A Robust Document Image Binarization Technique for Degraded Document Images

Bolan Su, Shijian Lu *Member, IEEE,* Chew Lim Tan *Senior Member, IEEE,*

***Abstract*—Segmentation of text from badly degraded document images is a very challenging task due to the high inter/intra-variation between the document background and the foreground text of different document images. In this paper, we propose a novel document image binarization technique that addresses these issues by using adaptive image contrast. The adaptive image contrast is a combination of the local image contrast and the local image gradient that is tolerant to text and background variation caused by different types of document degradations. In the proposed technique, an adaptive contrast map is first constructed for an input degraded document image. The contrast map is then binarized and combined with Canny's edge map to identify the text stroke edge pixels. The document text is further segmented by a local threshold that is estimated based on the intensities of detected text stroke edge pixels within a local window. The proposed method is simple, robust and involves minimum parameter tuning. It has been tested on three public datasets that were used in the recent Document Image Binarization Contest (DIBCO) 2009 & 2011 and Handwritten Document Image Binarization Contest (H-DIBCO) 2010 and achieves accuracies of 93.5%, 87.8% and 92.03%, respectively, that are significantly higher than or close to that of the best-performing methods reported in the three contests. Experiments on the Bickley diary dataset that consists of several challenging bad quality document images also show the superior performance of our proposed method, compared with other techniques.**

***Index Terms*—Document Image Processing, Document Analysis, Pixel Classification, Degraded Document Image Binarization, Adaptive Image Contrast**

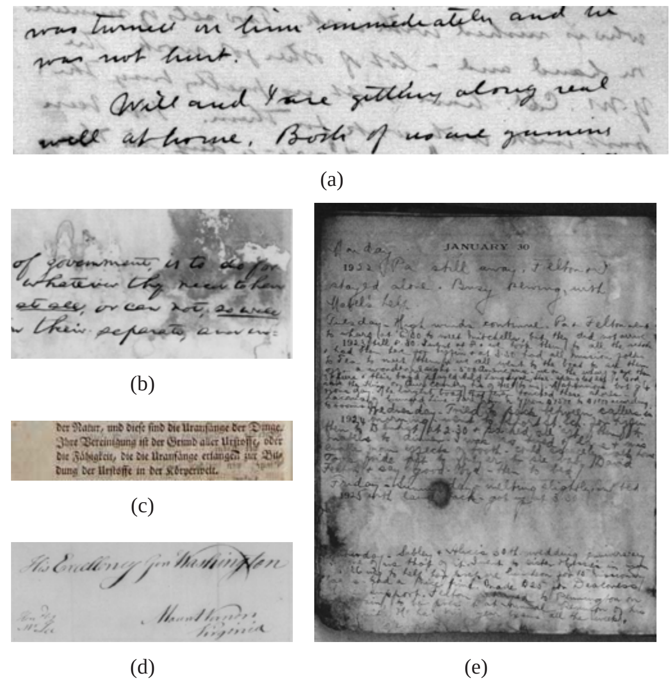


Fig. 1. Five degraded document image examples taken from DIBCO, H-DIBCO and Bickley diary datasets.

I. Introduction

Document Image Binarization is performed in the preprocessing stage for document analysis and it aims to segment the foreground text from the document background. A fast and accurate document image binarization technique is important for the ensuing document image processing tasks such as optical character recognition (OCR).

Though document image binarization has been studied for many years, the thresholding of degraded document images is still an unsolved problem due to the high inter/intra-variation between the text stroke and the document background across different document images. As illustrated in Figure 1, the handwritten text within the degraded documents often shows a certain amount of variation in terms of the stroke width, stroke brightness, stroke connection, and document background. In addition, historical documents are often degraded by the bleed-through as illustrated in Figure 1(a) and (c) where the ink of the other side seeps through to the front. In addition, historical documents are often degraded by different types of imaging artifacts as illustrated in Figure 1(e). These different types of document degradations tend to induce the document thresholding error and make degraded document image binarization a big challenge to most state-of-the-art techniques.

The recent Document Image Binarization Contest (DIBCO) [1], [2] held under the framework of the International Conference on Document Analysis and Recognition (ICDAR) 2009 & 2011 and the Handwritten Document Image Binarization Contest(H-DIBCO) [3] held under the framework of the International Conference on Frontiers in Handwritten Recognition show recent efforts on this issue. We participated in the DIBCO 2009 and our background estimation method [4] performs the best among entries of 43 algorithms submitted from 35 international research groups. We also participated in the H-DIBCO 2010 and our local maximum-minimum method [5] was one of the top two winners among 17 submitted algorithms. In the latest DIBCO 2011, our proposed method achieved second

Bolan Su (subolan@comp.nus.edu.sg) and Chew Lim Tan (tancl@comp.nus.edu.sg) are with School of Computing, National University of Singapore, Singapore. Shijian Lu (slu@i2r.a-star.edu.sg) is with Institute for Infocomm Research, A*STAR, Singapore.

best results among 18 submitted algorithms.

This paper presents a document binarization technique that extends our previous local maximum-minimum method [5] and the method used in the latest DIBCO 2011. The proposed method is simple, robust and capable of handling different types of degraded document images with minimum parameter tuning. It makes use of the adaptive image contrast that combines the local image contrast and the local image gradient adaptively and therefore is tolerant to the text and background variation caused by different types of document degradations. In particular, the proposed technique addresses the over-normalization problem of the local maximum minimum algorithm [5]. At the same time, the parameters used in the algorithm can be adaptively estimated.

The rest of this paper is organized as follows. Section 2 first reviews the current state-of-the-art binarization techniques. Our proposed document binarization technique is described in Section 3. Then experimental results are reported in Section 4 to demonstrate the superior performance of our framework. Finally, conclusions are presented in Section 5.

(a) Local Image Gradient

(b) Local Image Contrast

(c) Proposed Method

Fig. 2. Contrast Images constructed using the local image gradient [42] in (a), the local image contrast [5] in (b), and our proposed method in (c) of the sample document images in Figure 1(b) and (d), respectively.

II. Related Work

Many thresholding techniques [6], [7], [8], [9] have been reported for document image binarization. As many degraded documents do not have a clear bimodal pattern, global thresholding [10], [11], [12], [13] is usually not a suitable approach for the degraded document binarization. Adaptive thresholding [14], [15], [16], [17], [18], [19], [20], which estimates a local threshold for each document image pixel, is often a better approach to deal with different variations within degraded document images. For example, the early window-based adaptive thresholding techniques [18], [19] estimate the local threshold by using the mean and the standard variation of image pixels within a local neighborhood window. The main drawback of these window-based thresholding techniques is that the thresholding performance depends heavily on the window size and hence the character stroke width. Other approaches have also been reported, including background subtraction [4], [21], texture analysis [22], recursive method [23], [24], decomposition method [25], contour completion [26], [27], [28], Markov Random Field [29], [30], [31], [32], matched wavelet [33], cross section sequence graph analysis [34], self-learning [35], Laplacian energy [36] user assistance [37], [38] and combination of binarization techniques [39], [40]. These methods combine different types of image information and domain knowledge and are often complex.

