


RESEARCH

Open Access



An enhanced binarization framework for degraded historical document images

Wei Xiong^{1,2*} , Lei Zhou¹, Ling Yue¹, Lirong Li¹ and Song Wang^{2*}

* Correspondence: xw@mail.hbut.edu.cn; songwang@cec.sc.edu

¹School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, Hubei, China

²Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29201, USA

Abstract

Binarization plays an important role in *document analysis and recognition* (DAR) systems. In this paper, we present our winning algorithm in ICFHR 2018 *competition on handwritten document image binarization* (H-DIBCO 2018), which is based on background estimation and energy minimization. First, we adopt mathematical morphological operations to estimate and compensate the document background. It uses a disk-shaped structuring element, whose radius is computed by the minimum entropy-based *stroke width transform* (SWT). Second, we perform Laplacian energy-based segmentation on the compensated document images. Finally, we implement post-processing to preserve text stroke connectivity and eliminate isolated noise. Experimental results indicate that the proposed method outperforms other state-of-the-art techniques on several public available benchmark datasets.

Keywords: Document image binarization, Document image segmentation, Background estimation and compensation, Laplacian energy minimization, Minimum entropy-based *stroke width transform* (SWT), *Markov random fields* (MRFs), Graph cut

1 Introduction

Historical documents are precious cultural heritage and have important scientific and cultural values. The digitization of ancient manuscripts is an important way to address literature protection and cultural heritage. However, it takes time and effort to manually process these large volumes of documents, and is error-prone. Therefore, it is necessary to use computers to process historical manuscripts automatically. The *document analysis and recognition* (DAR) system has emerged at this purpose. It consists of image enhancement, segmentation, page layout analysis, *optical character recognition* (OCR), and indexing [1]. Document image binarization (also referred to as document image segmentation) is an important preprocessing step. It aims to segment the input document image into text (foreground) and non-text (background). The segmentation performance will directly affect subsequent tasks in the DAR system.

The thresholding of high-quality images is very simple, but the binarization of historical document images is quite challenging because historical document images are subject to severe degradation, such as ink bleed through, page stains, text stroke fading, and artifacts. In addition, changes of text stroke color, width, brightness, and

connectivity in degraded handwritten manuscripts further increase the difficulty of binarization. Figure 1 presents several degraded historical document image samples in recent DIBCO (*document image binarization competition*) and H-DIBCO (*handwritten document image binarization competition*) benchmark datasets.

The DIBCO and H-DIBCO series (DIBCO 2009 [2], 2011 [3], 2013 [4], and 2017 [5] and H-DIBCO 2010 [6], 2012 [7], 2014 [8], 2016 [9], and 2018 [10]) show the latest progress in document image binarization. We have participated in such academic competitions since 2017. Our energy-based segmentation method achieved the best performance in ICFHR 2018 competition on handwritten document image binarization [10], and the 2nd best performance in Challenge A of ICFHR 2018 competition on document image analysis tasks for Southeast Asian palm leaf manuscripts [11]. Later, our newly developed binarization method based on D-LinkNet [12] achieved the best performance in ICDAR 2019 time–quality binarization competition on photographed document images captured by Motorola Z1 and Galaxy Note4 with flash off, and the 2nd and 3rd best performances on binarization of photographed document images captured by the same mobile devices with flash on, respectively [13].

This paper presents our winning algorithm in ICFHR 2018 competition on handwritten document image binarization (H-DIBCO 2018). The proposed method is based on background estimation and energy minimization. As far as we know, the estimation and compensation procedure can eliminate most document degradation, and help extract text objects from complex document background in the following energy-based segmentation stage.

Our contributions are two folds. First, we present a document image binarization method that can achieve promising pixel-wise labeling results on various degraded historical document images. Second, we investigate a voting strategy to automatically



Fig. 1 Historical document image samples in recent DIBCO and H-DIBCO benchmark datasets

determine the correct directions for *stroke width transform* (SWT). The SWT direction determination has so far received little attention, but if done well, it offers many advantages. This method is simple and robust since users do not need to predefine document types, e.g., dark text on bright background or vice versa.

The rest of the paper is organized as follows. Section 2 reviews the related work on document image binarization. Section 3 introduces our proposed technique in detail. Section 4 presents the experimental results and discussion, and Section 5 concludes the paper.

2 Related work

Varieties of document image binarization or segmentation methods have been proposed over the past few decades. They can be roughly divided into global thresholding, local thresholding, and hybrid methods [1, 14].

A *global thresholding* approach, e.g., Otsu's method [15] computes an optimal threshold for the entire image to maximize the inter-class variance or equivalently minimize the intra-class variance between text and non-text pixels. Global thresholding can provide satisfactory results when the image quality is good enough, that is, the image histogram follows a bimodal distribution, but it will generally fail when dealing with low-quality images.

A *local thresholding* approach adapts the threshold value of each pixel to its neighborhood image features; for instance, local mean and local standard deviation are required for Niblack's [16], Sauvola's [17], and Wolf's [18] methods. In general, locally adaptive thresholding methods have better performance than global thresholding counterparts. However, the main disadvantages of these local methods are that the thresholding performance depends heavily on the sliding window size and hence on the text stroke width.

Ntirogiannis et al. [19] propose a *hybrid* method. First, Niblack's method is used for document background estimation via image inpainting, and then image normalization is adopted to compensate background variations. Otsu's method is then applied on the compensated image to remove background noise. Meanwhile, Niblack's method is also performed on the normalized image to detect faint characters and estimate the text stroke width. The two binarization results are finally combined at connected component level.

Su et al. [20] also present a *combined* framework, which integrates several existing document binarization methods to achieve better segmentation. This method divides image pixels into three groups, i.e., foreground, background, and uncertain pixels. Based on preselected foreground and background pixels, a classifier is then applied to iteratively classify the uncertain pixels as foreground or background.

In the rest of this section, we classify other document image binarization methods into following categories:

2.1 Contrast or edge-based segmentation

Early studies of document image binarization are often based on edge detection. Lu et al. [21] present a segmentation technique using *background estimation and stroke edges* (BESE). This method first uses two one-dimensional polynomial smoothing

procedures to estimate the document background, and then detects text stroke edges from the compensated document image based on the L1-norm image gradient. Finally, text strokes can be extracted based on the detected stroke edge pixels. Lu and Tan [22] also studied a similar technique for document background estimation via two-dimensional polynomial smoothing.

Su et al. [23] propose a binarization technique for historical document images. It first uses *local maximum and minimum* (LMM) to construct a contrast image, and then high-contrast pixels are extracted by using Otsu's method. Therefore, the document text pixels can be further segmented based on the detected high-contrast image pixels, which are located near the text strokes. Later, Su et al. [24] present a degraded document image binarization method based on adaptive image contrast, which is a combination of LMM and local gradient. The adaptive contrast image is first binarized by Otsu's method. The resulting binary contrast map is then combined with the Canny edge map to produce text stroke edges. Finally, the text pixels can be extracted based on the detected stroke edge pixels.

Jia et al. [25] present a document image binarization method based on *structural symmetric pixels* (SSPs), which are located along text stroke edges, and can be extracted from those with large gradient values and opposite gradient directions. Finally, a voting framework based on multiple local thresholds is adopted to further determine whether each pixel is text or non-text.

2.2 Energy-based segmentation

Markov random fields (MRFs) [26] and *conditional random fields* (CRFs) [27] are widely used in degraded document image binarization. Howe [28, 29] presents an energy-based segmentation that uses graph cut optimization [30] to solve the energy minimization problem of an objective function which combines the Laplacian operator and Canny edge detector. A fast algorithm for Howe's binarization method based on heterogeneous computing platform is implemented by Westphal et al. [31].

Mesquita et al. [32] present a document image binarization algorithm based on the *perception of objects by distance* (POD). The k -means clustering, and Otsu's thresholding methods are combined in the classification process. Later, Mesquita et al. [33] adopt the POD (with parameters tuned by I/F-Race) as a preprocessing stage of Howe's binarization method.

Kligler et al. [34] propose a document enhancement technique based on visibility detection. The main idea of the algorithm is to convert an image to a 3D point cloud. The classification of visible and invisible points provides guidance for background and foreground segmentation.

Another approach based on energy generalization, active contour model or snakes, is also used for document image binarization. Rivest-Hénault et al. [35] propose a local linear level set framework, and Hadjadj et al. [36] present a technique based on active contour evolution. The snakes model generally has a high time complexity and a tendency to fall into the nearest local minimum.

The method proposed in this paper belongs to this category. Like Mesquita's and Kligler's approaches, we integrate Howe's energy-based segmentation technique into our framework, but with several subtle and important differences described in

Subsection 3.3. Document background estimation and compensation are performed in the preprocessing stage, while de-noising and text stroke preservation are performed in the post-processing stage.

2.3 Statistical learning-based segmentation

Chen et al. [37] propose a parallel non-parametric binarization framework for degraded document images. This method first uses Sauvola's method with different parameters to generate many binary images. A *support vector machine* (SVM) is then used to recognize each pixel of those binarized images. Finally, the resulting binary image is reconstructed based on the recognition result.

After conducting local contrast enhancement, Xiong et al. [38] divide the document image into 5×5 blocks, and then adopt a SVM classifier to choose an optimal global threshold for each block. The document image is further segmented by Wolf's method to eliminate noise near text stroke edges.

Bhowmik et al. [39] present a *game theory inspired binarization* (GiB) technique for degraded document images. It first extracts pixel-level features using neighbor's collision, and then classifies each pixel as either text or non-text using the k -means clustering method.

In general, the main drawback of statistical learning-based methods is that only hand-crafted features are used to obtain segmentation results. Therefore, it is difficult to design representative features for different applications, and manually designed features work well for one type of images, but often fail on another.

2.4 Deep learning-based segmentation

Pastor-Pellicer et al. [40] explore the use of *convolutional neural networks* (CNNs) for document image binarization. It includes several convolutional and subsampling layers followed by *multilayer perceptron* (MLP) layers, and then classifies each pixel as foreground or background from a sliding window.

Tensmeyer and Martinez [41] present a multi-scale *fully convolutional network* (FCN) that combines F-measure and pseudo F-measure losses for document image binarization tasks. The raw gray scale, Howe's binarization, and *relative darkness* (RD) features are concatenated and fed into the networks.

