Binarization of Degraded Document Images Using Morphological Operators and Ternary Entropy-Based Approach

T. Hoang Ngan Le, Tien D. Bui, Ching Y. Suen

***Abstract*—This paper presents two novel adaptive binarization algorithms using a ternary entropy-based approach[1]. The proposed methods are able to deal with image degradations which can occur due to bleeding-through ink, large black border, fading ink, uneven illumination, contrast variation, smear, and various pattern backgrounds. Given an input image, the contrast of intensity is first estimated by a grayscale morphological closing operator. A double-threshold is generated by our Shannon entropy-based thresholding methods corresponding to 1-D histogram and 2-D histogram to classify pixels into text, near-text, and non-text categories. The pixels in the second group are relabeled by the local mean and the standard deviation values. Our proposed methods classify noise into two classes, which are dealt with by binary morphological operators, shrink and swell filters, and a graph searching strategy. The methods have been tested with four databases that have been used in DIBCO 2009, H-DBCIO 2010, ICFHR 2010, and Carlos et al.'s (2008) database. The evaluation is based upon nine distinct measures. The experimental results show that our proposed methods outperform other state-of-the-art schemes.**

Keywords* - *Binarization, Ternarization, Morphological Operators, Shannon Entropy-based 1-D histogram, Shannon Entropy-based 2-D histogram, Otsu Thresholding

I. INTRODUCTION *(HEADING 1)*

Binarization means converting a multi-coloured image into a bi-coloured one by threshold selection techniques that classify the pixels of the image into either one (white - background) or zero (black - foreground). It has been studied for many years and plays an important role in document analysis systems. As more and more documents are digitized, less time consuming and higher accuracy in document image binarization have become increasingly necessary. A vast number of historical and badly degraded document images can be easily found in libraries, public archives, and national archives. Due to the complex nature of different artifacts, such poor quality documents are hard to read and process. Figs. 1-4 show some examples of degraded document images which have been used in the Document Image Binarization Contest 2009 (DIBCO 2009), the Handwriting Document Image Binarization Contest 2010 (H-DBCIO 2010), the International Conference on Frontier in Handwriting Recognition 2010 (ICFHR 2010), and Carlos et al.'s database. Much research on binarization has been conducted and reported in the literature [6, 12, 20, 28]. Generally, binarization techniques can be divided into two categories, namely, global thresholding and local thresholding, which are described below.

A. Global Thresholding Techniques

The global thresholding techniques partition an entire image based on a single threshold value, which is obtained by the gray-level histogram of the image. The pixels of the image (I) are separated into either foreground or background using the global threshold (T), as shown in Eq.1, below.

$$B(x,y) = \begin{cases} 0 & if \quad I(x,y) \le T \\ 1 & if \quad I(x,y) > T \end{cases}, \quad (1)$$

where I(x, y) is the intensity of the input image and B(x, y) is the binarized value of the output image. Three popular global thresholding techniques with different approaches are described as follows:

Fixed Thresholding Binarization: This technique binarizes with respect to a specified threshold value. This approach works well when the background and foreground intensities are clearly distinct and uniform over the image. The default threshold value of 127 which is midway between 0 and 255was selected for grayscale image in the earlier methods.

Histogram Shape-based Binarization: This technique uses the minimum (or the valley) between two peaks of the histogram as a threshold value. This approach requires some knowledge of noise and the distance between peaks; thus, it is not suitable when peaks of the histogram are not clearly determined. To address this matter, some more pattern recognition methods are used to optimize the foreground/background separation. An example of this approach is the "convex hull thresholding" scheme proposed by Rosenfeld and Torre [1].

Optimal Thresholding Binarization: This technique applies a criterion function to each pixel in order to separate an image into two regions, corresponding to white and black. The techniques belonging to this group can be classified into some sub-categories, such as: discriminant method, entropy thresholding, moment preservation approach, and minimum error thresholding. Among all the schemes in the first sub-category, the discriminant method, Otsu's algorithm [22] is considered as the most popular one. In this system, the threshold value is formulated by a pattern recognition

[1] This paper has been developed from its preliminary version, which will be presented at the ICDAR 2011 conference.