The local image contrast and the local image gradient are very useful features for segmenting the text from the document background because the document text usually has certain image contrast to the neighboring document background. They are very effective and have been used in many document image binarization techniques [5], [14], [18], [19]. In Bernsen's paper [14], the local contrast is defined as follows:

$$C(i,j) = I_{max}(i,j) - I_{min}(i,j) \tag{1}$$

where $C(i,j)$ denotes the contrast of an image pixel (i,j), $I_{max}(i,j)$ and $I_{min}(i,j)$ denote the maximum and minimum intensities within a local neighborhood windows of (i,j), respectively. If the local contrast $C(i,j)$ is smaller than a threshold, the pixel is set as background directly. Otherwise it will be classified into text or background by comparing with the mean of $I_{max}(i,j)$ and $I_{min}(i,j)$. Bernsen's method is simple, but cannot work properly on degraded document images with a complex document background.

We have earlier proposed a novel document image binarization method [5] by using the local image contrast that is evaluated as follows [41]:

$$C(i,j) = \frac{I_{max}(i,j) - I_{min}(i,j)}{I_{max}(i,j) + I_{min}(i,j) + \epsilon} \tag{2}$$

where ϵ is a positive but infinitely small number that is added in case the local maximum is equal to 0. Compared with Bernsen's contrast in Equation 1, the local image contrast in Equation 2 introduces a normalization factor (the denominator) to compensate the image variation within the document background. Take the text within shaded document areas such as that in the sample document image in Figure 1(b) as an example. The small image contrast around the text stroke edges in Equation 1 (resulting from the shading) will be compensated by a small normalization factor (due to the dark document background) as defined in Equation 2.

III. Proposed Method

This section describes the proposed document image binarization techniques. Given a degraded document image, an adaptive contrast map is first constructed and the text stroke edges are then detected through the combination of the binarized adaptive contrast map and the canny edge map. The text is then segmented based on the local threshold that is estimated from the detected text stroke edge pixels. Some post-processing is further applied to improve the document binarization quality.

A. Contrast Image Construction

The image gradient has been widely used for edge detection [42] and it can be used to detect the text stroke edges of the document images effectively that have a uniform document background. On the other hand, it often detects many non-stroke edges from the background of degraded document that often contains certain image variations due to noise, uneven lighting, bleed-through, etc. To extract only the stroke edges properly, the image gradient needs to be normalized to compensate the image variation within the document background.

In our earlier method [5], The local contrast evaluated by the local image maximum and minimum is used to suppress the background variation as described in Equation 2. In particular, the numerator (i.e. the difference between the local maximum and the local minimum) captures the local image difference that is similar to the traditional image gradient [42]. The denominator is a normalization factor that suppresses the image variation within the document background. For image pixels within bright regions, it will produce a large normalization factor to neutralize the numerator and accordingly result in a relatively low image contrast. For the image pixels within dark regions, it will produce a small denominator and accordingly result in a relatively high image contrast.

However, the image contrast in Equation 2 has one typical limitation that it may not handle document images with the bright text properly. This is because a weak contrast will be calculated for stroke edges of the bright text where the denominator in Equation 2 will be large but the numerator will be small. To overcome this over-normalization problem, we combine the local image contrast with the local image gradient and derive an adaptive local image contrast as follows:

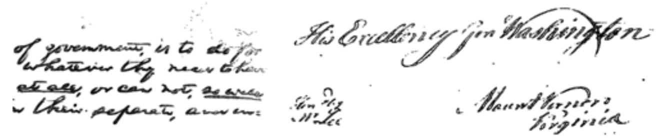
$$C_a(i,j) = \alpha C(i,j) + (1-\alpha)(I_{max}(i,j) - I_{min}(i,j)) \quad (3)$$

where $C(i,j)$ denotes the local contrast in Equation 2 and $(I_{max}(i,j) - I_{min}(i,j))$ refers to the local image gradient that is normalized to $[0,1]$. The local windows size is set to 3 empirically. α is the weight between local contrast and local gradient that is controlled based on the document image statistical information. Ideally, the image contrast will be assigned with a high weight(i.e. large α) when the document image has significant intensity variation. So that the proposed binarization technique depends more on the local image contrast that can capture the intensity variation well and hence produce good results. Otherwise, the local image gradient will be assigned with a high weight. The proposed binarization technique relies more on image gradient and avoid the over normalization problem of our previous method [5].

We model the mapping from document image intensity variation to α by a power function as follows:

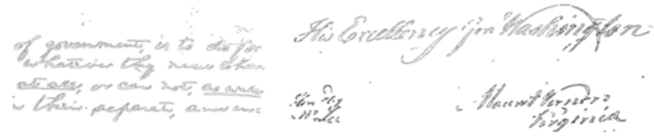
$$\alpha = \left(\frac{Std}{128}\right)^{\gamma}. \quad (4)$$

where Std denotes the document image intensity standard deviation, and γ is a pre-defined parameter. The power function has a nice property in that it monotonically and smoothly increases from 0 to 1 and its shape can be easily controlled by different γ.γ can be selected from $[0,\infty]$, where the power function becomes a linear function when $\gamma = 1$. Therefore, the local image gradient will play the major role in Equation 3 when γ is large and the local image contrast will play the major role when γ is small. The setting of parameter γ will be discussed in Section 4.



(a) Binary Contrast Map

(b) Canny Edge Map



(c) Combined Text Stroke Edge Map

Fig. 3. Binary contrast images, canny edge maps and their corresponding combined Edge maps of the sample document images in Figure 1(b) and (d), respectively.

Figure 2 shows the contrast map of the sample document images in Figure 1 (b) and (d) that are created by using local image gradient [43], local image contrast [5] and our proposed method in Equation 3, respectively.

For the sample document with a complex document background in Figure 1(b), the use of the local image contrast produces a better result as shown in Figure 2(b) compared with the result by the local image gradient as shown in Figure 2(a) (because the normalization factors in Equation 2 helps to suppress the noise at the upper left area of Figure 2(a)). But for the sample document in Figure 1(d) that has small intensity variation within the document background but large intensity variation within the text strokes, the use of the local image contrast removes many light text strokes improperly in the contrast map as shown in Figure 2(b) whereas the use of local image gradient is capable of preserving those light text strokes as shown in Figure 2(a).

As a comparison, the adaptive combination of the local image contrast and the local image gradient in Equation 3 can produce proper contrast maps for document images with different types of degradation as shown in Figure 2(c). In particular, the local image contrast in Equation 3 gets a high weight for the document image in Figure 1(a) with high intensity variation within the document background whereas the local image gradient gets a high weight for the document image in Figure 1(b).

