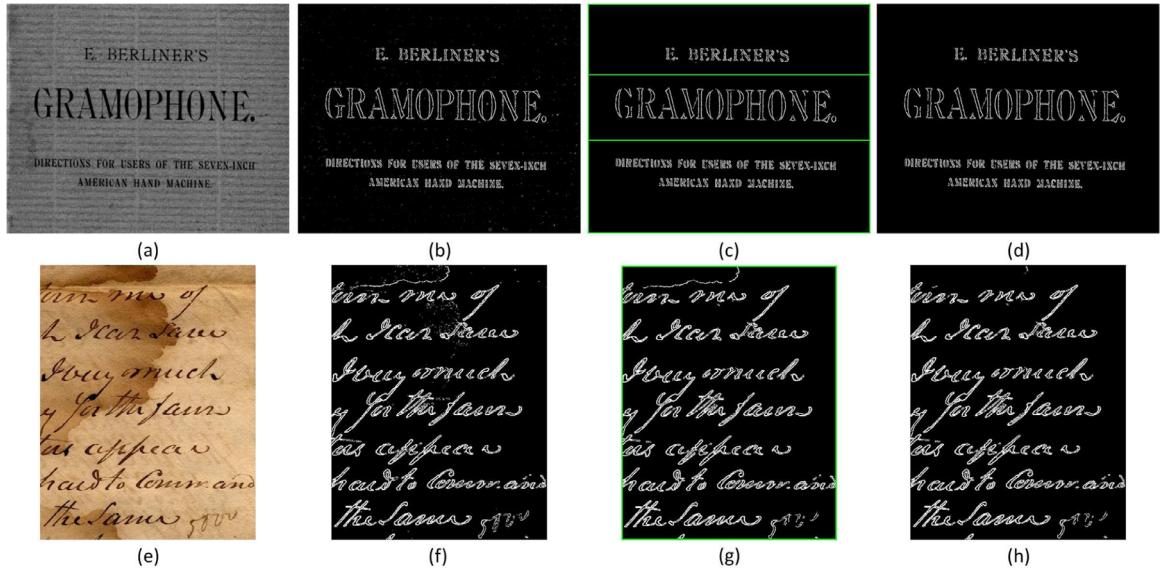
$$I_{bin}(i, j) = \begin{cases} 1 & N(i, j) > 0 \\ 0 & \text{else} \end{cases} \quad (11)$$

if  $N(i, j)$  is bigger than zero, the pixel  $(i, j)$  will be labeled as text. The reason to do so is twofold. The final label of the pixel with multiple thresholds follows the majority voting results. The labels are background for those pixels with no threshold, in other word, the value in map  $N$  is always the initial value zero.

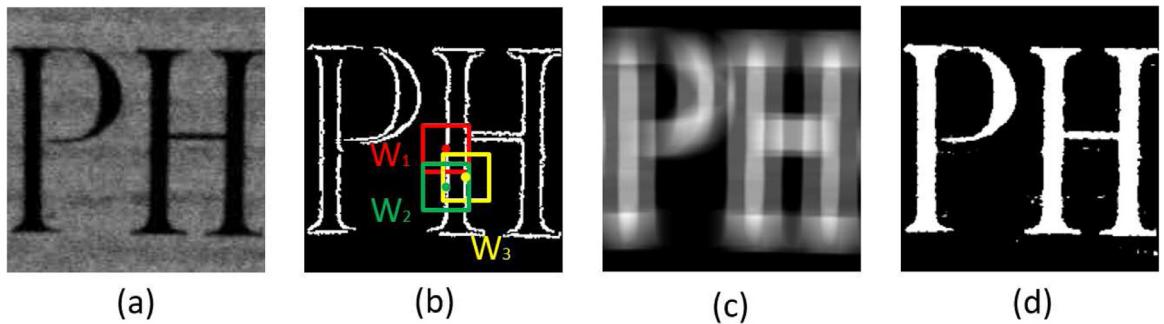
Once the initial binary image is derived from [Eq. \(11\)](#) described above, the binarization result can be further improved by removing the noisy artifacts of small size. In this paper, if the size of one text connected component is smaller than half of the estimated stroke width and this component is not connected to any pixel in SSP, we can call it a noisy artifact. The reason for this is twofold. First, the edges of big artifacts have already been removed in the SSP extraction process by using the gradient symmetry filter. Second, the text pixels are binarized only near the SSPs because the size of the neighborhood window which is used in the local threshold binarization stage is only two times the corresponding stroke width. Therefore, the initial binary image cannot contain the artifacts of very large size and we only need to remove the small noisy artifacts in the post-processing step.

## 4. Evaluation

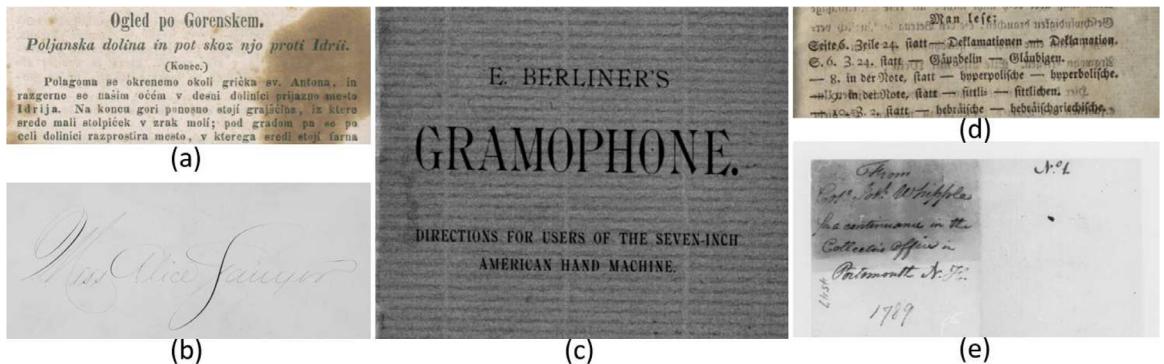
The proposed method is evaluated on several public datasets from the previous document image binarization competitions:



**Fig. 6.** The SSP extraction procedure. (a)(e) Original images. (b)(f) Binarized image  $G_{t^*}$ : SSP candidates. (c)(g) Binarized image  $G_{SW}$ : filtered SSP candidates, the green lines are the border between different stroke width estimations which are 6, 13 and 6 from top to bottom in (c) and 5 in(g). (d)(h) The locations of SSPs illustrated with white pixels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 7.** The illustration of the voting framework. (a) Original image. (b) Three SSP-centered neighborhood windows  $W_1$ ,  $W_2$  and  $W_3$  which are used to calculate the local threshold values. (c) The total computation number of local threshold for each pixel: the lighter the color, the greater the number. (d) The possibility of text for each pixel: the lighter the color, the greater the possibility.