Vo et al. [42] propose a supervised binarization for historical document images based on hierarchical *deep supervised networks* (DSNs). By extracting both low-level and high-level features, the networks can differentiate text pixels from background noise, and thus can deal with severe degradations occurring in document images.

Calvo-Zaragoza and Gallego [43] choose the *residual encoder-decoder network* (RED-Net) [44] as the backbone of their *selectional auto-encoder* (SAE) architecture for document image binarization. The encoder contains 5 convolution layers, while the decoder contains 5 transposed convolution layers, each with a stride value of 2. The RED-Net has an input image patch size of 256×256 , the number of filters in the first layer is 64, and the kernel size of all layers is 7×7 .

Bezmaternykh et al. [45] present a historical document image binarization method based on U-Net [46], originally designed for biomedical image segmentation. The U-

Net architecture is derived from FCNs, but its architecture has been modified and extended to use fewer training images and produce more accurate segmentation.

Zhao et al. [47] formulate binarization as an image-to-image generation task, using *conditional generative adversarial networks* (cGANs) to solve multi-scale information combination problems in binarization tasks.

It is worth mentioning that deep learning is a subset of machine learning, which combines feature learning and metric learning in a deep network model. Although the purpose of metric learning is to reduce the distance between similar sample features while increasing the distance between different sample features, the intrinsic difference between deep learning-based methods and other non-deep learning-based approaches is that the former can be trained to automatically extract both fine-grained, shallow, low-level visual features and coarse-grained, deep, high-level semantic features, while the latter can only use handcrafted or manually designed features to obtain segmentation results, and no training is required.

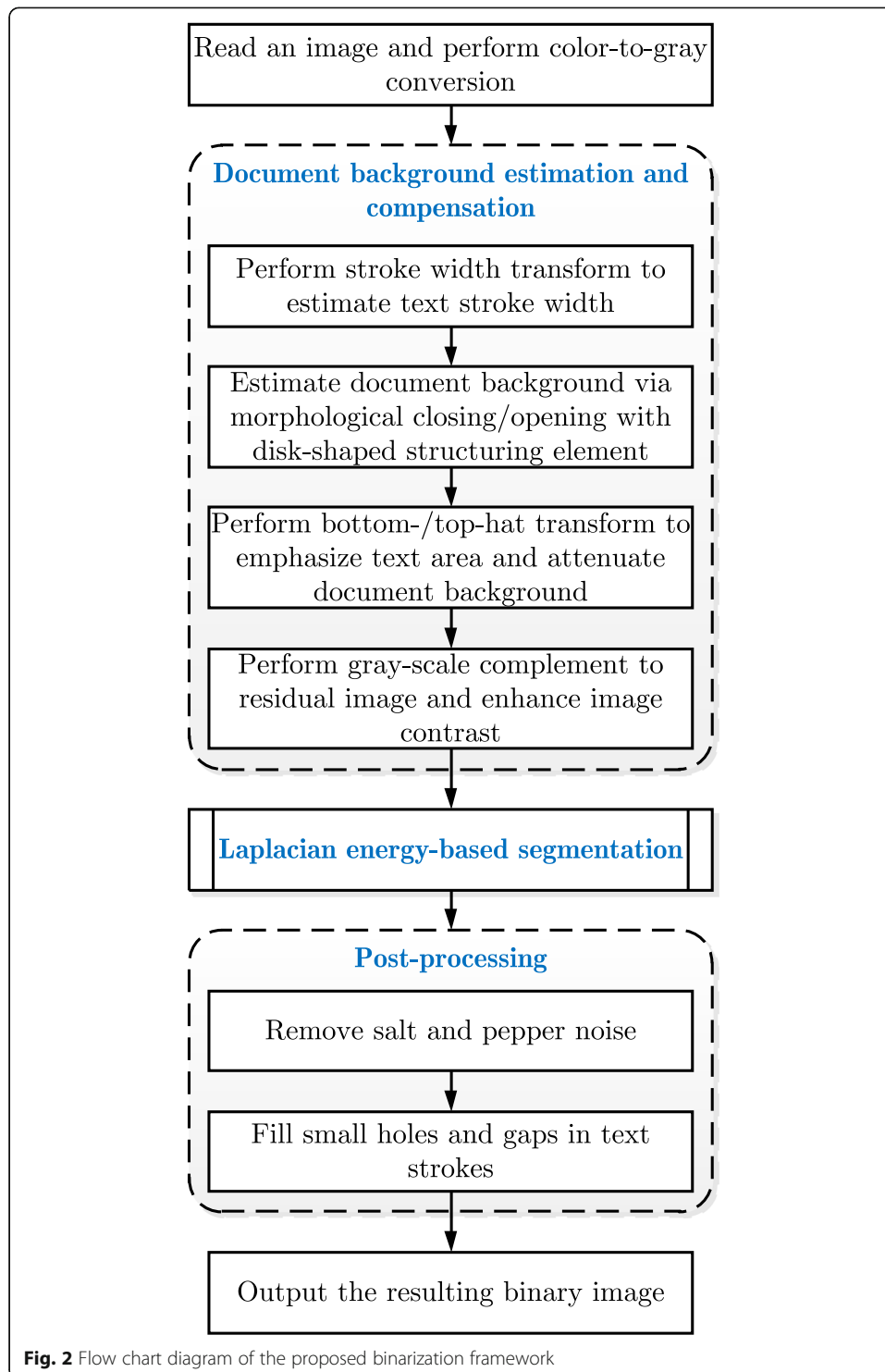
3 Motivation and proposed method

Figure 2 depicts our proposed binarization framework for degraded historical document images based on background estimation and energy minimization. It consists of three main steps. First, we apply morphological operations on gray scale images to estimate and compensate document background. It utilizes a disk-shaped structuring element, whose radius is estimated by a minimum entropy-based *stroke width transform* (SWT). Second, we perform the Laplacian energy-based segmentation on the compensated document images. Finally, we implement post-processing to preserve text stroke connectivity and eliminate isolated noise.

The motivations behind this method are as follows: First, historical document images generally contain severe degradation, such as page stains, ink bleed through, text stroke fading, and artifacts, which is not conducive to the correct extraction of text pixels from the images. The document background estimation and compensation technique can effectively eliminate the impacts of these degradation factors. Second, inspired by the image information entropy, the minimum entropy-based SWT is able to automatically detect the document image type, for instance, dark text on bright background or bright text on dark background. Third, the graph cut is a group of MRF algorithms that uses max-flow and min-cut algorithms to solve discrete energy minimization problems and has been widely used in many image analysis tasks, such as image restoration and reconstruction, edge detection, texture segmentation, optical flow, and stereo vision. Last but not least, we combine the use of preprocessing and post-processing, which has proven to be the gold standard for document image binarization.

3.1 Stroke width transform with minimum entropy

Text stroke width is a crucial attribute that can distinguish true text from possible degradation. Most previous approaches perform a locally adaptive thresholding (e.g., the Sauvola's or Niblack's method) on a given document image, and then estimate text stroke width by using the resulting binary image. We take a different approach and use Canny edge detector to generate text edge maps by extracting main edge features of



the input image while minimizing irrelevant details such as various types of degradation described above.

We first apply a Gaussian filter with a sigma value $\sigma = 1$ on the gray scale document image, and then compute the magnitude $(g_x^2 + g_y^2)^{\frac{1}{2}}$ and direction $\tan^{-1}\left(\frac{g_y}{g_x}\right)$ of the local

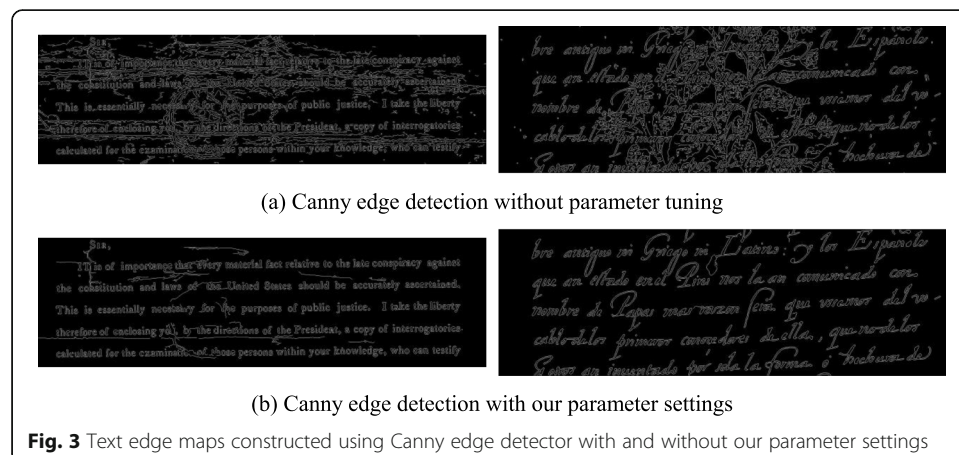
gradient at each pixel. An edge pixel is defined as the local maxima of the image gradient and determined by non-maximum suppression along the direction of the image gradient. The algorithm chooses to use a hysteresis thresholding with two thresholds (t_{high} and t_{low}) to preserve those true edges. Based on observations that true text edges often have higher contrast than possible degradation, so the following parameter settings can be used as reasonable default values in our implementation: $t_{\text{high}} = 0.4$ and $t_{\text{low}} = 0$.

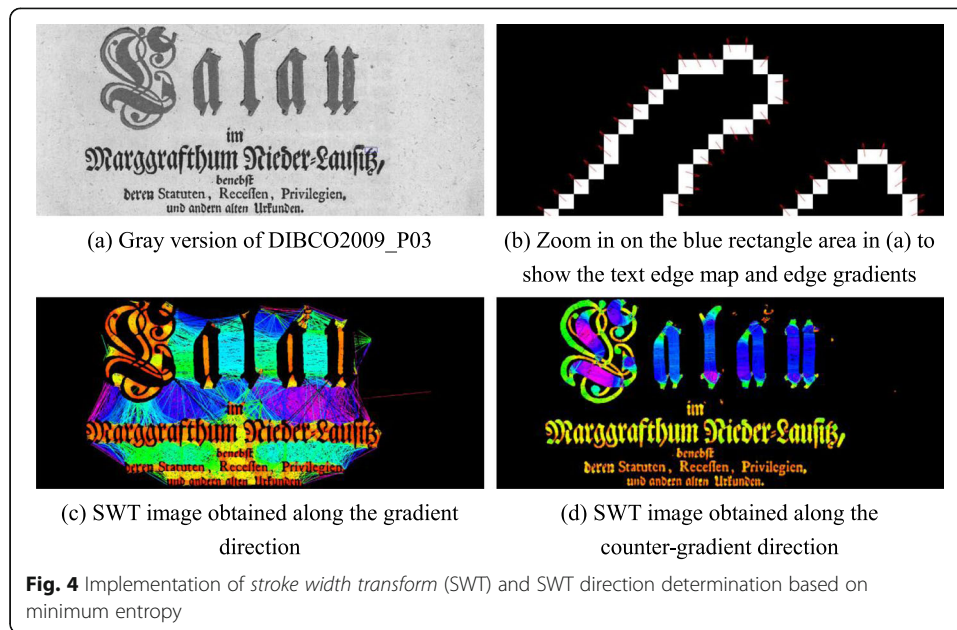
Figure 3 displays text edge maps of the sample document images in Figs. 1e and d, which are constructed by using Canny edge detector with and without our parameter settings, respectively. It can be seen from the figures that the text edge maps, obtained by using the default syntax for “canny” options in Matlab and without parameter tuning, extract a large number of non-text edges. But the optimized Canny edge detection with our specified parameter settings produces a cleaner text edge map.