TABLE I
EVALUATION RESULTS OF THE DATASET OF DIBCO 2009

Methods	F-Measure(%)	PSNR	NRM($\times 10^{-2}$)	MPM($\times 10^{-3}$)	Rank Score
OTSU [12]	78.72	15.34	5.77	13.3	196
SAUV [18]	85.41	16.39	6.94	3.2	177
NIBL [19]	55.82	9.89	16.4	61.5	251
BERN [14]	52.48	8.89	14.29	113.8	313
GATO [21]	85.25	16.5	10	0.7	176
LMM [5]	91.06	18.5	7	0.3	126
BE [4]	91.24	18.6	4.31	0.55	101
Proposed method	93.5	19.65	3.74	0.43	100

B. Text Stroke Edge Pixel Detection

The purpose of the contrast image construction is to detect the stroke edge pixels of the document text properly. The constructed contrast image has a clear bi-modal pattern [5], where the adaptive image contrast computed at text stroke edges is obviously larger than that computed within the document background. We therefore detect the text stroke edge pixel candidate by using Otsu's global thresholding method. For the contrast images in Figure 2(c), Figure 3(a) shows a binary map by Otsu's algorithm that extracts the stroke edge pixels properly.

As the local image contrast and the local image gradient are evaluated by the difference between the maximum and minimum intensity in a local window, the pixels at both sides of the text stroke will be selected as the high contrast pixels. The binary map can be further improved through the combination with the edges by Canny's edge detector [43], because Canny's edge detector has a good localization property that it can mark the edges close to real edge locations in the detecting image. In addition, canny edge detector uses two adaptive thresholds and is more tolerant to different imaging artifacts such as shading [44]. It should be noted that Canny's edge detector by itself often extracts a large amount of non-stroke edges as illustrated in Figure 3(b) without tuning the parameter manually. In the combined map, we keep only pixels that appear within both the high contrast image pixel map and canny edge map. The combination helps to extract the text stroke edge pixels accurately as shown in Figure 3(c).

C. Local Threshold Estimation

The text can then be extracted from the document background pixels once the high contrast stroke edge pixels are detected properly. Two characteristics can be observed from different kinds of document images [5]: First, the text pixels are close to the detected text stroke edge pixels. Second, there is a distinct intensity difference between the high contrast stroke edge pixels and the surrounding background pixels. The document image text can thus be extracted based on the detected text stroke edge pixels as follows:

$$R(x,y) = \begin{cases} 1 & I(x,y) \le E_{mean} + \frac{E_{std}}{2} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where E_{mean} and E_{std} are the mean and standard deviation of the intensity of the detected text stroke edge pixels within a neighborhood window W, respectively.

The neighborhood window should be at least larger than the stroke width in order to contain stroke edge pixels. So the size of the neighborhood window W can be set based on the

---

Algorithm 1 Edge Width Estimation

Require: The Input Document Image I and Corresponding Binary Text Stroke Edge Image Edg

Ensure: The Estimated Text Stroke Edge Width EW

1: Get the $width$ and $height$ of I
2: **for** Each Row $i = 1$ to $height$ in Edg **do**
3: Scan from left to right to find edge pixels that meet the following criteria:
- its label is 0(background)
- the next pixel is labeled as 1(edge).

4: Examine the intensities in I of those pixels selected in Step 3, and remove those pixels that have a lower intensity than the following pixel next to it in the same row of I.
5: Match the remaining adjacent pixels in the same row into pairs, and calculate the distance between the two pixels in pair.
6: **end for**
7: Construct a histogram of those calculated distances.
8: Use the most frequently occurring distance as the estimated stroke edge width EW.

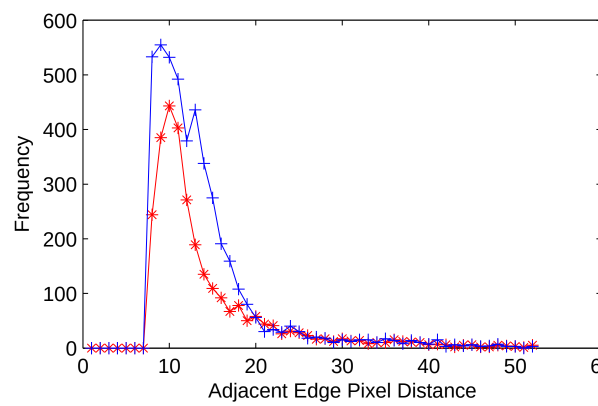


Fig. 4. The histogram of the distance between adjacent edge pixels. The '+++' line denotes the histogram of the image in Figure 1(b), The '***' line denotes the histogram of the image in Figure 1(d)

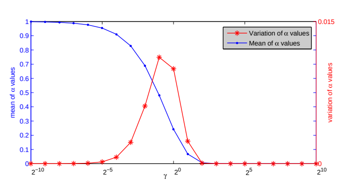
stroke width of the document image under study, EW, which can be estimated from the detected stroke edges (shown in Figure 3(b)) as stated in Algorithm 1.

Since we do not need a precise stroke width, we just calculate the most frequently distance between two adjacent edge pixels (which denotes two sides edge of a stroke) in horizontal direction and use it as the estimated stroke width. First the edge image is scanned horizontally row by row and the edge pixel candidates are selected as described in step 3. If the edge pixels, which are labeled 0 (background) and the pixels next to them are labeled to 1 (edge) in the edge map (Edg), are correctly detected, they should have higher intensities than the following few pixels (which should be the text stroke pixels). So those improperly detected edge pixels are removed in step 4. In the remaining edge pixels in the same row, the two adjacent edge pixels are likely the two sides of a

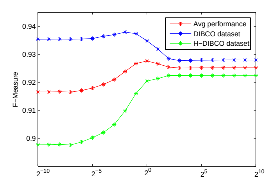
Algorithm 2 Post Processing Procedure

Require: The Input Document Image I, Initial Binary Result B and Corresponding Binary Text Stroke Edge Image Edg
Ensure: The Final Binary Result B_f

1: Find out all the connect components of the stroke edge pixels in Edg
2: Remove those pixels that do not connect with other pixels.
3: **for** Each remaining edge pixels (i, j): **do**
4: Get its neighborhood pairs: $(i-1, j)$ and $(i+1, j)$; $(i, j-1)$ and $(i, j+1)$
5: **if** The pixels in the same pairs belong to the same class (both text or background) **then**
6: Assign the pixel with lower intensity to foreground class (text), and the other to background class.
7: **end if**
8: **end for**
9: Remove single-pixel artifacts [4] along the text stroke boundaries after the document thresholding.
10: Store the new binary result to B_f.



(a) The mean and variation of α values on DIBCO 2009 & H-DIBCO 2010 datasets



(b) F-Measure Performance on DIBCO 2009 & H-DIBCO 2010 datasets

Fig. 5. (a) shows the means and variations of the α values of the twenty images on DIBCO and H-DIBCO dataset under different γ values. (b) illustrates the F-Measure Performance on DIBCO 2009 & H-DIBCO 2010 datasets using different γ power functions

stroke, so these two adjacent edge pixels are matched to pairs and the distance between them are calculated in step 5. After that a histogram is constructed that records the frequency of the distance between two adjacent candidate pixels. The stroke edge width EW can then be approximately estimated by using the most frequently occurring distances of the adjacent edge pixels as illustrated in Figure 4.