The authors are with Centre for Pattern Recognition and Machine Intelligence, Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West Montreal, Quebec, Canada (H3G 1M8).

E-mail: {l_thihoa, bui, suen}@encs.concordia.ca

Figure 1. Document images (a-j) from DIBCO 2009 database

Figure 2. Handwritten document images (a-j) from H-DIBCO 2010 database

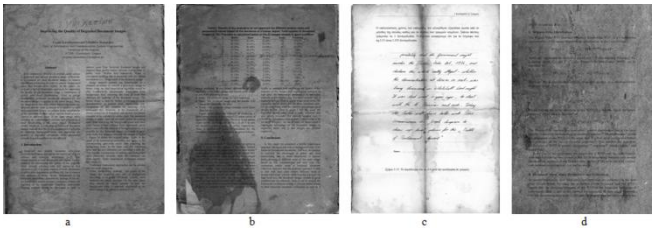


Figure 3. Representative images (a-d) from ICFHR 2010 database

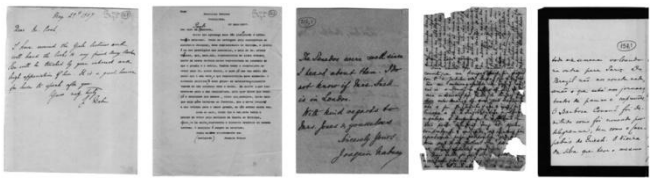


Figure 4. Representative images (a-e) from Carlos et al.'s database

procedure, in which a particular criterion function is used as a measure of the statistical separation between foreground and background classes. A common concept in Information Theory, named entropy, has received significant interest in image thresholding and has been categorized into two classes: either Shannon's entropy or relative entropy. The optimized threshold is selected to either maximize the Shannon's entropy over the image (Shannon-based) [23] or to minimize the difference between two different sources (relative-based) [9]. Researchers in [17, 19, 31, 32] have tried to choose an appropriate initial threshold and suitable parameters by testing them on numerous training and testing images. The third sub-category is called moment preservation, in which the thresholded image best preserves the mathematical moment of the original image [3, 13, 30]. The last sub-category, called minimum error thresholding, makes use of a histogram, which is considered to be a combination of two normally distributed classes of pixel values. An iterative process is used to adjust the two classes and to minimize the classification error [18]. This approach is really robust and effective when the sizes of classes are completely different, i.e., one class is much smaller or bigger than the other.

B. Local Thresholding Techniques

Local thresholding techniques apply a separate threshold value ($T(x, y)$) to a single pixel or a particular region. The binarization process can be expressed as in Eq. 2, below.

$$B(x,y)=\begin{cases}0 & if \quad I(x,y)\le T(x,y)\\ 1 & if \quad I(x,y)>T(x,y)\end{cases}. \quad (2)$$

The local threshold value can be calculated by using different information from the input image. This is also known as dynamic thresholding, and can be divided into different approaches, such as: background subtraction [8], water flow model [11, 24], illumination model [24, 16], the mean and standard derivation of pixel values within a local window [15], and local image contrast [7]. Some drawbacks of the local thresholding techniques include: region size dependency, individual image characteristics, and time-consuming. Therefore, some researchers have used a hybrid approach that applies both global and local thresholding methods [5]. In general, local methods achieve better results than global ones, however, they often rely on some specific parameters and have a high computational cost.