After Canny edge detection, we can estimate the text stroke width from the detected text edge pixels and the directed edge gradients. It has been observed that (1) the text stroke width or its mathematical expectation remains nearly constant throughout individual characters, and (2) the gradient direction of each edge pixel is approximately perpendicular to the direction of the text stroke. Therefore, the text stroke width can be estimated along the gradient or counter-gradient directions.

The proposed technique adopts the similar idea of stroke width transform presented in [48], but deviates from the original in several ways. If text pixels are darker than background pixels as illustrated in Fig. 4a, the edge gradients will point to the exterior of strokes as shown in Fig. 4b; therefore, the search path is opposite to the gradient direction. However, if text pixels are brighter than background pixels, the edge gradients will point to the interior of strokes, and then the search path will follow the gradient direction. In order to detect either dark text on bright background or vice versa, the algorithm can be executed twice in parallel on the same image to achieve that effect, but the original paper does not inform how to implement.

Figure 4c and d illustrate the resulting stroke width transform along the gradient direction, referred to as SWT_{grad} , and along counter-gradient direction, referred to as $\text{SWT}_{\text{cnt-grad}}$, respectively. Each color in the SWT_{grad} or $\text{SWT}_{\text{cnt-grad}}$ image corresponds to a specific stroke width (black corresponds to background), so pixels with the same





stroke width are represented with the same color. We can see that $SWT_{\text{cnt-grad}}$ in this example is more compact and contains less colors than SWT_{grad} . This is reasonable since $SWT_{\text{cnt-grad}}$ corresponds to the true text regions with a uniform stroke width distribution, but SWT_{grad} corresponds to the non-text regions with randomly distributed “strokes”.

Inspired by the above observations, we propose a minimum entropy-based technique to help determine the correct SWT direction. Specifically, the entropy S is defined as a logarithmic measure of the number of connected components with significant probability of being occupied:

$$S = -s_w \sum_i p_i \log p_i \quad (1)$$

where p_i is the inverse of the number of connected components, and s_w equals to the mathematical expectation of stroke widths in the corresponding SWT image. The summation is over all the connected components of the SWT image. We modify the conventional connected component labeling algorithm [49] by converting the association rule from a binary mask to a predication that compares the SWT values of neighboring pixels. If two pixels are adjacent and have similar stroke width values, they are in the same connected component, and it can be empirically verified that the two neighboring pixels belong to the same component if the SWT ratio does not exceed 3. This local rule ensures that strokes with a smooth width change will be grouped into the same component, and therefore be robust to various text sizes, fonts, and orientations.

As mentioned before, we perform the SWT algorithm twice in parallel, once along the gradient direction and once along the counter-gradient direction, and then two entropies S_{grad} and $S_{\text{cnt-grad}}$ are computed. We vote to determine the correct SWT direction in the following fashion:

$$\text{SWT}_{\text{correct-direction}} = \arg \min \{S_{\text{grad}}, S_{\text{cnt-grad}}\} \quad (2)$$

Consequently, the minimum entropy corresponds to the correct SWT direction, and in our implementation, the text stroke width is computed as the average of the corresponding non-zero stroke widths.

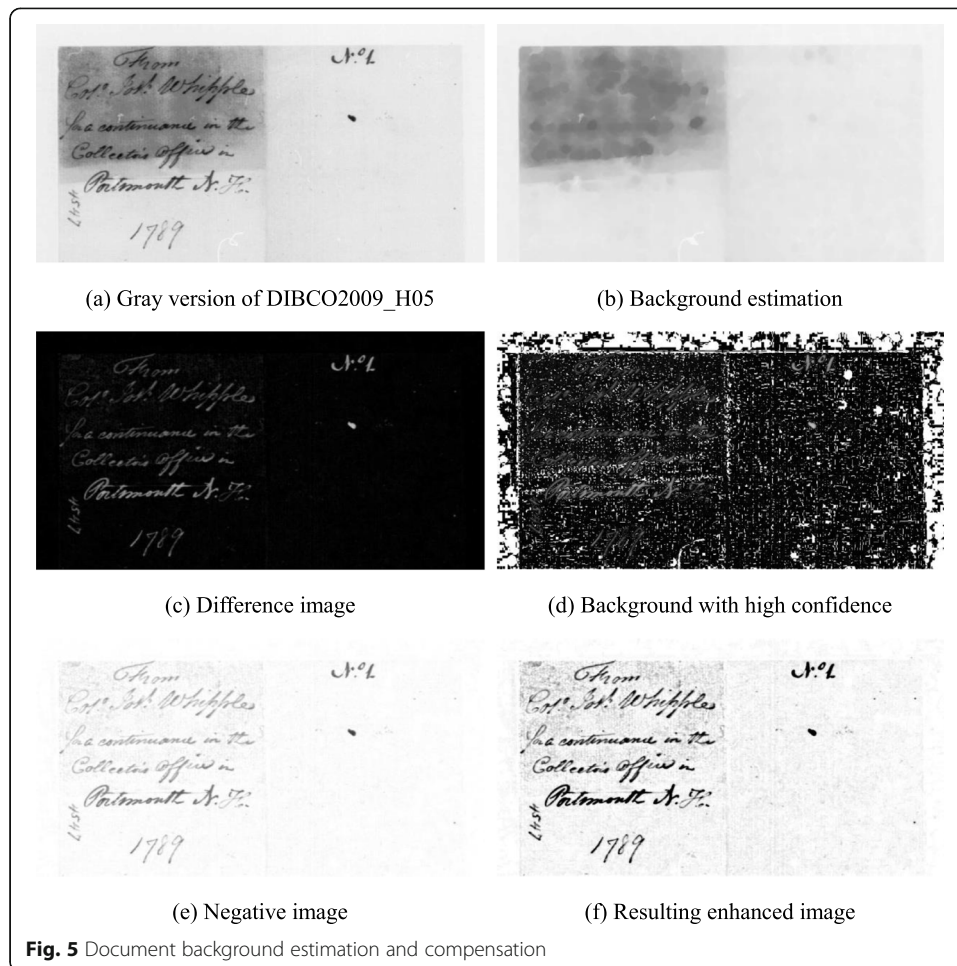
3.2 Document background estimation and compensation

Mathematical morphology has been used for background estimation. Yang et al. [50] compare various segmentation and background estimation methods used on cDNA microarray data. They find that morphological opening provides a more reliable estimation of background than other methods. Su et al. [23] adopt gray scale morphological dilation and erosion, which are referred to as maximum and minimum in the original paper, to eliminate the document background. Mesquita et al. [32, 33] make use of morphological closing and image resizing to estimate the document background based on POD, initially proposed by Mello [51].

We implement the mathematical morphology for document background estimation and compensation in a different way. The document background is estimated by using gray scale morphological opening or closing according to previous SWT direction determination. The structuring element is a disk-shaped mask. If text pixels are darker than document background as illustrated in Fig. 5a, the morphological closing is used as illustrated in Fig. 5b. If text pixels are brighter than background pixels, the morphological opening is then used. In this context, the morphological closing operator can suppress dark details that are smaller than the structuring element, while the morphological opening operator can suppress bright details that are smaller than the structuring element. Therefore, the size of the structuring element should be larger than the width of the text stroke, and the radius parameter settings will be discussed in Subsection 4.3.

Morphological closing or opening can produce reasonable document background estimation for the entire image, and then we perform a morphological bottom-hat or top-hat transform to emphasize the text area and attenuate the document background. The bottom-hat transform, as illustrated in Fig. 5c, is defined as the difference between the closing and the gray scale images, while the top-hat transform is the difference between the gray scale image and its opening. The two related operations produce exactly the same images, and therefore, the subsequent algorithm will no longer distinguish the two types of documents accordingly.

In a difference image, background pixels are referred to as those whose intensity values are equal to 0. We convert the background pixels into white (assigned the maximum pixel value of 255), as illustrated in Fig. 5d, and then apply the gray scale image complement to the remaining pixels (subtracted from the maximum pixel value), as illustrated in Fig. 5e. Finally, we adjust the image contrast so that 1% of the image data is saturated at low and high intensities. Figure 5f depicts the resulting image of this preprocessing stage applied to the original image of Fig. 5a.



We can see that the document background has been compensated and the contrast between foreground and background pixels has also been enhanced.

3.3 Laplacian energy-based segmentation

Once the document background estimation and compensation is completed, we then perform *Markov random fields* (MRFs) for image segmentation. The MRF models have been widely used to solve both low-level and high-level vision problems, including document image binarization [26, 52, 53].

Howe's methods [28, 29] regard document image binarization as a max-flow and min-cut graph optimization problem [30]. The unary terms are determined by the Laplace operator, and the pairwise terms are given by the Canny edge detector. The exact optimal solution can be obtained by finding the maximum flow on a special defined graph network, constructed according to the energy function.