D. Post-Processing

Once the initial binarization result is derived from Equation 5 as described in previous subsections, the binarization

TABLE II
EVALUATION RESULTS OF THE DATASET OF H-DIBCO 2010

Methods	F-Measure(%)	pseudo F-Measure(%)	PSNR	NRM $(\times 10^{-2})$	MPM $(\times 10^{-3})$	Rank Score
OTSU [12]	85.27	90.83	17.51	9.77	1.35	188
SAUV [18]	75.3	84.22	15.96	16.31	1.96	225
NIBL [19]	74.1	85.4	15.73	19.06	1.06	263
BERN [14]	41.3	44.4	8.57	21.18	115.98	244
GATO [21]	71.99	74.35	15.12	21.89	0.41	284
LMM [5]	85.49	92.64	17.83	11.46	0.37	216
BE [4]	86.41	88.25	18.14	9.06	1.11	202
Proposed method	92.03	94.85	20.12	6.14	0.25	178

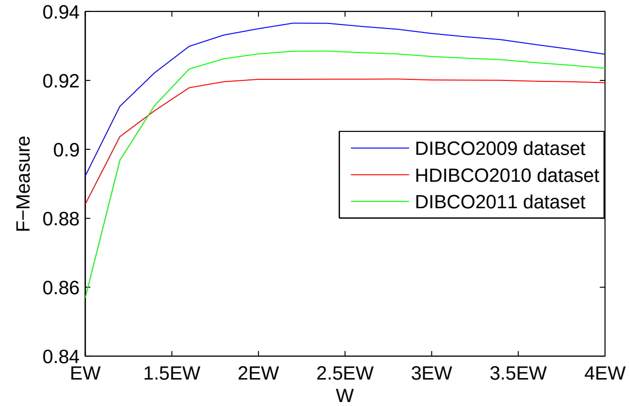


Fig. 6. F-Measure Performance on DIBCO 2009, H-DIBCO 2010 and DIBCO 2011 datasets using different Local Window Size W (the EW denotes the estimated text stroke width).

result can be further improved by incorporating certain domain knowledge as described in Algorithm 2. First, the isolated foreground pixels that do not connect with other foreground pixels are filtered out to make the edge pixel set precisely. Second, the neighborhood pixel pair that lies on symmetric sides of a text stroke edge pixel should belong to different classes (i.e., either the document background or the foreground text). One pixel of the pixel pair is therefore labeled to the other category if both of the two pixels belong to the same class. Finally, some single-pixel artifacts along the text stroke boundaries are filtered out by using several logical operators as described in [4].

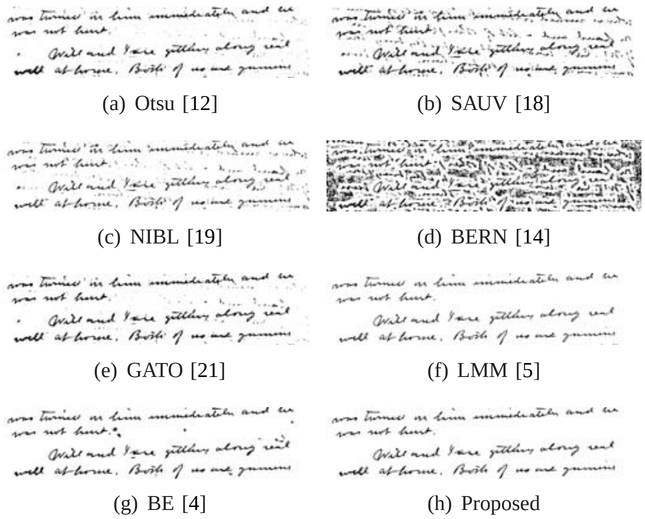
IV. EXPERIMENTS AND DISCUSSION

A few experiments are designed to demonstrate the effectiveness and robustness of our proposed method. We first analyze the performance of the proposed technique on public datasets for parameter selection. The proposed technique is then tested and compared with state-of-the-art methods over on three well-known competition datasets: DIBCO 2009 dataset [1], H-DIBCO 2010 dataset [3], and DIBCO 2011 dataset [2]. Finally, the proposed technique is further evaluated over a very challenging Bickley diary dataset [37].

The binarization performance are evaluated by using F-Measure, pseudo F-Measure, Peak Signal to Noise Ratio (PSNR), Negative Rate Metric (NRM), Misclassification Penalty Metric (MPM), Distance Reciprocal Distortion (DRD) and rank score that are adopted from DIBCO 2009, H-DIBCO 2010 and DIBCO 2011 [1], [2], [3]. Due to lack of ground truth data in some datasets, no all of the metrics are applied on every images.

A. Parameter Selection

In the first experiment, we apply different γ to obtain different power functions and test their performance under the DIBCO 2009 and H-DIBCO 2010 datasets. The γ increases from 2^{-10} to 2^{10} exponentially and monotonically as shown in Figure 5(a). In particular, α is close to 1 when γ is small



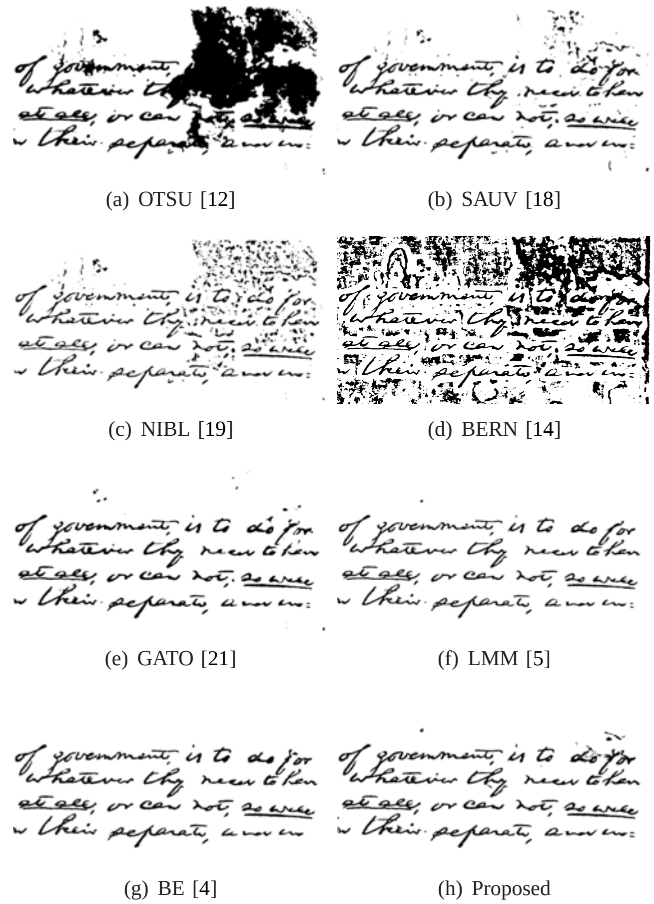
(a) Otsu [12] (b) SAUV [18]

(c) NIBL [19] (d) BERN [14]

(e) GATO [21] (f) LMM [5]

(g) BE [4] (h) Proposed

Fig. 7. Binarization Results of the sample document image in Figure 1(a) produced by different methods.