**Fig. 8.** Some document image examples taken from (H)DIBCO datasets.

DIBCO'09, DIBCO'11, DIBCO'13, H-DIBCO'10, H-DIBCO'12, H-DIBCO'14 and H-DIBCO'16 [6–12]. The first three datasets contain both handwritten and machine-printed images while the rest ones only contain the handwritten images. Every dataset provides a collection of hand-annotated ground truth binarization results. All these document images contain challenging noise, various fonts and different types of degradations. Some examples are shown

in Fig. 8. To demonstrate the effectiveness of every improvement proposed in this paper, we do a series of experiments by using the improvements and the basic processes respectively. Then, we compare our proposed binarization method with some traditional and state-of-the-art algorithms and the top ranking algorithm in each competition.

#### 4.1. Evaluation metrics

There are many metrics available for the evaluation of image binarization methods in the contest reports [6–12]. Six of them used in our study are as follows:

- F-Measure (FM)

$$FM = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (12)$$

where  $Precision = \frac{TP}{TP+FP}$  and  $Recall = \frac{TP}{TP+FN}$ . The  $TP$ ,  $FP$  and  $FN$  denote the true positive, false positive and false negative values.

- pseudo F-Measure (p-FM)

$$p-FM = \frac{2 \times pRecall \times Precision}{pRecall + Precision} \quad (13)$$

where  $pRecall$  is defined as the percentage of the skeletonized ground truth image described in [7,9].

- Peak Signal to Noise Ratio (PSNR)

$$PSNR = 10 \log \left( \frac{C^2}{MSE} \right) \quad (14)$$

where  $MSE = \frac{\sum_{x=1}^M \sum_{y=1}^N (l_{bin}(x,y) - l'_{bin}(x,y))^2}{MN}$ ,  $C$  denotes the difference between text and background. PSNR measures how close two images are.

- Negative Rate Metric (NRM)

$$NRM = \frac{NR_{FN} + NR_{FP}}{2} \quad (15)$$

where  $NR_{FN} = \frac{FN}{FN+TP}$  and  $NR_{FP} = \frac{FP}{FP+TN}$ . The NRM measures how mismatched between the GT and predicted image.

- Misclassification Penalty Metric (MPM)

$$MPM = \frac{MP_{FN} + MP_{FP}}{2} \quad (16)$$

where  $MP_{FN} = \frac{\sum_i d_{FN}^i}{D}$  and  $MP_{FP} = \frac{\sum_j d_{FP}^j}{D}$ .  $d_{FN}^i$  and  $d_{FP}^j$  denote the distance of the  $i_{th}$  false negative and the  $j_{th}$  false positive pixel from the contour of the text in the GT image. The normalization factor  $D$  is the sum over all the pixel-to-contour distances of the GT object. A lower MPM score indicates a better binarization performance.

- Distance Reciprocal Distortion Metric (DRD)

$$DRD = \frac{\sum_k DRD_k}{NUBN} \quad (17)$$

where  $DRD_k$  is the distortion of the  $k$ -th flipped pixel and  $NUBN$  is the number of the non-uniform (not all black or white pixels)  $8 \times 8$  blocks in the GT image. The DRD is used to measure the visual distortion in binary document images.

#### 4.2. Evaluation of the proposed method's performance

##### 4.2.1. Parameter selection

As introduced in the previous Section 3.3, the proposed algorithm requires two input parameters:  $\alpha$  and  $k$ . In the proposed method, we assign the size of neighborhood window to be  $\alpha$  times the corresponding estimated stroke width. The neighborhood window should be at least larger than the stroke width to contain stroke edge pixels. If  $\alpha$  is too small, the number of SSPs in neighborhood will be too few to lose statistical significance. Conversely, if the  $\alpha$  value is too large, the threshold value will deviate from the pixel itself due to the far distance. The mean and standard deviation in Eq. (9) are computed by the intensity of SSPs in this neighborhood. The weight of the two statistics is parameter  $k$  which could control the binarized result: the higher the value of  $k$ , the more pixels will be predicted as text.

In order to select the optimal values of the two parameters, we create a training dataset by randomly choose fifty percent of the total images for each dataset and a series of experiments are conducted on them. The average FM for all the seven training datasets with the change of  $\alpha$  and  $k$  are used as the evaluation criterion. Judging from the results shown in Fig. 9(a), we can pick the values of  $\alpha = 2$  and  $k = 0.6$ , which perform best based on the average FM across the training datasets. Fig. 9(b) and (c) show the variation of FM with variable  $\alpha$  and  $k$  on training datasets, respectively. There is a similar trend on the complete datasets shown in Fig. 9(d) and (e), which demonstrates the robustness of the two parameters. These two selected values are not adapted any further in the subsequent experiments. It is worth noting that the optimized neighborhood window is a square with 2 times the stroke width, thus the distance between the boundary and center pixels is just one stroke width. Since all the center pixels in our framework are the SSPs, i.e. the stroke edge pixels, the optimized neighborhood contains approximately one text stroke.

##### 4.2.2. Evaluation under different experimental settings

Compared with our previous work [5], there are three main procedures to improve the quality of binarization results in this paper. We did a series of experiments to prove the effectiveness of the three procedures.

- (a) Adaptive global threshold selection algorithm

To extract the SSP candidates more precisely, we propose an adaptive global threshold selection algorithm (i.e. Algorithm 1) to compute the global threshold. The threshold is used to binarize the compensated gradient magnitude image. In order to demonstrate the effectiveness of this algorithm, we binarize the image using Otsu's [15] global threshold.