Due to the superior performance of Howe's binarization method [28], we decide to integrate it as part of our framework, but there are several subtle and important differences. Given an $h \times w$ gray scale image I , the quadratic pseudo-Boolean energy function can be defined as

$$\begin{aligned} \mathcal{E}_I(B) = & \sum_{i=0}^h \sum_{j=0}^w [L_{ij}^0 \bar{B}_{ij} + L_{ij}^1 B_{ij}] + \sum_{i=0}^{h-1} \sum_{j=0}^w C_{ij}^H (B_{ij} \neq B_{i+1,j}) \\ & + \sum_{i=0}^h \sum_{j=0}^{w-1} C_{ij}^V (B_{ij} \neq B_{i,j+1}) \end{aligned} \tag{3}$$

where $\bar{B}_{ij} = 1 - B_{ij}$ is the negation of $B_{ij} \in \{0, 1\}$ at (i, j) , L_{ij}^0 and L_{ij}^1 denote the costs of assigning label 0 or 1 to each pixel, and C_{ij}^H and C_{ij}^V denote the costs of label mismatch between B_{ij} and its horizontal or vertical neighbors, respectively.

The unary potentials L_{ij}^0 and L_{ij}^1 are given by the Laplace operator ∇^2 :

$$L_{ij}^0 = \nabla^2 I_{ij} \tag{4}$$

$$L_{ij}^1 = \begin{cases} \phi, & \text{background with high confidence} \\ -\nabla^2 I_{ij}, & \text{otherwise} \end{cases} \tag{5}$$

We set background pixels with high confidence, found in the previous stage, to a negative constant ϕ , which is twice the maximum pixel value in our implementation.

The pairwise potentials C_{ij}^H and C_{ij}^V are given by the Canny edge detector:

$$C_{ij}^H = \begin{cases} 0, & E_{ij} \wedge (I_{i-1,j} \geq I_{ij}) \vee E_{ij} \wedge (I_{ij} < I_{i+1,j}) \\ \psi, & \text{otherwise} \end{cases} \tag{6}$$

$$C_{ij}^V = \begin{cases} 0, & E_{ij} \wedge (I_{i,j-1} \geq I_{ij}) \vee E_{ij} \wedge (I_{ij} < I_{i,j+1}) \\ \psi, & \text{otherwise} \end{cases} \tag{7}$$

where E_{ij} denotes the Canny edges, and non-edge pixels with label mismatch are set to a positive constant ψ . We have noticed that, among all the parameters, the high threshold t_{high} and the neighbor mismatch penalty ψ are the most important; therefore, we follow the automatic parameter tuning strategy proposed in [29].

3.4 Post-processing

After obtaining the segmentation based on Laplacian energy minimization, we proceed to the post-processing stage to preserve text stroke connectivity and eliminate isolated noise. Our post-processing algorithm is described in detail below.

3.4.1 Step 1

We perform foreground *connected component analysis* (CCA) to eliminate isolated noise from document background. The CCA operator scans the binary image and examines each foreground connected component. When it comes to an unlabeled foreground pixel p , we use the flood fill algorithm [54] to label all other pixels in the connected component that contains p . After completing the scan, we count the number of pixels in each connected component and generate a binary image B_1 with an area greater than t_{noise} , where t_{noise} is the area threshold for isolated noise.

3.4.2 Step 2

We perform background CCA to fill small holes in text strokes, which may preserve text stroke connectivity. Consider using the same CCA framework as described in Step 1, we first apply the binary image complement on B_1 , and then follow Step 1 to

generate a new binary image B_2 with an area less than t_{hole} , where t_{hole} is the area threshold for text stroke holes. The resulting binary image B is

$$B = B_1 \vee B_2 \quad (8)$$

4 Experimental results and discussion

We have conducted extensive experiments to evaluate the performance of our proposed framework. In this section, we first determined the size of disk-shaped morphological structuring elements, and then quantitatively compared our binarization method with other state-of-the-art segmentation techniques on recent DIBCO and H-DIBCO benchmark datasets.

4.1 Datasets

This study uses nine document image binarization competition datasets from 2009 to 2018, such as DIBCO 2009 [2], 2011 [3], 2013 [4], and 2017 [5] and H-DIBCO 2010 [6], 2012 [7], 2014 [8], 2016 [9], and 2018 [10] benchmark datasets, covering 31 machine-printed and 85 handwritten document images as well as their corresponding *ground truth* (GT) images. The historical documents in these datasets are originated from the *recognition and enrichment of archival documents* (READ) project, which contains a variety of collections from the 15th to 19th century. The datasets contain representative document degradation, such as ink bleed through, page smudges, text stroke fading, background texture, and artifacts.

4.2 Evaluation metrics

We adopt evaluation measures used in recent international document image binarization competitions, including FM (*F-measure*), pFM (*pseudo F-measure*), PSNR (*peak signal-to-noise ratio*), NRM (*negative rate metric*), DRD (*distance reciprocal distortion*), and MPM (*misclassification penalty metric*). The first two metrics, namely FM and pFM, reach their best values at 1 and the worst at 0. The PSNR measures how close a binary image to the GT image, so the higher the PSNR value, the better. In contrast to the former three metrics, the binarization performance is better for lower NRM, DRD, and MPM values. Due to space limitations, we omit definitions of those evaluation measures, but readers may refer to [6, 10] for more information.

4.3 Comparison results on the size of morphological structuring elements

This experiment demonstrates how to determine the size of the morphological structuring element, which is an essential part of morphological operations. In the document background estimation and compensation stage, we perform morphological closing or opening operation with disk-shaped structuring element to estimate document background, and then perform morphological bottom-hat or top-hat transform to emphasize text area and attenuate document background.

As mentioned in Subsection 3.2, the size of the disk-shaped structuring element should be no smaller than the width of text strokes. Therefore, we set different radius values to obtain different sizes of disk-shaped structuring elements, and then evaluate the binarization performance of our proposed method on the DIBCO 2009 and H-DIBCO 2010 benchmark datasets.

Figure 6 compares the FM of our proposed technique when the radius increases from 1 to 5 times the estimated text stroke width. As can be seen from the figure, the FM becomes stable when the radius value is 2 times larger than the estimated text stroke width on the two datasets. In our implementation, we therefore set the radius value to about 3.5 times the estimated text stroke width, as it gives the best score for the F-measure on both datasets.

4.4 Comparison results on the DIBCO and H-DIBCO benchmark datasets

In the first experiment, we quantitatively compared the proposed method with those that achieved TOP 3 performance in the annual document image binarization competition during 2009–2018. The evaluation results are provided in Table 1, and those for the TOP 3 winners of the year are copied from the corresponding competition reports [2–10], respectively. Readers may also refer to the same competition reports for a brief description of each winning method involved in this experiment. From the data in Table 1, we can see that our proposed method achieves the best performance in almost all the evaluation measures, except for DIBCO 2017, in which the TOP 3 winners are all based on deep learning architectures.

It is worth noting that our proposed method is based on graph cut, which is an efficient and powerful graph-based segmentation technique before the deep learning era. It consists of two main parts, namely the data part, which measures the consistency of image data within the segmented region that includes the features of the image, and the regularization part, which smooths the boundaries of the segmented region by maintaining the spatial information of the image. The graph cut is considered as an energy minimization process of the constructed graph when segmenting the image using a set of seeds (e.g., text stroke edges and document background), and no training is required. However, deep learning-based network models follow hierarchy architecture. Images or patches are fed into the network,

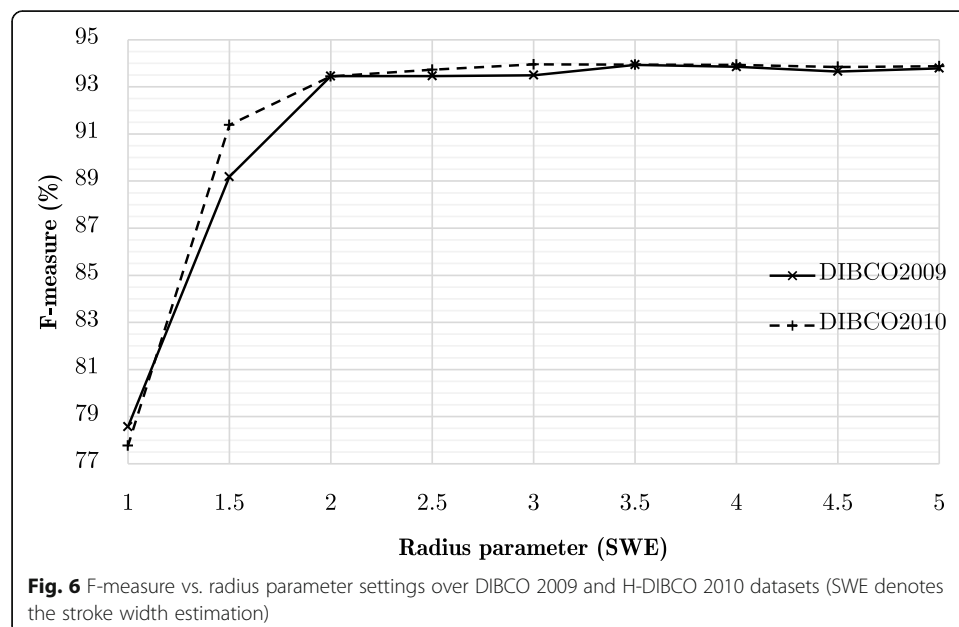


Table 1 Performance evaluation results of our proposed method against the TOP 3 winners of the annual DIBCO or H-DIBCO competitions (best results highlighted in bold)