(a) OTSU [12] (b) SAUV [18]

(c) NIBL [19] (d) BERN [14]

(e) GATO [21] (f) LMM [5]

(g) BE [4] (h) Proposed

Fig. 8. Binarization Results of the sample document image in Figure 1(b) produced by different methods.

and the local image contrast C dominates the adaptive image contrast C_a in Equation 3. On the other hand, C_a is mainly influenced by local image gradient when γ is large. At the same time, the variation of α for different document images increases when γ is close to 1. Under such circumstance, the power function becomes more sensitive to the global image intensity variation and appropriate weights can be assigned to images with different characteristics.

As shown in Figure 5(b), Our proposed method produces better results on DIBCO dataset when the γ is much smaller than 1 and the local image contrast dominates. On the other hand, the F-Measure performance of H-DIBCO dataset improves significantly when γ increases to 1. Therefore the proposed method can assign more suitable α to different images when γ is closer to 1. Parameter γ should therefore be set around 1 when the adaptability of the proposed technique is maximized and better and more robust binarization results can be derived from different kinds of degraded document images.

Another parameter, i.e., the local window size W, is tested in the second experiment on the DIBCO 2009, H-DIBCO 2010 and DIBCO 2011 datasets. W is closely related to the stroke width EW. Figure 6 shows the thresholding results when W varies from EW to $4EW$. Generally, a larger local window size will help to reduce the classification error that is often induced by the lack of edge pixels within the local neighborhood window. In addition, the performance of the proposed method becomes stable when the local window size is larger than $2EW$ consistently on the three datasets. W can therefore be set around $2EW$ because a larger local neighborhood window will increase the computational load significantly.

B. Testing on competition datasets

In this experiment, we quantitatively compare our proposed method with other state-of-the-art techniques on DIBCO 2009, H-DIBCO 2010 and DIBCO 2011 datasets. These methods include Otsu's method (OTSU) [12], Sauvola's method (SAUV) [18], Niblack's method (NIBL) [19], Bernsen's method (BERN) [14], Gatos et al.'s method (GATO) [21], and our previous methods(LMM [5], BE [4]). The three datasets are composed of the same series of document images that suffer from several common document degradations such as smear, smudge, bleed-through and low contrast. The DIBCO 2009 dataset contains ten testing images that consist of five degraded handwritten documents and five degraded printed documents. The H-DIBCO 2010 dataset consists of ten degraded handwritten documents. The DIBCO 2011 dataset contains eight degraded handwritten documents and eight degraded printed documents. In total, we have 36 degraded document images with ground truth.

The evaluation results are shown in Table III-C, III-C and IV-B. As Table III-C and III-C show, our proposed method achieves the highest scores in F-Measure, pseudo F-Measure, PSNR, and NRM and its MPM is only slightly lower than LMM [5] under DIBCO dataset. This means that our proposed method produces a higher overall precision and preserves the text strokes better. In addition, our proposed method also performs better than the 43 document thresholding algorithms submitted to the DIBCO 2009 [1] under DIBCO 2009 dataset and the 17 submitted algorithms in the H-DIBCO 2010 [3] under the H-DIBCO 2010 dataset. Figures 7, 8, 9 and 10 further show the binarization results of the four example document images in Figure 1 by using the eight document binarization methods. We also use rank score to evaluate the methods, which is to accumulate the ranking of the method within all the testing methods over all testing images and the four evaluation metrics as described in DIBCO 2011 [2]. Based on this ranking score scheme, the performance of our proposed method is relative to other methods to compare. It's clear that our proposed method extracts the text better than the other comparison methods.

Besides the comparison methods mentioned above, our proposed method is also compared with the top three algorithms, namely Lelore et al.'s method (LELO) [45], the method submitted by our team (SNUS) and N. Howe's method (HOWE) [36] for the DIBCO 2011 dataset. The quantitative results are shown in Table IV-B. As Table IV-B shown, Our proposed technique performs the best in terms of DRD and MPM, which means that our proposed technique maintains good text stroke contours and provides best visual quality. In addition, our proposed method also performs well when being evaluated in pixel level. The F-Measure and PSNR of our proposed method are very close to the highest scores, which is also shown in Table IV-B. Although it does not reach the lowest ranking score, our proposed technique produces good results on all the testing images, which is reflected on the high F-measure score.

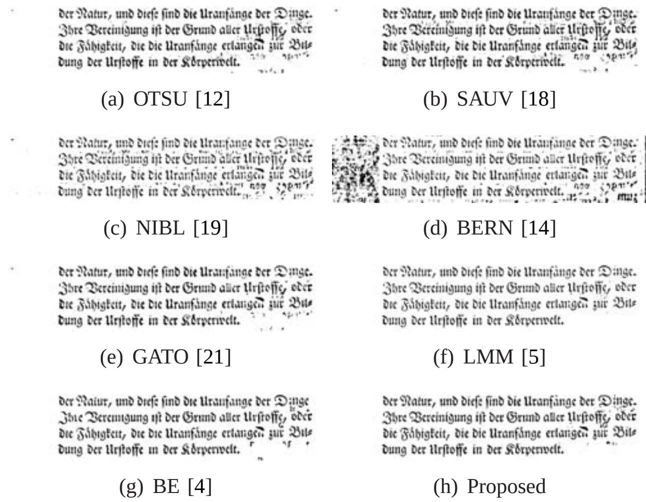
Figure 11, 12, 13 further show three example images (PR06, PR07, and HW06) from the DIBCO 2011 dataset and its corresponding binary results produced by different methods. As shown in Figure 11, BERN, NIBL and LELO method fail to produce reasonable results. In addition, most of the methods including HOWE method induce some background noise in the final results. LMM and SNUS instead remove too much character strokes. On the other hand, our proposed method produces a binary result with better visual quality and contains most of the text information. Figure 12 and 13 are more challenging, some of the methods fail, including LELO, BERN, NIBL and SNUS method. HOWE method and our proposed method produce quite reasonable results with a little noise remains, compared with other methods. However, the binary result of our proposed method in Figure 13 is a little over-binarized due to the high text stroke variation of the input image. We will improve it in our future study.

In addition, we test the computation time of our proposed method and other state-of-the-art techniques implemented in Matlab. Experiments over DIBCOs test dataset shown that the average execution time of the proposed method is around 21 seconds. The execution time of OTSU, BERN, NIBL, SAUV, GATO, BE and LMM methods are around 0.5 seconds, 18 seconds, 27 seconds, 28 seconds, 100 seconds, 24 seconds and 20 seconds, respectively. The proposed technique is comparable to the state-of-art adaptive document thresholding methods.