- (b) Iterative stroke width estimation algorithm

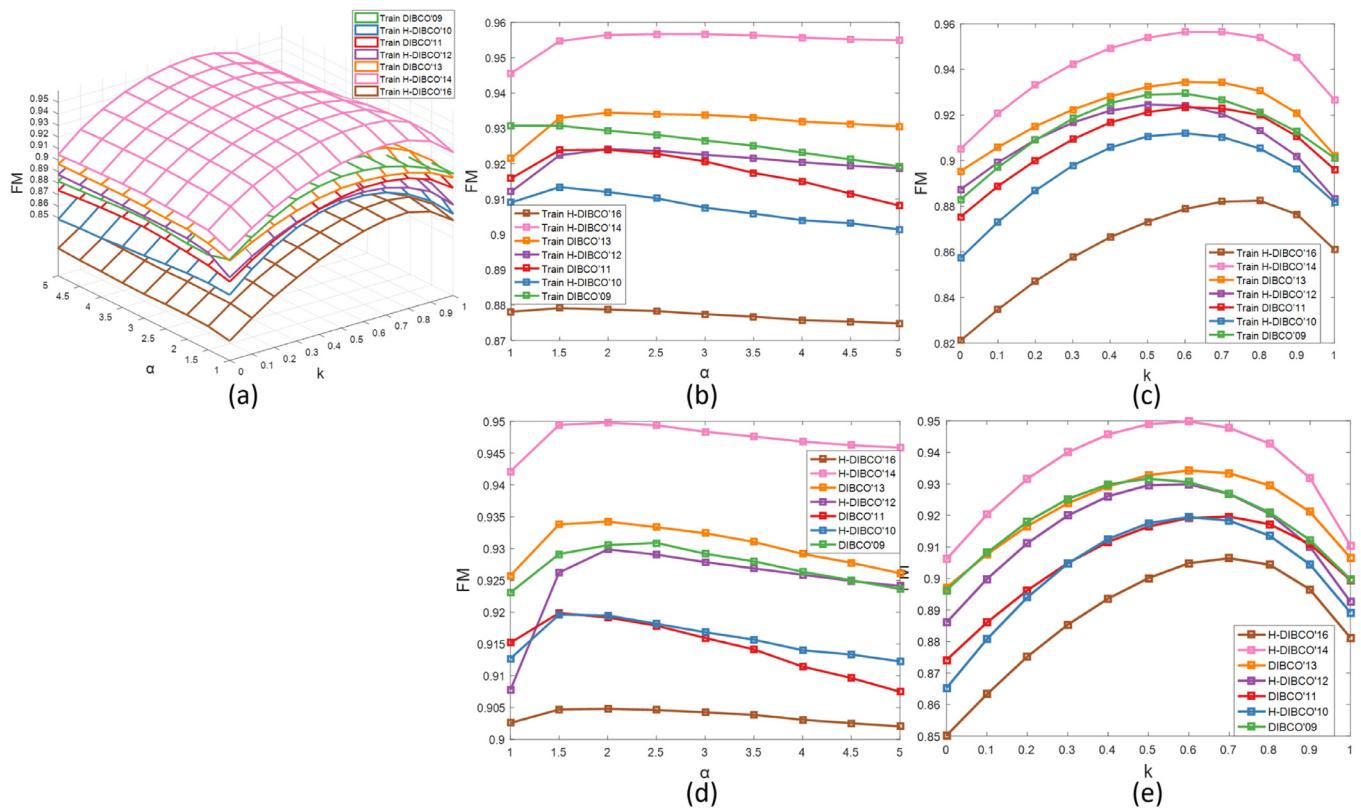
The stroke width in the document image is the basis of the neighborhood size used in our binarization method. Its effect is not only embodied on the orientation symmetric SSP extraction but also on the local threshold calculation. To demonstrate the necessity of the stroke width estimation procedure, we do a comparative experiment in which the stroke width for each image is set to a manually selected value. Fig. 10 shows the FM measures on all the datasets for various values of the stroke width. We can pick the best value of 4, which performs best in most datasets.

- (c) Multiple local threshold vote based framework

Since the SSP extraction causes errors inevitably, the subsequent binarization framework needs to have the ability to compensate for the lack or error of the detected SSPs. We use the multiple local threshold vote based framework to achieve this goal. In the comparative experiment, we compute the unique local threshold using the SSPs in neighborhood for each pixel in the image.

The average FM measures for all the seven datasets under different experimental settings are shown in Table 1. The first line of the table lists the result obtained by the proposed method. The next three lines represent the results obtained by the three comparative experiments described above.

As we can see from the results shown in the second line in Table 1, the Algorithm 1 outperforms Otsu's method on each dataset especially on DIBCO'11 and DIBCO'13. Both of the two datasets contain many degraded images with faint characters and bleed-through background. According to the definition of the evaluation function  $E(t)$ , there are two peaks in those degraded images. We list the two-peak and total number of images in Table 2. The results indicate that FM measure increases significantly in the dataset which contains two-peak degraded images. This demonstrates the effectiveness of our algorithm using the geographic



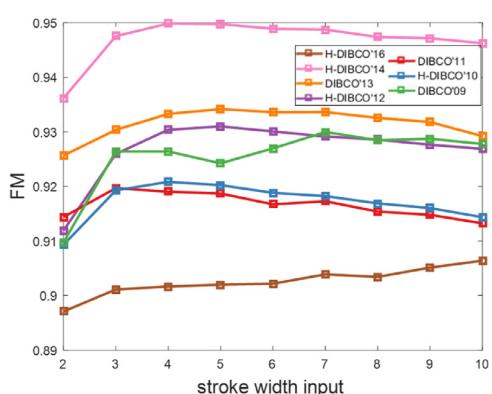
**Fig. 9.** Effect of parameter  $\alpha$  and  $k$  on average FM measure for the seven datasets. (a) The variation of FM with the change of both  $\alpha$  and  $k$  on the training datasets. (b)(d) The variation with the change of  $\alpha$  when  $k = 0.6$  on the training or complete datasets. (c)(e) The variation with the change of  $k$  when  $\alpha = 2$  on the training or complete datasets.

**Table 1**  
FM measure (%) under different experimental settings.

	DIBCO'09	H-DIBCO'10	DIBCO'11	H-DIBCO'12	DIBCO'13	H-DIBCO'14	H-DIBCO'16	Average
Proposed	<b>93.05</b>	91.95	<b>91.92</b>	92.99	<b>93.42</b>	94.98	<b>90.48</b>	<b>92.68</b>
(a)	92.69	91.28	86.05	92.27	89.64	94.47	89.31	90.82
(b)	92.64	<b>92.08</b>	91.90	<b>93.04</b>	93.33	<b>94.99</b>	90.17	92.60
(c)	92.94	91.55	91.33	90.17	92.16	94.27	84.61	91.00

**Table 2**  
The number of images with two-peak  $E(t)$  in each dataset.

	DIBCO'09	H-DIBCO'10	DIBCO'11	H-DIBCO'12	DIBCO'13	H-DIBCO'14	H-DIBCO'16
Increased FM (%)	0.37	0.67	5.87	0.71	3.78	0.51	1.17
Two-peak number	0	0	6	5	4	0	1
Total number	10	10	16	14	16	10	10



**Fig. 10.** Effect of different stroke width values for the seven datasets.

distribution to select the optimal threshold. In this way we can suppress the bleed-through background while preserving the faint characters. For the datasets containing no two-peak curve at all, the FM measure still increases by using the main peak of  $E(t)$ , which further demonstrates the superiority of our adaptive algorithm based on the completeness of stroke edges.