Dataset	Method	FM (%)	pFM (%)	PSNR (dB)	NRM (%)	DRD	MPM (‰)
DIBCO 2009	Rank 1st [21]	91.24		18.66	4.31		0.55
	Rank 2nd [55]	90.06		18.23	4.75		0.89
	Rank 3rd [56]	89.34		17.79	5.32		1.90
	Proposed	93.46		20.01	2.59		1.54
H-DIBCO 2010	Joint 1st [23]	91.50	93.58	19.78	5.98		0.49
	Joint 1st [57]	89.70	95.15	19.15	8.18		0.29
	Rank 2nd [58]	91.78	94.43	19.67	4.77		1.33
	Rank 3rd [28]	89.73	90.11	18.90	5.78		0.41
	Proposed	93.73	95.18	20.97	3.64		0.29
DIBCO 2011	Rank 1st [59]	80.86		16.13		104.48	64.43
	Rank 2nd [24]	85.20		17.16		15.66	9.07
	Rank 3rd [28]	88.74		17.84		5.36	8.68
	Proposed	90.72		18.85		4.47	7.87
H-DIBCO 2012	Rank 1st [29]	89.47	90.18	21.80		3.44	
	Rank 2nd [59]	92.85	93.34	20.57		2.66	
	Rank 3rd [24]	91.54	93.30	20.14		3.05	
	Proposed	94.26	95.16	21.68		2.08	
DIBCO 2013	Rank 1st [24]	92.12	94.19	20.68		3.10	
	Rank 2nd [29]	92.70	93.19	21.29		3.18	
	Rank 3rd [60]	91.81	92.67	20.68		4.02	
	Proposed	93.51	94.54	21.32		2.77	
H-DIBCO 2014	Rank 1st [33]	96.88	97.65	22.66		0.90	
	Rank 2nd [29]	96.63	97.46	22.40		1.00	
	Rank 3rd [61]	93.35	96.05	19.45		2.19	
	Proposed	96.77	97.73	22.47		0.95	
H-DIBCO 2016	Rank 1st [34]	87.61	91.28	18.11		5.21	
	Rank 2nd [62, 63]	88.72	91.84	18.45		3.86	
	Rank 3rd [62]	88.47	91.71	18.29		3.93	
	Proposed	89.64	93.56	18.69		4.03	
DIBCO 2017	Rank 1st [45]	91.04	92.86	18.28		3.40	
	Rank 2nd	89.67	91.03	17.58		4.35	
	Rank 3rd [42]	89.42	91.52	17.61		3.56	
	Proposed	89.37	90.80	17.99		5.51	
H-DIBCO 2018	Rank 1st [64]	88.34	90.24	19.11		4.92	
	Rank 2nd	73.45	75.94	14.62		26.24	
	Rank 3rd	70.01	74.68	13.58		17.45	
	Proposed	88.34	90.37	19.11		4.93	

and then features are extracted by different layers. The shallow layer extracts fine-grained low-level visual features, which are minor details of the input, such as edges and blobs, while the deep layer extracts coarse-grained high-level semantic features, which are more abstract and built on top of low-level features to detect or recognize objects. Although the FM, pFM, PSNR, and DRD metrics of the

proposed method are slightly worse than or comparable to those of the TOP 3 winners in the DIBCO 2017 competition, it still illustrates that our proposed method can better segment text pixels and preserve text strokes.

In the second experiment, we have also quantitatively compared our proposed method with Otsu's global thresholding method [15], locally adaptive thresholding (e.g., Niblack's [16], Sauvola's [17], and Wolf's [18]) methods, contrast or edge-based segmentation (e.g., Lu's BESE [21], Su's LMM [23, 24], and Jia's SSP [25]) methods, energy-based segmentation (e.g., Howe's [28, 29], Mesquita's [33], and Kligler's [34]) methods, Bhowmik's game theory-inspired binarization [39], as well as deep learning-based segmentation (e.g., Tensmeyer's FCN [41], Vo's DSN [42], Gallego's SAE [43], and Zhao's cGAN [47]) methods for all the nine DIBCO and H-DIBCO testing datasets. The running codes for the methods involved in this comparison are provided by the original author(s), and the evaluation results are listed in Table 2. The first, second, and third best results for each evaluation measure are bolded in red, green, and blue, respectively. As can be seen from the table, our proposed method outperforms all other non-deep learning-based approaches, and is even comparable to several state-of-the-art deep learning-based techniques. This also implies that the proposed method is robust to various types and levels of document degradation, and can preserve text strokes better.

Table 2 Performance evaluation results of our proposed method against the state-of-the-art techniques on the nine DIBCO and H-DIBCO testing datasets (the first, second, and third best results in bold red, green, and blue font, respectively)

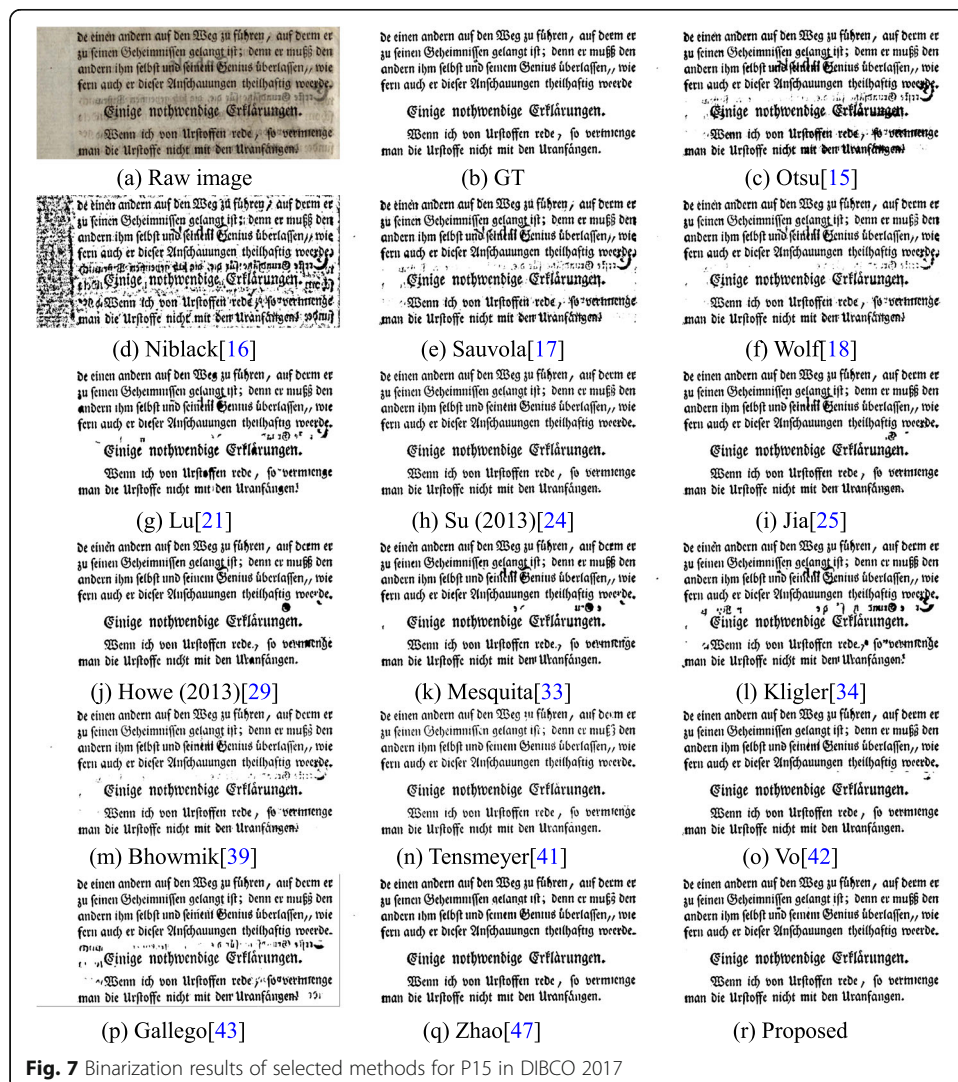
CATEGORY	METHOD	YEAR	FM(%)	pFM(%)	PSNR(dB)	NRM(%)	DRD	MPM(‰)
Global thresholding	Otsu[15]	1979	78.77	81.96	15.48	8.57	16.72	22.22
Local thresholding	Niblack[16]	1986	42.47	42.98	6.79	14.14	88.99	162.31
	Sauvola[17]	2000	82.37	86.61	16.69	10.16	6.35	4.52
	Wolf[18]	2003	81.40	86.01	16.82	11.83	5.98	3.24
Contrast or edge-based	Lu[21] [†]	2010	85.76	87.31	17.63	7.01	7.86	11.08
	Su[23]	2010	87.94	89.77	18.24	6.58	5.10	5.76
	Su[24]	2013	88.90	91.57	18.67	6.50	6.06	4.53
	Jia[25]	2018	90.08	92.58	19.04	4.77	3.87	3.09
Energy-based	Howe(Conf)[28]	2011	87.36	89.05	17.75	5.89	5.37	6.33
	Howe(Base)[29]		89.71	91.17	19.16	5.18	4.79	5.87
	Howe(Alg1)[29]	2013	89.22	91.26	19.15	5.40	4.83	5.79
	Howe(Alg2)[29]		90.25	91.72	19.97	5.06	4.08	4.30
	Howe(Alg3)[29]		90.97	92.40	19.88	4.43	4.11	5.53
	Mesquita[33]	2015	91.46	92.80	19.86	3.73	3.97	5.28
Kligler[34]	2018	89.16	90.71	19.28	5.18	4.97	5.60	
Unsupervised learning-based	Bhowmik[39]	2019	85.14	90.27	17.16	9.17	7.46	5.42
Deep learning-based	Tensmeyer[41]	2017	91.45	95.28	20.07	5.96	2.78	1.30
	Vo[42]	2018	91.89	95.01	19.77	4.59	2.91	2.60
	Gallego[43]	2019	84.68	86.82	16.94	7.23	6.31	7.70
	Zhao[47]	2019	91.73	93.35	19.92	3.89	3.46	3.86
	Proposed			92.02	93.53	19.99	3.84	3.40

[†]Note: H01.tif in H-DIBCO 2010, H03.png in H-DIBCO 2012, H08.bmp in HDIBCO 2016, H01-H05.bmp in DIBCO 2017, and H10.bmp in H-DIBCO 2018 are excluded, because the original Lu's BESE method encounters errors when processing these document images

Figures 7 and 8 display two sample images (P15 in DIBCO 2017 and H06 in H-DIBCO 2018 datasets) and the resulting binary images generated by selected comparison methods. As you can see from the figures, Otsu’s [15] global thresholding and Niblack’s [16] local thresholding methods generally fail to produce reasonable results. Sauvola’s [17] and Wolf’s [18] locally adaptive thresholding methods tend to remove too many text strokes. Lu’s BESE [21], Su’s LMM [23, 24], as well as Howe’s [28, 29], Mesquita’s [33], and Kligler’s [34] energy-based methods fail to extract low-contrast text strokes. Compared with Jia’s SSP [25], Bhowmik’s GiB [39], and other state-of-the-art CNN-based segmentation methods (e.g., Tensmeyer’s FCN [41], Vo’s DSN [42], Gallego’s SAE [43], and Zhao’s cGAN [47]), our proposed method can better preserve text strokes and produce better visual quality.