C. Testing on Bickley diary dataset

In the last experiment, we evaluate our method on the Bickley diary dataset to show its robustness and superior performance. The images from Bickley diary dataset are taken from a photocopy of a diary that is written about 100 years ago. These images suffer from different kinds of degradation, such as water stains, ink bleed-through, and significant foreground text intensity and are more challenging than the previous two DIBCO and H-DIBCO datasets.

We use seven ground truth images that are annotated manually using Pix Labeler [46] to evaluate our proposed method with the other methods. Our proposed method achieves average 78.54% accuracy in terms of F-measure, which is at least 10% higher than the other seven methods. Detailed evaluation results are illustrated in Table IV. Figure 14 shows one example image from the Bickley diary dataset and its corresponding binarization results generated by different meth-



(a) OTSU [12] (b) SAUV [18] (c) NIBL [19] (d) BERN [14] (e) GATO [21] (f) LMM [5] (g) BE [4] (h) Proposed

Fig. 9. Binarization Results of the sample document image in Figure 1(c) produced by different methods.

(a) OTSU [12] (b) SAUV [18] (c) NIBL [19] (d) BERN [14] (e) GATO [21] (f) LMM [5] (g) BE [4] (h) Proposed

Fig. 10. Binarization Results of the sample document image in Figure 1(d) produced by different methods.

TABLE III
EVALUATION RESULTS OF THE DATASET OF DIBCO 2011

Methods	F-Measure(%)	PSNR	DRD	MPM	Rank Score
OTSU [12]	82.22	15.77	8.72	15.64	412
SAUV [18]	82.54	15.78	8.09	9.20	403
NIBL [19]	68.52	12.76	28.31	26.38	362
BERN [14]	47.28	7.92	82.28	136.54	664
GATO [21]	82.11	16.04	5.42	7.13	353
LMM [5]	85.56	16.75	6.02	6.42	516
BE [4]	81.67	15.59	11.24	11.40	376
LELO [45]	80.86	16.13	104.48	64.43	252
SNUS	85.2	17.16	15.66	9.07	279
HOWE [36]	88.74	17.84	5.37	8.64	299
Proposed	87.8	17.56	**4.84**	**5.17**	307

(a) Input Image (b) OTSU [12] (c) SAUV [18] (d) NIBL [19]
(e) BERN [14] (f) GATO [21] (g) LMM [5] (h) BE [4]
(i) LELO [45] (j) SNUS (k) HOWE [36] (l) Proposed

Fig. 11. Binarization Results of the sample document image (PR 06) in DIBCO 2011 dataset produced by different methods.

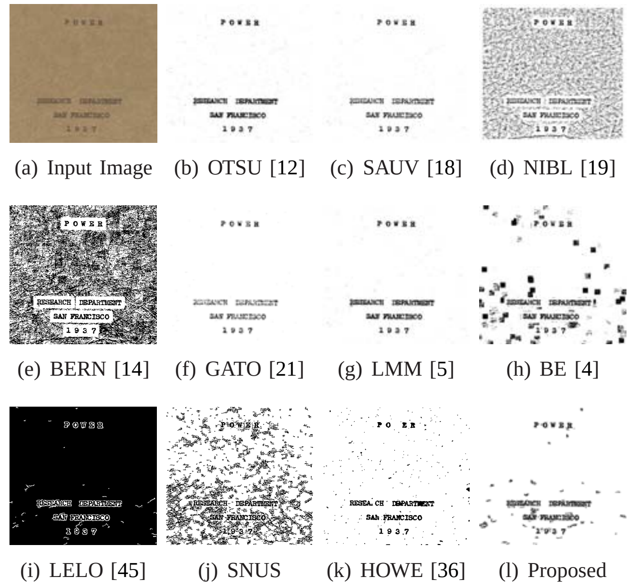
ods. It's clear that the proposed algorithm performs better than other methods by preserving most textual information and producing the least noise.

D. Discussion

As described in previous sections, the proposed method involves several parameters, most of which can be automatically estimated based on the statistics of the input document image. This makes our proposed technique more stable and easy-to-use for document images with different kinds of degradation. The superior performance of our proposed method can be explained by several factors. First, the proposed method combines the local image contrast and the local image gradient that help to suppress the background variation and avoid the over-normalization of document images with less variation. Second, the combination with edge map helps to produce a precise text stroke edge map. Third, the proposed method makes use of the text stroke edges that help to extract the foreground text from the document background accurately. But the performance on Bickley diary dataset and some images of DIBCO contests still needs to be improved, we will explore it in future.

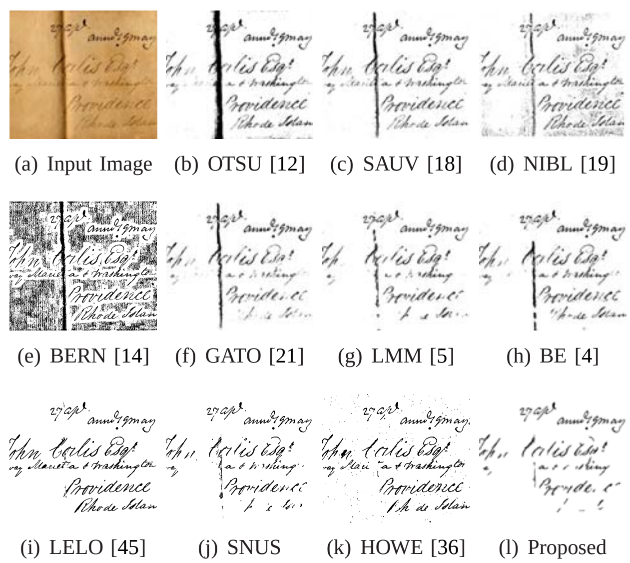
V. CONCLUSIONS

This paper presents an adaptive image contrast based document image binarization technique that is tolerant to different types of document degradation such as uneven illumination



(a) Input Image (b) OTSU [12] (c) SAUV [18] (d) NIBL [19]
(e) BERN [14] (f) GATO [21] (g) LMM [5] (h) BE [4]
(i) LELO [45] (j) SNUS (k) HOWE [36] (l) Proposed

Fig. 12. Binarization Results of the sample document image (PR 07) in DIBCO 2011 dataset produced by different methods.