The FM measures corresponding to the manually selected stroke width are shown in the third line in Table 1. The results show that our method performs better in DIBCO'09, DIBCO'11, DIBCO'13 and H-DIBCO'16 but performs worse in rest ones. Note that the datasets DIBCO'09, DIBCO'11 and DIBCO'13 contain both handwritten and machine-printed images while the other four datasets only contain handwritten images. The reason may lie in the fact that the stroke width of the handwritten image is not consistent which leads to inaccurate estimation of the stroke width. To prove this conjecture, we separate the three mixed datasets into handwritten and machine-printed ones and the results are shown

**Table 3**  
FM measure (%) on machine-printed and handwritten datasets.

	H-DIBCO'09	P-DIBCO'09	H-DIBCO'11	P-DIBCO'11	H-DIBCO'13	P-DIBCO'13
Proposed	91.88	94.22	91.28	92.55	93.05	93.79
Without(b)	92.56	92.72	91.22	92.59	93.03	93.63

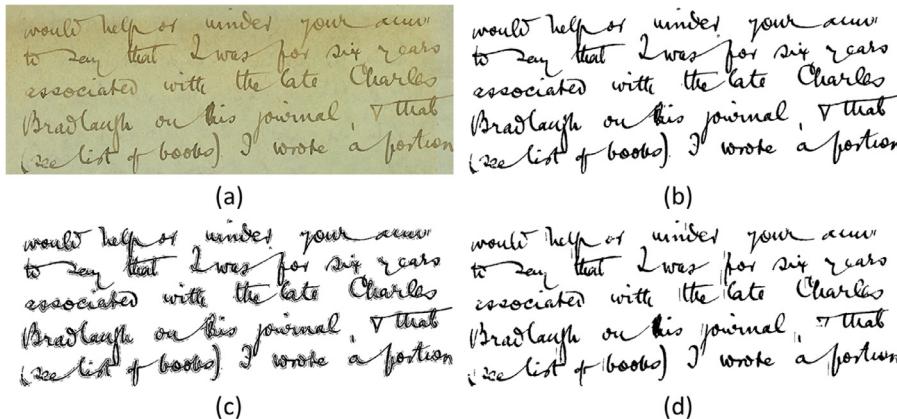


Fig. 11. Binarization results of a sample image (H09 in H-DIBCO'12) using different methods. (a) Original image. (b) Proposed. (c) Unique local threshold per pixel. (d) Lu [19].

in Table 3. We can observe that our method achieves a better performance on machine-printed document images than handwritten images. In conclusion, compared with the method using manually selected stroke width, our method using adaptive estimated ones achieves the superior performance on the printed images and it is comparable on the handwritten documents. Besides, the manual method is too sensitive to image resolution and can not work well when the value is selected inappropriately. Therefore our adaptive algorithm is a good choice to avoid the risk of improper manual selection of stroke width and deal with the machine-printed images with different font sizes.

The results without applying the multiple threshold voting strategy are shown in the last line in Table 1. As we can see, the FM measure on each dataset increases obviously through the voting framework. The reason for the better performance is twofold. On the one hand, through increasing the number of local thresholds for the pixels to be judged, we can avoid the wrong judgement caused by the single local threshold which might be inaccurate. On the other hand, the voting framework can effectively eliminate the boundary noise in whose neighborhood the SSPs usually happen to be all background pixels. Fig. 11(c) shows a binarization result based on unique local threshold per pixel. There are many wrong text pixels around the correct ones. The distance from the wrong text pixel to its nearest SSP is about half of the neighborhood window. Under such conditions, the count of SSPs in its neighborhood is relatively small and mostly belongs to background. This causes the estimated local threshold to be so high that the background pixel is erroneously judged to be foreground. Lu et.al [19] also use a unique local threshold to obtain the binarization result such as Fig. 11(d). We can see from the picture that there are also some boundary noise in their result. In the multiple local threshold vote based framework, we estimate the local threshold only in the SSP-centered neighborhood which means that there are enough symmetric SSPs to estimate a suitable local threshold.

Compared with our previous work [5], we also improve the image compensation method from the following two aspects: we use a new algorithm [26] to estimate the background surface and we use the deformation of Sobel operator to compute the gradient map. The average FM values across the seven datasets are 91.32% and 91.55% when we use the background estimation method in

[5] and the original Sobel operator respectively. They are both smaller than the FM value of the proposed method (92.68%). Therefore, the two improvements are effective to generate the compensated image.

#### 4.3. Comparison with other binarization methods

In this section, we compare our method with some traditional and state-of-the-art binarization algorithms including Otsu [15], Sauvola [4], Lu [19], Su [22], Howe [28], Lelore [23] Mitianoudis [35] and Jia [5]. First the global thresholding method of Otsu and the local thresholding technique of Sauvola are used in our benchmarking experiments. For the Sauvola's method, we use a  $30 \times 30$  neighborhood window and a value of  $k = 0.2$  to calculate the local threshold. Then two binarization methods based on text stroke edges, as proposed by Lu and Su<sup>2</sup> are used to compare with ours. For Howe's<sup>3</sup> method using the Laplacian energy, we choose its automated threshold version. For the FAIR method of Lelore<sup>4</sup> and the LCM method of Mitianoudis<sup>5</sup>, the default parameters in their demos are used to produce the binarization results. Finally, we also compare the proposed method with the 1st rank method of each contest.

##### 4.3.1. Time complexity

Since the speed of our method depends on the image size, we estimate the runtime in second per megapixel. Considering that the methods of Lelore [23] and Mitianoudis [35] put demo programs online but do not provide their executable codes, besides, the methods of Lu [19] and Howe [28] have only MATLAB codes, we only compare the average runtimes of the proposed method with the other four methods with C/C++ code as shown in Table 4. All the experiments are performed under such an environment: quad-core Intel Core i5 at 3.1 GHz with 8GB RAM. Since the same processing is required for each global threshold in Algorithm 1, with no dependencies in the computations, the algorithm can be

<sup>2</sup> Code kindly provided at <https://sites.google.com/site/flydreamersu/research>.

<sup>3</sup> Code kindly provided at <http://www.cs.smith.edu/~nhowe/research/code/>.

<sup>4</sup> Demo kindly provided at <http://lelore.lsis.univ-tln.fr/demoFAIR/>.

<sup>5</sup> Binarization samples and demo provided at <http://utopia.duth.gr/nmitiano/publications.html>.

**Table 4**  
Average runtimes(seconds/megapixel) of different algorithms.

Method	Proposed	Proposed(parallel)	Otsu [15]	Sauvola [4]	Jia [5]	Su [22]
Runtime	4.4488	2.4568	0.0010	0.0454	0.1194	3.7847

**Table 5**  
Comparisons of the performance of the proposed and the other algorithms against DIBCO'09 dataset (best values highlighted in bold).