4.5 Comparison results on the time complexity of each binarization method

Since the proposed method mainly consists of several stages, namely preprocessing (including image reading, standard color-to-gray conversion, and image normalization),



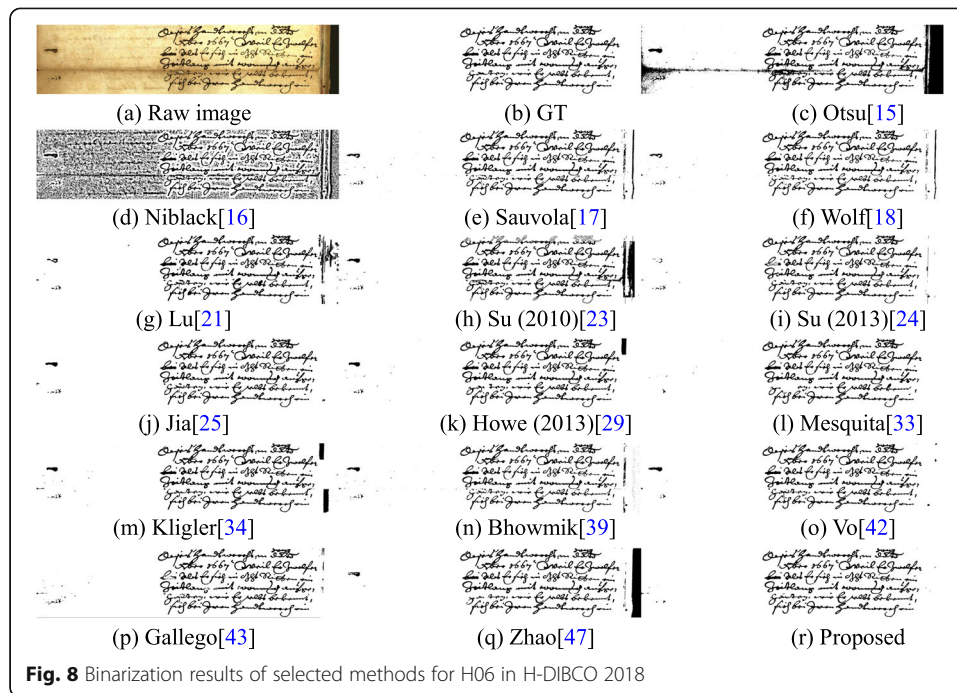


Fig. 8 Binarization results of selected methods for H06 in H-DIBCO 2018

stroke width transform, background estimation and compensation, graph cut-based segmentation, and post-processing, among which stroke width transform and graph cut-based segmentation are the two most time-consuming stages, we analyze the computational complexity of these two stages theoretically.

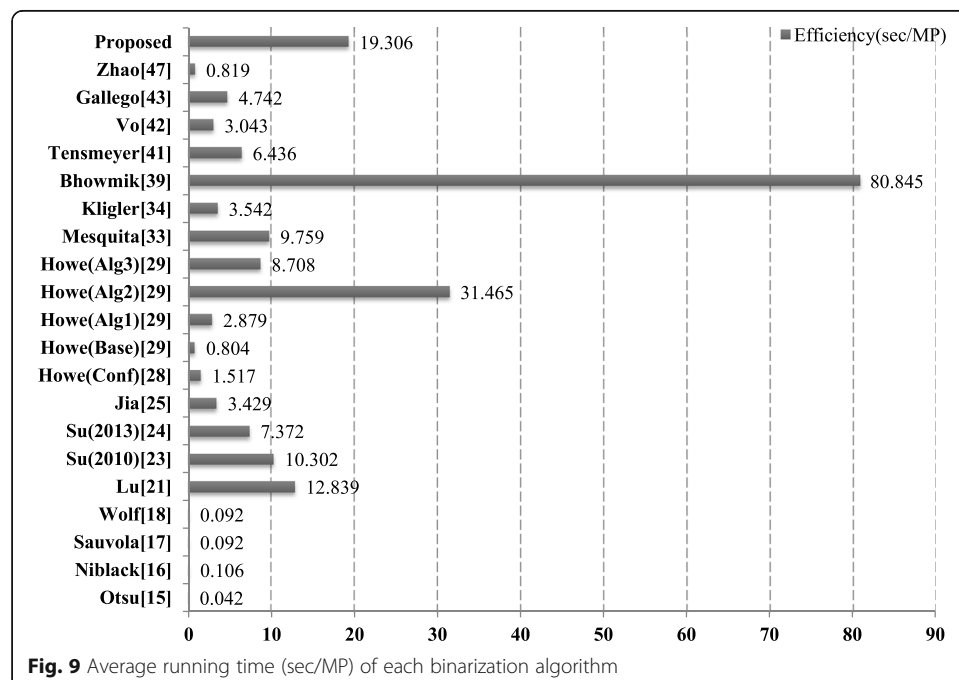
To estimate text stroke width accurately, we apply the stroke width transform operator presented by Epshtein et al. [48]. It is a text region detection algorithm that extracts text strokes from noisy images, which are finite width shapes and consist of two roughly parallel edges. Starting from a text edge pixel and by exploring pixels perpendicular to the direction of text edges, we may locate another text edge pixel, whose gradient direction is approximately opposite to that of the previous one, and then a stroke cross-section is formed from these two edge pairs. By joining the stroke cross-sections of similar widths, a complete text stroke is thus produced. Therefore, the theoretical worst case complexity of the stroke width transform algorithm is $O(n|P|)$, where n is the number of pixels in the image, and $|P|$ is the length of stroke cross-sections.

For graph cut-based segmentation, we adopt the max-flow and min-cut algorithm presented by Boykov and Kolmogorov [30] for energy minimization. It belongs to the group of algorithms based on augmenting paths. To detect augmenting paths, two non-overlapping search trees rooted at the source and the sink are built; and to detect a new augmenting path (not necessarily the shortest one), these trees are no longer rebuilt from scratch. As a result, the theoretical worst case complexity of the Boykov-Kolmogorov algorithm is $O(mn^2|C|)$, where $|C|$ is the cost of the minimum cut, n and m are the number of nodes and edges in the graph, respectively.

To give the reader a clearer picture about the execution efficiency of each method, we adopt the average running time in *second per megapixel* (sec/MP) to evaluate the time complexity of each binarization algorithm. All experiments are conducted on my Dell Alienware 17 R5 laptop. The system hardware configurations are Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz with 16 GB RAM and NVIDIA GeForce GTX 1080 with 8 GB GDDR5X Video RAM.

In terms of the programming language used by each method, Otsu's [15] global thresholding and Niblack's [16], Sauvola's [17], and Wolf's [18] locally adaptive thresholding methods are reproduced in Matlab scripts. Lu's BESE [21] is in Matlab pcode format, while Su's LMM [23, 24] and Jia's SSP [25] methods are written in C++ with OpenCV. Howe's [28, 29], Mesquita's [33], and Kligler's [34] energy-based segmentation (including our proposed method) are implemented in Matlab scripts. Bhowmik's GiB [39] also uses Matlab, but is compiled into an executable. The deep learning methods are all Python-based. However, the deep learning framework used by Tensmeyer's FCN [41] and Vo's DSN [42] is Caffe. Gallego's SAE [43] adopts TensorFlow and Zhao's cGAN [47] uses PyTorch. Deep learning-based methods run on the GPU, while non-deep learning-based counterparts run on the CPU. Therefore, we can only roughly evaluate the average running time of each binarization algorithm, as shown in Fig. 9.

It can be seen from the figure that binarization methods based on simple statistical features, such as Otsu's, Niblack's, Sauvola's, and Wolf's, are relatively less computationally intensive and faster to process, but the binarization performance is poor. The processing speed of our proposed method is comparable to that of most other contrast/edge-based or energy-based segmentation algorithms, and is significantly faster than that of Bhowmik's game theory-inspired binarization.



4.6 Discussion

The superior performance of the proposed method can be explained by the following factors:

First, our proposed method estimates the text stroke width feature based on stroke width transform with minimum entropy. It can detect either bright text on dark background or vice versa, and can distinguish true text in various sizes, fonts, and orientations from possible degradation.

Second, mathematical morphology is used to compensate the document background and then the Laplacian energy-based segmentation is performed on the compensated document images. The estimation and compensation procedure can remove most of the document degradation and help to extract text objects from complex document background in the subsequent energy-based segmentation stage.

Last but not least, the proposed method employs post-processing operations to eliminate possible noise and preserve text stroke connectivity by removing isolated text pixels and filling breaks, gaps, or holes inside text strokes.

Of course, we also need to emphasize that there is no single binarization algorithm that works for all types of historical document images. The method proposed in this paper is no exception, and it also has some limitations:

(1) Manual extraction of text stroke features using traditional feature engineering is somewhat inadequate and subject to bias, especially when dealing with extremely degraded or badly damaged pages of historical antiquities. Therefore, deep learning-based approaches are not only a good alternative, but also the current trend.

(2) The memory usage of graph cuts will increase rapidly as the image size increases; for instance, the well-known Boykov-Kolmogorov's max-flow and min-cut algorithm [30] that we used for energy minimization allocates $24n + 14m$ bytes, where n and m are the number of nodes and edges in the graph, respectively.

5 Conclusion

We propose an enhanced historical document image binarization method based on background estimation and energy minimization. It first adopts mathematical morphology algorithms to compensate document background. The size of the disk-shaped structuring element is determined by the stroke width transform with minimum entropy. We then perform Laplacian energy-based segmentation on the compensated document images. Finally, we implement post-processing to preserve text stroke connectivity and eliminate isolated noise. The proposed method is robust to various types and levels of document degradation and leads to high accuracy. Experimental results show that the overall performance of our proposed method is far superior to other state-of-the-art segmentation techniques.

For future work, we intend to conduct further research in the following aspects. First, we can improve the contrast between text and background by using machine learning or deep learning techniques to effectively achieve degraded document image enhancement in the preprocessing stage, and then take fully connected CRFs [65] or convolutional CRFs [66] for further segmentation. Second, to address the problems such as poor robustness of handcrafted or manually designed text features, we hope to perform

multi-scale and adaptive feature extraction and learning through deep network models, so as to improve the discriminative property of text regions.