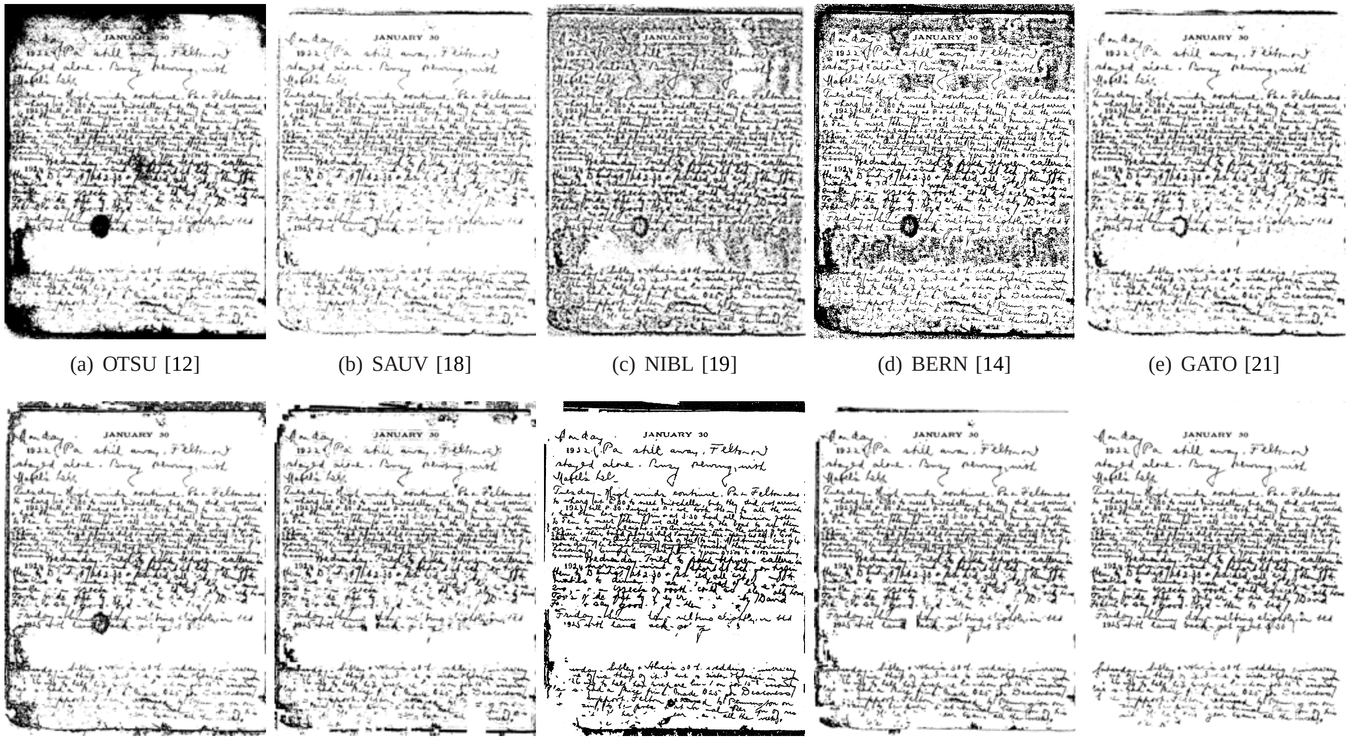
(a) Input Image (b) OTSU [12] (c) SAUV [18] (d) NIBL [19]
(e) BERN [14] (f) GATO [21] (g) LMM [5] (h) BE [4]
(i) LELO [45] (j) SNUS (k) HOWE [36] (l) Proposed

Fig. 13. Binarization Results of the sample document image (HW 06) in DIBCO 2011 dataset produced by different methods.

and document smear. The proposed technique is simple and robust, only few parameters are involved. Moreover, it works for different kinds of degraded document images. The proposed technique makes use of the local image contrast that is evaluated based on the local maximum and minimum. The proposed method has been tested on the various datasets. Experiments show that the proposed method outperforms most reported document binarization methods in term of the F-measure, pseudo F-measure, PSNR, NRM, MPM and DRD.

ACKNOWLEDGEMENTS

The Bickley diary dataset is originally made from the diary of Ms. Anna Felton Bickley from January to March, 1922 to



(a) OTSU [12] (b) SAUV [18] (c) NIBL [19] (d) BERN [14] (e) GATO [21]

(f) LMM [5] (g) BE [4] (h) HOWE [36] (i) Proposed (j) Ground Truth Image

Fig. 14. Binary results of the badly degraded document image from Bickley diary dataset shown in Figure 1(e) produced by different binarization methods and the ground truth image.

TABLE IV
EVALUATION RESULTS OF BICKLEY DIARY DATASET

Methods	F-Measure(%)	PSNR	NRM	MPM
OTSU [12]	50.42	7.58	21.41	196.98
SAUV [18]	64.60	11.62	23.26	28.97
NIBL [19]	67.71	9.79	9.52	105.17
BERN [14]	52.97	7.71	18.86	193.35
GATO [21]	69.13	11.44	21.89	36.57
LMM [5]	66.44	10.76	17.50	72.08
BE [4]	34.65	3.54	40.78	370.15
Proposed	78.54	13.15	12.92	16.71

1926, which is a Private Journal (unpublished) copy donated to the Methodist Church Archives, Singapore by Ervin Bickley, Jr. We gratefully acknowledge Mr. Ervin Bickley, Jr (grandson of Bishop Bickley) to make the dataset public for us to use. We also thank Prof. Michael S. Brown and his team for making the ground truth images of the Bickley diary dataset.

REFERENCES

[1] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "ICDAR 2009 document image binarization contest(DIBCO 2009)," *International Conference on Document Analysis and Recognition*, pp. 1375–1382, July 2009.

[2] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2011 document image binarization contest (DIBCO 2011)," *International Conference on Document Analysis and Recognition*, September 2011.

[3] ——, "H-DIBCO 2010 handwritten document image binarization competition," *International Conference on Frontiers in Handwriting Recognition*, pp. 727–732, November 2010.

[4] S. Lu, B. Su, and C. L. Tan, "Document image binarization using background estimation and stroke edges," *International Journal on Document Analysis and Recognition*, vol. 13, pp. 303–314, December 2010.

[5] B. Su, S. Lu, and C. L. Tan, "Binarization of historical handwritten document images using local maximum and minimum filter," *International Workshop on Document Analysis Systems*, pp. 159–166, June 2010.

[6] G. Leedham, C. Yan, K. Takru, J. Hadi, N. Tan, and L. Mian, "Comparison of some thresholding algorithms for text/background segmentation in difficult document images," *International Conference on Document Analysis and Recognition*, vol. 13, pp. 859–864, 2003.

[7] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–165, 2004.

[8] O. D. Trier and A. K. Jain, "Goal-directed evaluation of binarization methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1191–1201, 1995.

[9] O. D. Trier and T. Taxt, "Evaluation of binarization methods for document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 312–315, 1995.

[10] A. Brink, "Thresholding of digital images using two-dimensional entropies," *Pattern Recognition*, vol. 25, no. 8, pp. 803–808, 1992.

[11] J. Kittler and J. Illingworth, "On threshold selection using clustering criteria," *IEEE transactions on Systems, Man, and Cybernetics*, vol. 15, pp. 652–655, 1985.

[12] N. Otsu, "A threshold selection method from gray level histogram," *IEEE Transactions on System, Man, Cybernetics*, vol. 19, no. 1, pp. 62–66, January 1978.

[13] N. Papamarkos and B. Gatos, "A new approach for multithreshold selection," *Computer Vision Graphics and Image Processing*, vol. 56, no. 5, pp. 357–370, 1994.