	FM(%)	p-FM(%)	PSNR	NRM( $10^{-2}$ )	MPM( $10^{-3}$ )	DRD
Otsu [15]	78.60	80.53	15.31	5.64	13.69	22.57
Sauvola [4]	85.37	89.08	16.37	6.91	3.67	7.08
Lu [19]	91.13	92.28	18.66	4.31	0.54	3.05
Su [22]	93.02	94.61	19.41	4.05	0.82	2.64
Howe [28]	<b>94.04</b>	<b>95.06</b>	<b>20.43</b>	<b>2.72</b>	1.36	<b>2.10</b>
Lelore [23]	93.93	95.10	20.21	2.76	0.44	2.17
Mitianoudis [35]	90.27	92.69	18.08	5.58	0.80	3.71
Jia [5]	91.37	94.56	18.49	4.58	0.65	3.06
1st rank of contest[6]	91.24	-	18.66	4.31	0.55	-
Proposed	93.05	94.60	19.29	3.18	<b>0.45</b>	2.40

**Table 6**  
Comparisons of the performance of the proposed and the other algorithms against H-DIBCO'10 dataset (best values highlighted in bold).

	FM(%)	p-FM(%)	PSNR	NRM( $10^{-2}$ )	MPM( $10^{-3}$ )	DRD
Otsu [15]	85.43	90.64	17.52	9.36	1.58	4.05
Sauvola [4]	75.18	84.08	15.94	16.28	2.46	7.22
Su [22]	91.36	93.18	19.78	5.98	0.49	2.42
Howe [28]	93.59	<b>94.81</b>	21.08	4.20	0.37	<b>1.72</b>
Lelore [23]	<b>93.82</b>	94.27	<b>21.09</b>	<b>3.53</b>	<b>0.24</b>	1.79
Mitianoudis [35]	88.97	91.16	18.32	4.67	1.38	3.38
Jia [5]	89.46	93.94	18.86	7.14	0.84	2.93
1st rank of contest[7]	91.50	93.58	19.78	5.98	0.49	-
Proposed	91.95	94.75	19.83	4.37	0.39	2.17

**Table 7**  
Comparisons of the performance of the proposed and the other algorithms against DIBCO'11 dataset (best values highlighted in bold).

	FM(%)	p-FM(%)	PSNR	NRM( $10^{-2}$ )	MPM( $10^{-3}$ )	DRD
Otsu [15]	82.10	85.96	15.72	8.16	16.09	8.95
Sauvola [4]	82.14	87.70	15.65	9.09	11.25	8.50
Lu [19]	79.71	81.54	15.55	9.18	39.50	21.47
Su [22]	87.83	90.24	17.71	7.33	5.33	4.66
Howe [28]	90.79	92.28	19.01	4.30	8.03	4.46
Lelore [23]	<b>92.48</b>	94.11	<b>19.37</b>	<b>3.91</b>	2.65	2.97
Mitianoudis [35]	89.13	93.79	17.90	7.62	2.15	3.47
Jia [5]	84.34	88.11	16.60	9.74	2.52	5.64
1st rank of contest[8]	80.86	-	16.14	-	64.42	104.48
Proposed	91.92	<b>95.09</b>	18.98	4.80	<b>1.37</b>	<b>2.64</b>

executed in parallel. And in this way, the runtime of our method is reduced from 4.45 to 2.46 seconds per megapixel as shown in Table 4.

It is worth mentioning that the optimal global threshold  $t^*$  in Algorithm 1 for most document images is not too high: the highest value among all the images from the seven datasets is 131. Therefore, the proposed algorithm would run faster while the binarization results remain the same if we narrow down the traversal range of global threshold  $t$  appropriately. Actually, the smaller the traversal range, the faster the speed of the algorithm. But the optimal global threshold  $t^*$  may be selected incorrectly when the range is too small. For all the images in seven datasets, the range of 0–190 is a good trade-off between speed and accuracy. In this case, the runtime is 3.73 (2.08 in parallel).

#### 4.3.2. Binarization quality

In our experiments, the six well-known evaluation measures FM, p-FM, PSNR, NRM, MPM and DRD which have been introduced in the Section 4.1 are used to evaluate the various algorithms. The source code of these measures is available in [36]. Tables 5, 6, 7, 8, 9, 10, 11 show the evaluation results obtained by the proposed method and the other binarization algorithms for each dataset. Table 12 shows the mean and standard deviation of the evaluation measures across the seven datasets.

We can see from the Tables 5–11 that our method performs much better than the two traditional threshold binarization methods proposed by Otsu [15] and Sauvola [4] on every dataset in all measures. Similar to the proposed method, the binarization algorithms of Lu [19] and Su [22] are also based on the text

**Table 8**

Comparisons of the performance of the proposed and the other algorithms against H-DIBCO'12 dataset (best values highlighted in bold).

	FM(%)	p-FM(%)	PSNR	NRM( $10^{-2}$ )	MPM( $10^{-3}$ )	DRD
Otsu [15]	75.07	78.14	15.03	9.82	32.10	26.46
Sauvola [4]	81.56	87.35	16.88	11.74	2.66	6.46
Su [22]	89.76	89.61	19.55	6.58	1.93	4.19
Howe [28]	93.73	94.24	<b>21.85</b>	4.15	<b>0.29</b>	<b>2.10</b>
Lelore [23]	<b>94.05</b>	94.42	21.43	<b>3.31</b>	0.43	2.11
Mitianoudis [35]	89.71	92.24	18.73	5.02	1.32	3.88
Jia [5]	88.65	91.92	18.75	8.01	0.69	3.88
1st rank of contest[9]	89.47	90.18	21.80	-	-	3.44
Proposed	92.99	<b>95.10</b>	20.37	3.51	0.46	2.34

**Table 9**

Comparisons of the performance of the proposed and the other algorithms against DIBCO'13 dataset (best values highlighted in bold).

	FM(%)	p-FM(%)	PSNR	NRM( $10^{-2}$ )	MPM( $10^{-3}$ )	DRD
Otsu [15]	80.04	83.43	16.63	9.28	14.58	10.98
Sauvola [4]	82.71	87.74	17.02	8.99	5.62	7.64
Lu [19]	87.08	88.03	18.75	6.91	3.20	4.27
Su [22]	87.70	88.15	19.59	7.35	3.02	4.21
Howe [28]	91.34	91.79	<b>21.29</b>	4.47	3.55	3.18
Lelore [23]	90.78	91.47	20.54	5.08	3.35	3.59
Mitianoudis [35]	91.41	95.47	19.54	5.81	1.11	2.78
Jia [5]	89.50	92.93	19.30	6.47	2.91	3.82
1st rank of contest[10]	91.27	93.37	20.68	5.67	2.39	3.10
Proposed	<b>93.42</b>	<b>96.05</b>	20.78	<b>3.82</b>	<b>0.33</b>	<b>2.03</b>

**Table 10**

Comparisons of the performance of the proposed and the other algorithms against H-DIBCO'14 dataset (best values highlighted in bold).