Abbreviations

DAR: Document analysis and recognition; OCR: Optical character recognition; DIBCO: Document image binarization competition; H-DIBCO: Handwritten document image binarization competition; SWT: Stroke width transform; CCA: Connected component analysis; BESE: Background estimation and stroke edges; LMM: Local maximum and minimum; SSPs: Structural symmetric pixels; MRFs: Markov random fields; CRFs: Conditional random fields; SVM: Support vector machine; GiB: Game theory inspired binarization; CNNs: Convolutional neural networks; FCNs: Fully convolutional networks; DSNs: Deep supervised networks; SAE: Selectional auto-encoder; cGANs: Conditional generative adversarial networks; READ: Recognition and enrichment of archival documents; FM: F-measure; pFM: Pseudo F-measure; PSNR: Peak signal-to-noise ratio; NRM: Negative rate metric; DRD: Distance reciprocal distortion; MPM: Misclassification penalty metric; GT: Ground truth; sec/MP: Second per megapixel

Authors' note

A preliminary version of this paper was presented in the *24th International Conference on Pattern Recognition (ICPR 2018)*, Beijing, China, August 20–24, 2018 [64]. This version includes a concrete analysis of the voting strategy, namely the minimum entropy-based *stroke width transform* (SWT), as well as more binarization performance comparison experiments on the DIBCO and H-DIBCO benchmark datasets.

Authors' contributions

Wei Xiong devised the project and provided the main conceptual ideas. He also made revisions to the manuscript. *Lei Zhou* developed the theory and conducted the suggested experiments. She wrote the first draft of this paper. *Ling Yue* helped design the algorithm and computational framework. She also validated the analytical methods. *Lirong Li* directed the experiments. *Song Wang* revised and polished the manuscript. He also gave several positive suggestions. The authors discussed the results and contributed to the final manuscript. The authors read and approved the final manuscript.

Authors' information

Wei Xiong received the B.S. degree in electronic engineering and the Ph.D. degree in signal and information processing, both from Wuhan University, Hubei, China, in 2003 and 2010, respectively. He is an Associate Professor with the School of Electrical and Electronic Engineering, Hubei University of Technology, Hubei, China. He was a visiting professor with the Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA. His research interests include computer vision, pattern recognition, deep learning, and artificial intelligence. *Lei Zhou* received the B.S. degree in communication engineering from Liren College of Yanshan University, Hebei, China, in 2019. She is a M.S. candidate in control engineering with the School of Electrical and Electronic Engineering, Hubei University of Technology, Hubei, China. Her research interests include computer vision, deep learning, and semantic segmentation.

Ling Yue received the B.S. degree in electronic information engineering from Hubei University of Technology, Hubei, China, in 2019. She is a M.S. candidate in control engineering with the School of Electrical and Electronic Engineering, Hubei University of Technology, Hubei, China. Her research interests include computer vision, deep learning, and person re-identification.

Lirong Li received the M.S. degree and Ph.D. degree, both in pattern recognition and artificial intelligence, from Huazhong University of Science and Technology, Wuhan, Hubei, China, in 2004 and 2017, respectively. She has been working as a teacher in the School of Electrical and Electronic Engineering, Hubei University of Technology since July 2004. Her research interests include computer vision and artificial intelligence, multispectral image processing.

Song Wang (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC) in 2002. From 1998 to 2002, he also worked as a Research Assistant with Image Formation and Processing Group, Beckman Institute, UIUC. In 2002, he joined the Department of Computer Science and Engineering, University of South Carolina, where he is currently a Professor. His research interests include computer vision, medical image processing, and machine learning. He is a Senior Member of the IEEE Computer Society. He is currently serving as an Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Pattern Recognition Letters, and Electronics Letters.

Funding

This research was supported by National Natural Science Foundation of China (61571182, 61601177), Natural Science Foundation of Hubei Province, China (2019CFB530), China Scholarship Council (201808420418), and Hubei Provincial Department of Education (B2019042).

Availability of data and materials

The DIBCO and H-DIBCO benchmark datasets, along with the corresponding ground truth images as well as objective evaluation methodologies, are available for public download at each competition site. Our Laplacian energy-based segmentation method, which achieved the best performance in H-DIBCO 2018, is available at <https://github.com/beargolden/H-DIBCO-2018>.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Received: 1 October 2020 Accepted: 13 April 2021

Published online: 10 May 2021

References:

1. S. Eskenazi, P. Gomez-Krämer, J.-M. Ogier, A comprehensive survey of mostly textual document segmentation algorithms since 2008. *Pattern Recognit.* **64**, 1–14 (2017). <https://doi.org/10.1016/j.patcog.2016.10.023>
2. B. Gatos, K. Ntirogiannis, I. Pratikakis, "ICDAR 2009 document image binarization contest (DIBCO 2009)," in Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR 2009), Barcelona, SPAIN, 2009, pp. 1375-1382. doi: <https://doi.org/10.1109/icdar.2009.246>
3. I. Pratikakis, B. Gatos, K. Ntirogiannis, "ICDAR 2011 document image binarization contest (DIBCO 2011)," in Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR 2011), Beijing, CHINA, 2011, pp. 1506-1510. doi: <https://doi.org/10.1109/icdar.2011.299>
4. I. Pratikakis, B. Gatos, K. Ntirogiannis, "ICDAR 2013 document image binarization contest (DIBCO 2013)," in Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013), Washington, DC, USA, 2013, pp. 1471-1476. doi: <https://doi.org/10.1109/icdar.2013.219>
5. I. Pratikakis, K. Zagoris, G. Barlas, B. Gatos, "ICDAR 2017 competition on document image binarization (DIBCO 2017)," in Proceedings of the 14th International Conference on Document Analysis and Recognition (ICDAR 2017), Kyoto, JAPAN, 2017, pp. 1395-1403. doi: <https://doi.org/10.1109/icdar.2017.228>
6. I. Pratikakis, B. Gatos, K. Ntirogiannis, "H-DIBCO 2010 - handwritten document image binarization competition," in Proceedings of the 12th International Conference on Frontiers in Handwriting Recognition (ICFHR 2010), Kolkata, INDIA, 2010, pp. 727-732. doi: <https://doi.org/10.1109/icfhr.2010.118>
7. I. Pratikakis, B. Gatos, K. Ntirogiannis, "ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012)," in Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition (ICFHR 2012), Monopoli, ITALY, 2012, pp. 817-822. doi: <https://doi.org/10.1109/icfhr.2012.216>
8. K. Ntirogiannis, B. Gatos, I. Pratikakis, "ICFHR 2014 competition on handwritten document image binarization (H-DIBCO 2014)," in Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition (ICFHR 2014), Hersonissos, GREECE, 2014, pp. 809-813. doi: <https://doi.org/10.1109/icfhr.2014.141>
9. I. Pratikakis, K. Zagoris, G. Barlas, B. Gatos, "ICFHR 2016 handwritten document image binarization contest (H-DIBCO 2016)," in Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition (ICFHR 2016), Shenzhen, CHINA, 2016, pp. 619-623. doi: <https://doi.org/10.1109/icfhr.2016.110>
10. I. Pratikakis, K. Zagoris, P. Kaddas, B. Gatos, "ICFHR 2018 competition on handwritten document image binarization (H-DIBCO 2018)," in Proceedings of the 16th International Conference on Frontiers in Handwriting Recognition (ICFHR 2018), Niagara Falls, USA, 2018, pp. 489-493. doi: <https://doi.org/10.1109/icfhr-2018.2018.00091>
11. M. W. A. Kesiman, D. Vally, J.-C. Burie, E. Paulus, M. Suryani, S. Hadi, M. Verleysen, S. Chhun, J.-M. Ogier, "ICFHR 2018 competition on document image analysis tasks for Southeast Asian palm leaf manuscripts," in Proceedings of the 16th International Conference on Frontiers in Handwriting Recognition (ICFHR 2018), Niagara Falls, USA, 2018, pp. 483-488. doi: <https://doi.org/10.1109/icfhr-2018.2018.00090>
12. L. Zhou, C. Zhang, M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in Proceedings of the 31st Meeting of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR 2018), Salt Lake City, UT, USA, 2018, pp. 192-196. doi: <https://doi.org/10.1109/cvprw.2018.00034>
13. R. D. Lins, E. Kavallieratou, E. B. Smith, R. B. Bernardino, D. M. d. Jesus, "ICDAR 2019 time-quality binarization competition," in Proceedings of the 15th International Conference on Document Analysis and Recognition (ICDAR 2019), Sydney, AUSTRALIA, 2019
14. M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **13**(1), 146–168 (2004). <https://doi.org/10.1117/1.1631316>
15. N. Otsu, A threshold selection method from gray-level histograms. *IEEE Trans.Syst. Man Cybern.* **9**(1), 62–66 (1979). <https://doi.org/10.1109/tsmc.1979.4310076>
16. W. Niblack, *An introduction to digital image processing* (Prentice-Hall International Inc., Englewood Cliffs, New Jersey, 1986)
17. J. Sauvola, M. Pietikäinen, Adaptive document image binarization. *Pattern Recognit.* **33**(2), 225–236 (2000). [https://doi.org/10.1016/s0031-3203\(99\)00055-2](https://doi.org/10.1016/s0031-3203(99)00055-2)
18. C. Wolf, J.-M. Jolion, Extraction and recognition of artificial text in multimedia documents. *Pattern Anal. Appl.* **6**(4), 309–326 (2003). <https://doi.org/10.1007/s10044-003-0197-7>
19. K. Ntirogiannis, B. Gatos, I. Pratikakis, A combined approach for the binarization of handwritten document images. *Pattern Recognit. Lett.* **35**, 3–15 (2014). <https://doi.org/10.1016/j.patrec.2012.09.026>
20. B. Su, S. Lu, C. L. Tan, "Combination of document image binarization techniques," in Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR 2011), Beijing, CHINA, 2011, pp. 22-26. doi: <https://doi.org/10.1109/icdar.2011.14>
21. S. Lu, B. Su, C.L. Tan, Document image binarization using background estimation and stroke edges. *Int. J. Doc. Anal. Recognit.* **13**(4), 303–314 (2010). <https://doi.org/10.1007/s10032-010-0130-8>
22. S. Lu, C. L. Tan, "Binarization of badly illuminated document images through shading estimation and compensation," in Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007), Curitiba, BRAZIL, 2007, pp. 312-316. doi: <https://doi.org/10.1109/icdar.2007.4378723>