[14] J. Bernsen, "Dynamic thresholding of gray-level images," *Internation Conference on Pattern Recognition*, pp. 1251–1255, October 1986.

[15] L. Eikvil, T. Taxt, and K. Moen, "A fast adaptive method for binarization of document images," *International Conference on Document Analysis and Recognition*, pp. 435–443, September 1991.

[16] I.-K. Kim, D.-W. Jung, and R.-H. Park, "Document image binarization based on topographic analysis using a water flow model," *Pattern Recognition*, vol. 35, no. 1, pp. 265–277, 2002.

[17] J. Parker, C. Jennings, and A. Salkauskas, "Thresholding using an illumination model," *International Conference on Document Analysis and Recognition*, pp. 270–273, October 1993.

[18] J. Sauvola and M. Pietikainen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225–236, 2000.

[19] W. Niblack, *An Introduction to Digital Image Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1986.

[20] J.-D. Yang, Y.-S. Chen, and W.-H. Hsu, "Adaptive thresholding algorithm and its hardware implementation," *Pattern Recognition Letters*, vol. 15, no. 2, pp. 141–150, 1994.

[21] B. Gatos, I. Pratikakis, and S. Perantonis, "Adaptive degraded document image binarization," *Pattern Recognition*, vol. 39, no. 3, pp. 317–327, 2006.

[22] Y. Liu and S. Srihari, "Document image binarization based on texture features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 540–544, May 1997.

[23] M. Cheriet, J. N. Said, and C. Y. Suen, "A recursive thresholding technique for image segmentation," *IEEE Transactions on Image Processing*, pp. 918–921, June 1998.

[24] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Iterative multimodel subimage binarization for handwritten character segmentation," *IEEE Transactions on Image Processing*, vol. 13, pp. 1223–1230, September 2004.

[25] Y. Chen and G. Leedham, "Decompose algorithm for thresholding degraded historical document images," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 152, no. 6, pp. 702–714, December 2005.

[26] Q. Chen, Q. Sun, H. Pheng Ann, and D. Xia, "A double-threshold image binarization method based on edge detector," *Pattern Recognition*, vol. 41, no. 4, pp. 1254–1267, 2008.

[27] R. Cao, C. L. Tan, Q. Wang, and P. Shen, "Double-sided handwritten archival documents," *International Workshop on Document Analysis Systems*, pp. 147–158, 2000.

[28] I. Blayvas, A. Bruckstein, and R. Kimmel, "Efficient computation of adaptive threshold surface for image binarization," *Pattern Recognition*, vol. 39, no. 1, pp. 89–101, 2006.

[29] S. Nicolas, J. Dardenne, T. Paquet, and L. Heutte, "Document image segmentation using a 2d conditional random field model," *International Conference on Document Analysis and Recognition*, pp. 407–411, September 2007.

[30] C. Wolf and D. Doermann, "Binarization of Low Quality Text using a Markov Random Field Model," *International Conference on Pattern Recognition*, pp. 160–163, 2002.

[31] T. Lelore and F. Bouchara, "Document image binarisation using markov field model," *International Conference on Document Analysis and Recognition*, pp. 551–555, July 2009.

[32] J. G. Kuk, N. I. Cho, and K. M. Lee, "Map-MRF approach for binarization of degraded document image," *International Conference on Image Processing*, pp. 2612–2615, 2008.

[33] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text extraction and document image segmentation using matched wavelets and mrf model," *IEEE Transactions on Image Processing*, vol. 16, pp. 2117–2128, 2007.

[34] A. Dawoud, "Iterative cross section sequence graph for handwritten character segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2150–2154, August 2007.

[35] B. Su, S. Lu, and C. L. Tan, "A self-training learning document binarization framework," *International Conference on Pattern Recognition*, pp. 3187–3190, August 2010.

[36] N. Howe, "A laplacian energy for document binarization," *International Conference on Document Analysis and Recognition*, September 2011.

[37] F. Deng, Z. Wu, Z. Lu, and M. S. Brown, "Binarizatioinshop: A user-assisted software suite for converting old documents to black-and-white," *Annual Joint Conference on Digital Libraries*, 2010.

[38] H. Yi, B. M. S., and X. Dong, "User-assisted ink-bleed reduction," *IEEE Transactions on Image Processing*, vol. 19, pp. 2646–2658, October 2010.

[39] E. Badekas and N. Papamarkos, "Optimal combination of document binarization techniques using a selforganizing map neural network." *Engineering Applications of Artificial Intelligence*, vol. 20, pp. 11–24, 2007.

[40] B. Gatos, I. Pratikakis, and S. Perantonis, "Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information," *International Conference on Pattern Recognition*, pp. 1–4, 2008.

[41] M. van Herk, "A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels," *Pattern Recognition Letters*, vol. 13, no. 7, pp. 517–521, 1992.

[42] D. Ziou and S. Tabbone, "Edge detection techniques - an overview," *International Journal of Pattern Recognition and Image Analysis*, vol. 8, no. 4, pp. 537–559, 1998.

[43] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, January 1986.

[44] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *International Jounal on Computer Vision*, vol. 30, no. 2, pp. 117–156, 1998.

[45] T. Lelore and F. Bouchara, "Super-resolved binarization of text based on the fair algorithm." in *International Conference on Document Analysis and Recognition*, September 2011, pp. 839–843.

[46] E. Saund, J. Lin, and P. Sarkar, "Pixlabeler: user interface for pixel-level labeling of elements in document images," *International Conference on Document Analysis and Recognition*, pp. 646–650, July 2009.

Bolan Su is currently a Research Associate in the Department of Computer Science, School of Computing, National University of Singapore. He received his B.Sc. degree in computer science in 2008 from the Fudan University, Shanghai, China, and his Ph.D. degree in computer science in 2012 from the National University of Singapore, Singapore. His research interests include document image analysis, medical image analysis and computer vision.

Shijian Lu is currently a Research Scientist in the Institute for Infocomm Research, A*STAR, Singapore. He received his Ph.D. degree in electrical and computer engineering in 2005 from the National University of Singapore, Singapore. His research interests include document image analysis, medical image analysis, and bio-inspired computer vision. He has published up to 80 research papers in these areas. He has served in the program committees of many international conferences such as International Conference on Document Analysis and Recognition (ICDAR), International Conference on Pattern Recognition (ICPR), International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), etc.

Chew Lim Tan is a Professor in the Department of Computer Science, School of Computing, National University of Singapore. He received his B.Sc. (Hons) degree in physics in 1971 from University of Singapore, his M.Sc. degree in radiation studies in 1973 from University of Surrey, UK, and his Ph.D. degree in computer science in 1986 from University of Virginia, U.S.A. His research interests include document image analysis and natural language processing. He has published more than 400 research publications in these areas. He is Associate Editor of ACM Transactions on Asian Language Information Processing, Pattern Recognition, and Editorial Board Member of International Journal on Document Analysis and Recognition. He is a fellow of the International Association of Pattern Recognition (IAPR) and a member of the Governing Board of IAPR. He is also a senior member of IEEE.