	FM(%)	p-FM(%)	PSNR	NRM( $10^{-2}$ )	MPM( $10^{-3}$ )	DRD
Otsu [15]	91.62	95.69	18.72	6.09	1.13	2.65
Sauvola [4]	84.70	87.88	17.81	10.36	1.00	4.77
Lu [19]	91.08	91.64	19.71	5.74	0.96	3.08
Su [22]	94.38	95.94	20.31	4.14	0.33	1.95
Howe [28]	96.49	<b>97.38</b>	22.24	2.20	0.33	1.08
Lelore [23]	96.14	96.73	21.88	<b>2.08</b>	0.29	1.25
Mitianoudis [35]	87.57	-	18.43	6.56	-	-
Jia [5]	93.34	96.60	19.51	4.72	0.38	2.15
1st rank of contest[11]	<b>96.88</b>	-	<b>22.66</b>	-	-	<b>0.90</b>
Proposed	94.98	97.18	20.56	2.87	<b>0.23</b>	1.50

**Table 11**

Comparisons of the performance of the proposed and the other algorithms against H-DIBCO'16 dataset (best values highlighted in bold).

	FM(%)	p-FM(%)	PSNR	NRM( $10^{-2}$ )	MPM( $10^{-3}$ )	DRD
Otsu [15]	86.59	89.92	17.79	7.39	5.52	5.58
Sauvola [4]	84.64	88.39	17.09	8.28	5.06	6.27
Su [22]	84.75	88.94	17.64	10.30	<b>4.61</b>	5.64
Howe [28]	87.47	92.28	18.05	7.45	9.30	5.35
Lelore [23]	87.21	88.48	17.36	4.93	7.86	5.27
Mitianoudis [35]	86.89	-	17.60	<b>1.76</b>	-	-
Jia [5]	85.56	91.24	17.57	9.32	10.91	6.11
1st rank of contest[12]	87.61	-	18.11	-	-	5.21
Proposed	<b>90.48</b>	<b>93.27</b>	<b>19.30</b>	4.95	6.90	<b>3.97</b>

**Table 12**

Comparisons of the means and standard deviations (M/STD) of evaluation measures of different methods over all the seven datasets (best values highlighted in bold).

	FM(%)	p-FM(%)	PSNR	NRM( $10^{-2}$ )	MPM( $10^{-3}$ )	DRD
Otsu [15]	82.78/4.75	86.33/5.29	16.67/1.19	7.96/1.42	12.10/9.25	11.61/7.99
Sauvola [4]	82.33/2.97	87.46/1.37	16.68/0.64	10.24/2.63	4.53/2.88	6.85/1.01
Su [22]	89.83/2.88	91.52/2.60	19.14/0.89	6.53/1.84	<b>2.36/1.73</b>	3.67/1.16
Howe [28]	92.49/2.49	93.98/1.72	<b>20.56</b> /1.31	4.21/1.44	3.32/3.29	2.86/1.33
Lelore [23]	92.63/2.48	93.51/2.32	20.27/1.31	<b>3.66</b> /0.93	2.18/2.41	2.74/1.16
Mitianoudis [35]	89.13/1.32	-/-	18.37/0.54	5.29/1.58	-/-	-/-
Jia [5]	88.89/2.68	92.76/2.32	18.44/0.88	7.14/1.75	2.70/3.21	3.94/1.24
Proposed	<b>92.68</b> / <b>1.21</b>	<b>95.15</b> / <b>1.05</b>	19.87/ <b>0.61</b>	3.93/ <b>0.69</b>	<b>1.45</b> /2.09	<b>2.44</b> / <b>0.66</b>

Ravno za to nam je pa toliko treba političnega lista. Samo, da bi se mož zato pripraven našel, kteri bi hotel težavno vredništvo na svoje rame vzeti. Dopisnikov, pravih iskrenih domoljubov, ki poznajo natanko stan našega ljudstva, mu se gotovo manjkalo ne bode. Oni spoznajo ime-

(a)

Ravno za to nam je pa toliko treba političnega lista. Samo, da bi se mož zato pripraven našel, kteri bi hotel težavno vredništvo na svoje rame vzeti. Dopisnikov, pravih iskrenih domoljubov, ki poznajo natanko stan našega ljudstva, mu se gotovo manjkalo ne bode. Oni spoznajo ime-

(c)

Ravno za to nam je pa toliko treba političnega lista. Samo, da bi se mož zato pripraven našel, kteri bi hotel težavno vredništvo na svoje rame vzeti. Dopisnikov, pravih iskrenih domoljubov, ki poznajo natanko stan našega ljudstva, mu se gotovo manjkalo ne bode. Oni spoznajo ime-

(e)

Ravno za to nam je pa toliko treba političnega lista. Samo, da bi se mož zato pripraven našel, kteri bi hotel težavno vredništvo na svoje rame vzeti. Dopisnikov, pravih iskrenih domoljubov, ki poznajo natanko stan našega ljudstva, mu se gotovo manjkalo ne bode. Oni spoznajo ime-

(b)

Ravno za to nam je pa toliko treba političnega lista. Samo, da bi se mož zato pripraven našel, kteri bi hotel težavno vredništvo na svoje rame vzeti. Dopisnikov, pravih iskrenih domoljubov, ki poznajo natanko stan našega ljudstva, mu se gotovo manjkalo ne bode. Oni spoznajo ime-

(d)

Ravno za to nam je pa toliko treba političnega lista. Samo, da bi se mož zato pripraven našel, kteri bi hotel težavno vredništvo na svoje rame vzeti. Dopisnikov, pravih iskrenih domoljubov, ki poznajo natanko stan našega ljudstva, mu se gotovo manjkalo ne bode. Oni spoznajo ime-

(f)

Fig. 12. Binarization results of image PR5 in DIBCO'13. (a)Lelore [23]. (b) Mitianoudis [35]. (c) Lu [19]. (d) Su [22]. (e) Howe [28]. (f) Proposed.

Now if there be any meaning in words at all  
as a sense of it is to give a picture to  
this must signify that the moralconduct of a  
Law is just a Being as this. —

(a)

Now if there be any meaning in words at all  
as a sense of it is to give a picture to  
this must signify that the moralconduct of a  
Law is just a Being as this. —

(c)

Now if there be any meaning in words at all  
as a sense of it is to give a picture to  
this must signify that the moralconduct of a  
Law is just a Being as this. —

(e)

Now if there be any meaning in words at all  
as a sense of it is to give a picture to  
this must signify that the moralconduct of a  
Law is just a Being as this. —

(b)

Now if there be any meaning in words at all  
as a sense of it is to give a picture to  
this must signify that the moralconduct of a  
Law is just a Being as this. —

(d)

Now if there be any meaning in words at all  
as a sense of it is to give a picture to  
this must signify that the moralconduct of a  
Law is just a Being as this. —

(f)