23. 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 (DAS 2010), Boston, Massachusetts, USA, 2010, pp. 159-165. doi: <https://doi.org/10.1145/1815330.1815351>
24. B. Su, S. Lu, C.L. Tan, Robust document image binarization technique for degraded document images. *IEEE Trans. Image Process.* **22**(4), 1408–1417 (2013). <https://doi.org/10.1109/tip.2012.2231089>
25. F. Jia, C. Shi, K. He, C. Wang, B. Xiao, Degraded document image binarization using structural symmetry of strokes. *Pattern Recognit.* **74**, 225–240 (2018). <https://doi.org/10.1016/j.patcog.2017.09.032>
26. Q.N. Vo, S.H. Kim, H.J. Yang, G. Lee, An MRF model for binarization of music scores with complex background. *Pattern Recognit. Lett.* **69**, 88–95 (2016). <https://doi.org/10.1016/j.patrec.2015.10.017>
27. E. Ahmadi, Z. Azimifar, M. Shams, M. Famouri, M.J. Shafiee, Document image binarization using a discriminative structural classifier. *Pattern Recognit. Lett.* **63**, 36–42 (2015). <https://doi.org/10.1016/j.patrec.2015.06.008>
28. N. R. Howe, "A Laplacian energy for document binarization," in Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR 2011), Beijing, CHINA, 2011, pp. 6-10. doi: <https://doi.org/10.1109/icdar.2011.11>
29. N.R. Howe, Document binarization with automatic parameter tuning. *Int. J. Doc. Anal. Recognit.* **16**(3), 247–258 (2013). <https://doi.org/10.1007/s10032-012-0192-x>
30. Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Machine Intell.* **26**(9), 1124–1137 (2004). <https://doi.org/10.1109/tpami.2004.60>
31. F. Westphal, H. Grahn, N. Lavesson, Efficient document image binarization using heterogeneous computing and parameter tuning. *Int. J. Doc. Anal. Recognit.* **21**(1-2), 41–58 (2018). <https://doi.org/10.1007/s10032-017-0293-7>
32. R.G. Mesquita, C.A.B. Mello, L.H.E.V. Almeida, A new thresholding algorithm for document images based on the perception of objects by distance. *Integr. Comput. Aided Eng.* **21**(2), 133–146 (2014). <https://doi.org/10.10233/ica-130453>
33. R.G. Mesquita, R.M.A. Silva, C.A.B. Mello, P.B.C. Miranda, Parameter tuning for document image binarization using a racing algorithm. *Expert Syst. Appl.* **42**(5), 2593–2603 (2015). <https://doi.org/10.1016/j.eswa.2014.10.039>
34. N. Kligler, S. Katz, A. Tal, "Document enhancement using visibility detection," in Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, Utah, USA, 2018
35. D. Rivest-Hénault, R.F. Moghaddam, M. Cheriet, A local linear level set method for the binarization of degraded historical document images. *Int. J. Doc. Anal. Recognit.* **15**(2), 101–124 (2012). <https://doi.org/10.1007/s10032-011-0157-5>
36. Z. Hadjadj, M. Cheriet, A. Meziane, Y. Cherfa, A new efficient binarization method: Application to degraded historical document images. *Signal Image Video Process.*, 1–8 (2017). <https://doi.org/10.1007/s11760-017-1070-2>
37. X. Chen, L. Lin, Y. Gao, Parallel nonparametric binarization for degraded document images. *Neurocomputing* **189**, 43–52 (2016). <https://doi.org/10.1016/j.neucom.2015.11.040>
38. W. Xiong, J. Xu, Z. Xiong, J. Wang, M. Liu, Degraded historical document image binarization using local features and support vector machine (svm). *Optik* **164**, 218–223 (2018). <https://doi.org/10.1016/j.ijleo.2018.02.072>
39. S. Bhowmik, R. Sarkar, B. Das, D. Doermann, Gib: A game theory inspired binarization technique for degraded document images. *IEEE Trans. Image Process.* **28**(3), 1443–1455 (2019). <https://doi.org/10.1109/tip.2018.2878959>
40. J. Pastor-Pellicer, S. España-Boquera, F. Zamora-Martínez, M. Z. Afzal, M. J. Castro-Bleda, "Insights on the use of convolutional neural networks for document image binarization," in Proceedings of the 13th International Workshop on Artificial Neural Networks (IWANN 2015), Palma de Mallorca, SPAIN, 2015, pp. 115-126. doi: https://doi.org/10.1007/978-3-319-19222-2_10
41. C. Tensmeyer, T. Martinez, "Document image binarization with fully convolutional neural networks," in Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR 2017), Kyoto, Japan, 2017, pp. 99-104. doi: <https://doi.org/10.1109/icdar.2017.25>
42. Q.N. Vo, S.H. Kim, H.J. Yang, G. Lee, Binarization of degraded document images based on hierarchical deep supervised network. *Pattern Recognit.* **74**, 568–586 (2018). <https://doi.org/10.1016/j.patcog.2017.08.025>
43. J. Calvo-Zaragoza, A.-J. Gallego, A selectional auto-encoder approach for document image binarization. *Pattern Recognit.* **86**, 37–47 (2019). <https://doi.org/10.1016/j.patcog.2018.08.011>
44. X.-J. Mao, C. Shen, Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in Proceedings of the 30th Annual Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 2016, pp. 2810-2818
45. P.V. Bezmaterniykh, D.A. Ilin, D.P. Nikolaev, U-net-bin: Hacking the document image binarization contest. *Comput. Opt.* **43**(5), 825–832 (2019). <https://doi.org/10.18287/2412-6179-2019-43-5-825-832>
46. O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, GERMANY, 2015, pp. 234-241. doi: https://doi.org/10.1007/978-3-319-24574-4_28
47. J. Zhao, C. Shi, F. Jia, Y. Wang, B. Xiao, Document image binarization with cascaded generators of conditional generative adversarial networks. *Pattern Recognit.* **96** (2019). <https://doi.org/10.1016/j.patcog.2019.106968>
48. B. Epshtein, E. Ofek, Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, CA, 2010, pp. 2963-2970. doi: <https://doi.org/10.1109/cvpr.2010.5540041>
49. L. He, X. Ren, Q. Gao, X. Zhao, B. Yao, Y. Chao, The connected-component labeling problem: A review of state-of-the-art algorithms. *Pattern Recognit.* **70**(Supplement C), 25–43 (2017). <https://doi.org/10.1016/j.patcog.2017.04.018>
50. Y.H. Yang, M.J. Buckley, S. Dudoit, T.P. Speed, Comparison of methods for image analysis on cDNA microarray data. *J. Comput. Graphical Stat.* **11**(1), 108–136 (2002). <https://doi.org/10.1198/106186002317375640>
51. C. A. B. Mello, "Segmentation of images of stained papers based on distance perception," in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC 2010), Istanbul, TURKEY, 2010, pp. 1636-1642. doi: <https://doi.org/10.1109/icsmc.2010.5642394>
52. H. Orii, H. Kawano, H. Maeda, N. Ikoma, Text-color-independent binarization for degraded document image based on MAP-MRF approach. *IEICE Trans. Fundamentals Electron. Commun. Comput. Sci.* **94**(11), 2342–2349 (2011). <https://doi.org/10.1587/transfun.E94.A.2342>
53. B. Su, S. Lu, C. L. Tan, "A learning framework for degraded document image binarization using markov random field," in Proceedings of the 21st International Conference on Pattern Recognition (ICPR 2012), Tsukuba, JAPAN, 2012, pp. 3200-3203

54. S. Torbert, *Applied computer science, 2nd ed* (Springer International Publishing, Switzerland, 2016)
55. J. Fabrizio, B. Marcotegui, M. Cord, "Text segmentation in natural scenes using toggle-mapping," in Proceedings of the 16th IEEE International Conference on Image Processing (ICIP 2009), Cairo, Egypt, 2009, pp. 2373-2376. doi: <https://doi.org/10.1109/icip.2009.5413435>
56. R. F. Moghaddam, D. Rivest-Hénault, M. Cheriet, "Restoration and segmentation of highly degraded characters using a shape-independent level set approach and multi-level classifiers," in Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR 2009), Barcelona, Spain, 2009, pp. 828-832. doi: <https://doi.org/10.1109/icdar.2009.107>
57. I. Bar-Yosef, I. Beckman, K. Kedem, I. Dinstein, "Binarization, character extraction, and writer identification of historical Hebrew calligraphy documents. *Int. J. Doc. Anal. Recognit.* **9**(2-4), 89-99 (2007). <https://doi.org/10.1007/s10032-007-0041-5>
58. T. Lelore, F. Bouchara, Fair: A fast algorithm for document image restoration. *IEEE Trans. Pattern Anal. Machine Intell.* **35**(8), 2039-2048 (2013). <https://doi.org/10.1109/tpami.2013.63>
59. T. Lelore, F. Bouchara, "Super-resolved binarization of text based on the fair algorithm," in Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR 2011), Beijing, China, 2011, pp. 839-843. doi: <https://doi.org/10.1109/icdar.2011.172>
60. R. F. Moghaddam, F. F. Moghaddam, M. Cheriet, "Unsupervised ensemble of experts (EoE) framework for automatic binarization of document images," in Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013), Washington, D.C., 2013, pp. 703-707. doi: <https://doi.org/10.1109/icdar.2013.144>
61. H.Z. Nafchi, R.F. Moghaddam, M. Cheriet, Phase-based binarization of ancient document images: Model and applications. *IEEE Trans. Image Process.* **23**(7), 2916-2930 (2014). <https://doi.org/10.1109/tip.2014.2322451>
62. A. Hassaïne, S. Al-Maadeed, A. Bouridane, "A set of geometrical features for writer identification," in Proceedings of the 19th International Conference on Neural Information Processing (ICONIP 2012), Doha, QATAR, 2012, pp. 584-591. doi: https://doi.org/10.1007/978-3-642-34500-5_69
63. A. Hassaïne, E. Decencière, B. Besserer, Efficient restoration of variable area soundtracks. *Image Anal. Stereol.* **28**(2), 113-119 (2011). <https://doi.org/10.5566/ias.v28.p113-119>
64. W. Xiong, X. Jia, J. Xu, Z. Xiong, M. Liu, J. Wang, "Historical document image binarization using background estimation and energy minimization," in Proceedings of the 24th International Conference on Pattern Recognition (ICPR 2018), Beijing, CHINA, 2018, pp. 3716-3721. doi: <https://doi.org/10.1109/icpr.2018.8546099>
65. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Machine Intell.* **40**(4), 834-848 (2018). <https://doi.org/10.1109/tpami.2017.2699184>
66. X. Peng, C. Wang, H. Cao, "Document binarization via multi-resolutional attention model with DRD loss," in Proceedings of the 15th IAPR International Conference on Document Analysis and Recognition (ICDAR 2019), Sydney, NSW, Australia, 2019, pp. 45-50. doi: <https://doi.org/10.1109/icdar.2019.00017>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