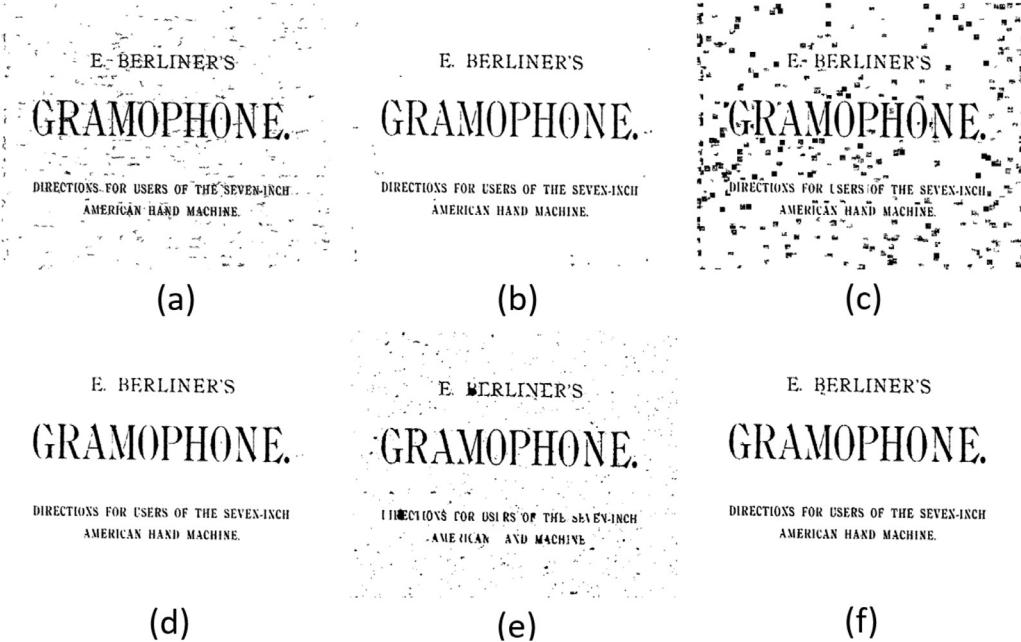
Fig. 13. Binarization results of image HW3 in DIBCO'13. (a)Lelore [23]. (b) Mitianoudis [35]. (c) Lu [19]. (d) Su [22]. (e) Howe [28]. (f) Proposed.

stroke edges. But they use Otsu threshold to detect the edges. Our method outperforms theirs on all the seven datasets especially on DIBCO'11, H-DIBCO'12 and DIBCO'13. As shown in Table 2, these three datasets contain many two-peak degraded images. This fact illustrates the effectiveness of our adaptive global threshold selection algorithm. From the results in the seven tables, we can see that the proposed method is superior to our previous work [5]. As listed in Table 1, the three modifications presented in this paper play an important role in the binarization improvement. Compared with Mitianoudis's [35] binarization method using local features and Gaussian mixture modeling, the performance of our method is better on all the datasets. Meanwhile our performance is superior to that of Howe [28] on the datasets of DIBCO'11, DIBCO'13 and H-DIBCO'16. Compared with Lelore's [23] method, ours performs better on the dataset DIBCO'13 and H-DIBCO'16. Besides, the results show that our method achieves better performance than the 1st rank method of contest on all datasets except the H-DIBCO'14.

In Table 12, the overall evaluation of all the seven datasets is presented. From the mean values shown in the table, we can see that the proposed method achieves the best performance in terms of FM, p-FM, MPM and DRD, while it ranks 3rd and 2nd in terms of PSNR and NRM respectively. Specially, our method obtains an excellent result of MPM: the MPM score obtained by this method is smaller than two thirds of the MPM score in the second place.

As introduced in the Section 4.1, the MPM measures how far the misclassified pixels deviate from the contour of the text in the GT image. This implies that the binarization result of our method has few obvious misclassifications. Considering the robustness of algorithm reflected in the form of standard deviation in Table 12, our method obtains the minimum value in terms of five measures. This demonstrates the robustness of our method. And it further indicates that the three proposed strategies including global threshold selection, stroke width estimation and voting framework make this method to have good adaptability and generality.

Figs. 12–14 further show some binarization results produced by different methods. Their corresponding original images have been introduced in Figs. 4 and 6. The binarization results produced by the proposed method are acceptable for all the three images based on visual criteria. Specifically, the advantage of our method is more obvious on the first image. We can see from these experimental results that other binarization methods perform differently for different images, whereas there is little difference between the performance of the proposed method on images with different degradations, which shows the robustness of our method. Indeed, due to the multiple adaptive procedures, such as the global threshold selection, stroke width estimation, voting framework and so on, our algorithm could handle the various degradations and achieve good binarization performance.



**Fig. 14.** Binarization results of image PR6 in DIBCO'11. (a)Lelore [23]. (b) Mitianoudis [35]. (c) Lu [19]. (d) Su [22]. (e) Howe [28]. (f) Proposed.

## 5. Conclusion

In this paper, we present a novel and effective local threshold binarization method based on SSP for degraded document images. There are three key strategies to improve the binarization result in this paper. First, for the SSP extraction, we propose an adaptive global threshold selection algorithm which can remove the bleed-through background while preserving the faint characters. Then the proposed stroke width estimation algorithm improves the adaptiveness of our binarization method. Finally, we use the multiple local threshold vote based framework to compensate the inaccurate detected SSPs and eliminate the boundary noise.

We test the method on seven public datasets covering numerous types of degradations. The results demonstrate that our method achieves promising performance compared with many traditional and state-of-the-art document binarization approaches based on various evaluation measures. Besides, there is little difference between the results we obtained using the proposed method on different datasets which shows the robustness of our method. Although the estimation of stroke width is more suitable for the printed text, it indeed works well on both printed and handwritten document images under the same experimental assignment. In summary, our method is effective, robust, and capable of handling machine-printed and handwritten images with various degradations thanks to the multiple adaptive strategies, such as the background removal, global threshold selection, stroke width estimation and the voting framework.

## Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant No. 61601462, No. 61531019 and No. 71621002.

## References

- [1] Y. Wu, P. Natarajan, S. Rawls, W. AbdAlmageed, Learning document image binarization from data, in: IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 3763–3767.
- [2] X. Chen, L. Lin, Y. Gao, Parallel nonparametric binarization for degraded document images, Neurocomputing 189 (2016) 43–52.
- [3] W. Niblack, An introduction to digital image processing, Strandberg Publishing Company, 1985.
- [4] J. Sauvola, M. Pietikäinen, Adaptive document image binarization, Pattern Recognit. 33 (2) (2000) 225–236.
- [5] F. Jia, C. Shi, K. He, C. Wang, B. Xiao, Document image binarization using structural symmetry of strokes, in: International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2016, in press
- [6] B. Gatos, K. Ntirogiannis, I. Pratikakis, ICDAR 2009 document image binarization contest (DIBCO 2009), in: International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2009, pp. 1375–1382.
- [7] I. Pratikakis, B. Gatos, K. Ntirogiannis, H-DIBCO 2010-handwritten document image binarization competition, in: International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2010, pp. 727–732.
- [8] I. Pratikakis, B. Gatos, K. Ntirogiannis, ICDAR 2011 document image binarization contest (DIBCO 2011), in: International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2011, pp. 1506–1510.
- [9] I. Pratikakis, B. Gatos, K. Ntirogiannis, ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012), in: International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2012, pp. 817–822.
- [10] I. Pratikakis, B. Gatos, K. Ntirogiannis, ICDAR 2013 document image binarization contest (DIBCO 2013), in: International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2013, pp. 1471–1476.
- [11] K. Ntirogiannis, B. Gatos, I. Pratikakis, ICFHR 2014 competition on handwritten document image binarization (H-DIBCO 2014), in: International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2014, pp. 809–813.
- [12] I. Pratikakis, K. Zagoris, G. Barlas, B. Gatos, Icfhr2016 handwritten document image binarization contest (H-DIBCO 2016), in: International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2016, pp. 619–623.
- [13] J. Kaur, R. Mahajan, A review of degraded document image binarization techniques, Int. J. Adv. Res. Comput. Commun. Eng. 3 (5) (2014) 6581–6586.
- [14] S. Pardhi, G. Kharat, An improved binarization method for degraded document, in: National Conference MOMENTUM, 17, 2017.
- [15] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 19 (1) (1979) 62–66.
- [16] J. Kittler, J. Illingworth, Minimum error thresholding, Pattern Recognit. 19 (1) (1986) 41–47.
- [17] J. Bernsen, Dynamic thresholding of grey-level images, in: International Conference on Pattern Recognition (ICPR), 1986, pp. 1251–1255.
- [18] Z. Hadjadj, M. Cheriet, A. Meziane, Y. Cherfa, A new efficient binarization method: application to degraded historical document images, Signal Image Video Process. (2017) 1–8.
- [19] S. Lu, B. Su, C.L. Tan, Document image binarization using background estimation and stroke edges, Int. J. Document Anal. Recognit. (IJDAR) 13 (4) (2010) 303–314.
- [20] S. Lu, C.L. Tan, Binarization of badly illuminated document images through shading estimation and compensation, in: International Conference on Document Analysis and Recognition (ICDAR), 1, IEEE, 2007, pp. 312–316.
- [21] B. Su, S. Lu, C.L. Tan, Binarization of historical document images using the local maximum and minimum, in: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, ACM, 2010, pp. 159–166.
- [22] B. Su, S. Lu, C.L. Tan, Robust document image binarization technique for degraded document images, IEEE Trans. Image Process. 22 (4) (2013) 1408–1417.

- [23] T. Lelore, F. Bouchara, Fair: a fast algorithm for document image restoration, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 2039–2048.
- [24] C.-H. Chou, W.-H. Lin, F. Chang, A binarization method with learning-built rules for document images produced by cameras, *Pattern Recognit.* 43 (4) (2010) 1518–1530.
- [25] B. Gatos, I. Pratikakis, S.J. Perantonis, Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information, in: *International Conference on Pattern Recognition (ICPR)*, IEEE, 2008, pp. 1–4.
- [26] K. Ntirogiannis, B. Gatos, I. Pratikakis, A combined approach for the binarization of handwritten document images, *Pattern Recognit. Lett.* 35 (2014) 3–15.
- [27] N.R. Howe, A Laplacian energy for document binarization, in: *International Conference on Document Analysis and Recognition (ICDAR)*, IEEE, 2011, pp. 6–10.
- [28] N.R. Howe, Document binarization with automatic parameter tuning, *Int. J. Doc. Anal. Recognit. (IJDAR)* 16 (3) (2013) 247–258.
- [29] A. Mishra, K. Alahari, C. Jawahar, Unsupervised refinement of color and stroke features for text binarization, *Int. J. Doc. Anal. Recognit. (IJDAR)* (2017) 1–17.
- [30] T. Lelore, F. Bouchara, Document image binarisation using markov field model, in: *International Conference on Document Analysis and Recognition (ICDAR)*, IEEE, 2009, pp. 551–555.
- [31] C. Shi, B. Xiao, C. Wang, Y. Zhang, Adaptive graph cut based binarization of video text images, in: *IAPR International Workshop on Document Analysis Systems (DAS)*, IEEE, 2012, pp. 58–62.
- [32] B. Su, S. Lu, C.L. Tan, A learning framework for degraded document image binarization using Markov random field, in: *International Conference on Pattern Recognition (ICPR)*, IEEE, 2012, pp. 3200–3203.
- [33] B. Su, S. Lu, C.L. Tan, Combination of document image binarization techniques, in: *International Conference on Document Analysis and Recognition (ICDAR)*, IEEE, 2011, pp. 22–26.
- [34] Y. Chen, L. Wang, Broken and degraded document images binarization, *Neurocomputing* (2017).
- [35] N. Mitianoudis, N. Papamarkos, Document image binarization using local features and gaussian mixture modeling, *Image Vision Comput.* 38 (2015) 33–51.
- [36] R.F. Moghaddam, H.Z. Nafchi, Objective evaluation of binarization methods [online], 2013, (<http://www.mathworks.com/matlabcentral/fileexchange/27652/>).

**Fuxi Jia** received the B.S. degree in Information Engineering from Harbin Engineering University in 2014. She is now a Ph.D. candidate in Pattern Recognition and Artificial Intelligence in the Institute of Automation, Chinese Academy of Sciences, Beijing, China. Her research interests include document image analysis, text segmentation and recognition.

**Cunzhao Shi** is an Assistant Professor at the Institute of Automation, Chinese Academy of Sciences. She received the B.S. degree in Electronic Engineering from Wuhan University, Wuhan, China, in 2009 and Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China in 2014. Her research interests include pattern recognition, computer vision, text detection, text extraction, character recognition and text recognition in natural images.

**Kun He** received the B.S. degree in Shandong University in 2014. He is now a Master candidate in Pattern Recognition and Artificial Intelligence in the Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include text detection and recognition in natural images.

**Chunheng Wang** received the B.S. degree and the M.E. degree in Electronic Engineering from Dalian University of Technology, Dalian, China and the Ph.D. degree in Pattern Recognition and Intelligent control from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1993, 1996 and 1999, respectively. From 2004, he has been a Professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include pattern recognition, image processing, neural networks and machine learning.

**Baihua Xiao** received the B.S. degree in Electronic Engineering from Northwestern Polytechnical University, Xian, China and the Ph.D. degree in Pattern Recognition and Intelligent control from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1995 and 2000, respectively. From 2005, he has been a Professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include pattern recognition, image processing and machine learning.