In this paper, we propose a novel nonparametric adaptive binarization method for poorly degraded document images. Our system firstly estimates the contrast image through a grayscale morphological operator. Secondly, two threshold values are generated by our Shannon entropy-based ternarization, based on either a 1-D histogram or a 2-D histogram, to maximize the entropy of the contrast image. The double-threshold value classifies the pixels into three regions corresponding to text, near-text, and non-text. Thirdly, the mean and standard deviation-based local thresholding technique is applied on the near-text areas, to best choose the document's text pixels. Finally, the binarized image is post-processed by two noise removal approaches, corresponding to different types of noise. In addition to binary morphological operators, other strategies such as shrink and swell filters, and a graph searching strategy are used to get rid of undesired regions and bridge the gaps, as well as to enhance the quality of the image. Experiments show that our proposed method outperforms other state-of-

the-art binarization methods for degraded document images when tested on four datasets: DIBCO 2009 [6], H-DIBCO 2010 [12], ICHFR 2010 [34], and Carlos et al.'s database [2]. The performance is evaluated by the nine distinct measures proposed in [6, 12, 21].

The rest of this paper is arranged as follows. We briefly review the basic morphological operators for grayscale and binary images in Section 2. The Shannon entropy-based thresholding technique, based on a 1-D histogram and a 2-D histogram, is described in Section 3. Section 4 presents the proposed degraded document binarization scheme in detail. Experiments along with comparisons are described and discussed in Section 5. Concluding remarks are provided in Section 6.

II. MORPHOLOGICAL OPERATORS

Four basic morphological operators, namely, dilation, erosion, opening, and closing, on both binary and gray-level images, are introduced in following sub-sections.

A. Binary morphology

Let *f* and *se* (a small window) be functions representing an $M \times N$ image, and a $(2w+1)\times(2w+1)$ structuring element, respectively, where *w* is the width of the character stroke. The four basic binary morphological operators [26, 27] are described as follows:

Dilation: The dilation of *f* by the structuring element *se* is given by Eq. 3 below.

$$g(x,y) = f \oplus se = \bigcup_{i=-M}^{M} \bigcup_{j=-N}^{N} se(i,j) \bigcap f(x-i, y-j) . \quad (3)$$

That is, if just one of the '1's in the structuring element *se* match with the input signal, then the output is '1'. This operator is very useful to connect gaps in an image as well as to expand an image. In binary images, the dilation operator gradually enlarges the boundaries of the foreground regions, typically from characters. Thus, the areas in the foreground grow in size, whereas the holes within those regions become smaller.

Erosion: The erosion of *f* by the same flat structuring element *se* is defined as in Eq. 4 below

$$g(x,y) = f \ominus se = \bigcap_{i=-M}^{M} \bigcap_{j=-N}^{N} se(i,j) \bigcup f(x+i, y+j) . \quad (4)$$

That is, if all of the '1's in the structuring element *se* match with the input signal, then the output is '1'. This operator clears away the boundaries of the foreground regions and thus, the foreground regions shrink in size. That means the holes within those areas become bigger. In binary images, the erosion operator gets rid of irrelevant details from the images. Figs.5 (b) and (c) show visual examples of dilation and erosion operators on the original binary image given in Fig.5(a).

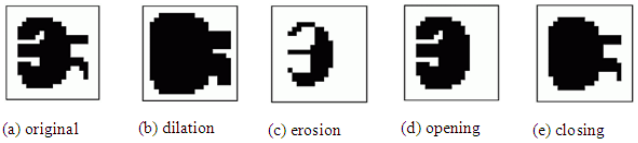


Figure 5. Binary morphological operators

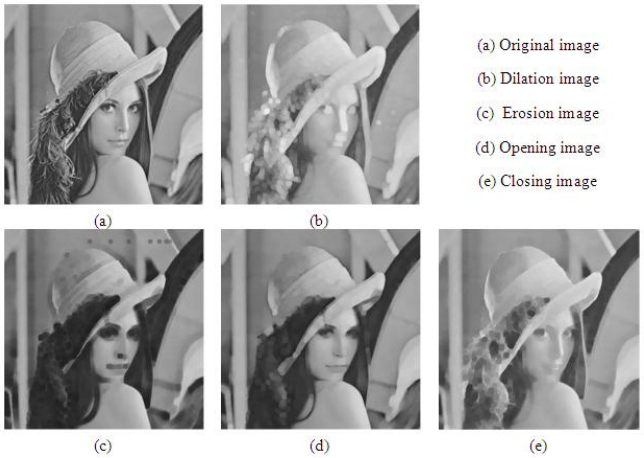


Figure 6. Illustration of grayscale morphological operators

By iteratively applying the dilation and erosion operators, one can eliminate details of the image, which are smaller than the structuring element *se*, without affecting its global geometric features. The morphological opening and closing operators are determined as follows:

Opening: is defined as $g = f \circ se = (f \ominus se) \oplus se$. This operator smoothes the original signal by eliminating peaks that are smaller than the size of the structuring element *se*. Fig. 5 (d) illustrates an example of the binary morphological opening operator. In image processing, this operator removes caps, spurs, and islands that are smaller than the structuring element *se*.

Closing: is defined as $g = f \bullet se = (f \oplus se) \ominus$ se . This operator smoothes the original signal by filling its gaps that are narrower than the given structuring element *se*. Fig. 5 (e) illustrates an example of the binary morphological closing operator. In image processing, this operator fills gulfs, channels, and lakes that are smaller than the structuring element *se*.

B. Grayscale morphology

In order to extend morphological concepts to grayscale images, gray-level extremum operators are used. The simple approach applies the maximum or minimum operators over a small neighbor, which is defined by the structuring element *se*.

Dilation: is defined as a *convolution-like* operation. A typical small structuring element is scanned over an image and, at each position, the maximums of point-by-point sums of the image and the structuring element are computed. If the original grayscale signal *f* and the structuring element *se* are sized as $M \times N$, and $(2w+1)\times(2w+1)$, then the dilation image *g* is defined as in Eq. 5.

$$g(x,y) = f \oplus se = \underset{i,j}{Max}[f(x-i, y-j) + se(i,j)] , \quad (5)$$

where $i, j \in [-w, w]$, $x \in [0, M-1]$ and $y \in [0, N-1]$. This operator increases the brightness of pixels which are surrounded by a higher intensity. Obviously, the neighborhood is determined by the structuring element.

Erosion: is defined as a *correlation-like* operation where the minimum operator is used and defined as in Eq. 6 below

$$g(x,y) = f \Theta se = \min_{i,j}[f(x+i, y+j) + se(i,j)] \quad (6)$$

where $i, j \in [-w, w]$, $x \in [0, M-1]$ and $y \in [0, N-1]$. This operator reduces the brightness of pixels which are surrounded by lower intensity. The grayscale erosion has the opposite effect of dilation because the resulting eroded image tends to be darker, whereas the other tends to be brighter. That means the dilation function expands the white regions of an image while the erosion function tends to shrink the white regions of an image.

Opening: is defined as $g = f \circ se = (f \Theta se) \oplus se$. That means the opening function consists of an erosion function followed by a dilation one. Because bright boundaries reduced by the erosion are restored by the dilation, this operation does not significantly alter the area and shape of regions. In general, small bright regions that vanish during the erosion do not reappear after the dilation; thus, this operator removes the bright isolated spots in the dark regions and it smoothes the boundaries.

Closing: is defined as $g = f \bullet se = (f \oplus se) \Theta se$. That means the closing function consists of a dilation function followed by an erosion one. Because bright boundaries expanded by the dilation are reduced by the erosion, this operation does not significantly alter the area and shape of regions. In general, small dark regions that vanish during the dilation do not reappear after the erosion; hence, this operator removes the dark isolated spots in the bright regions and it smoothes the boundaries. Fig. 6 shows the gray-level morphological operations, in which the original image is given in Fig. 6 (a).

III. SHANNON ENTROPY-BASED THRESHOLDING

In Information Theory, it is assumed that there are L possible symbols which occur with probability p_i, where i=0, 1, …, L-1. Denote that the entropy can be measured in bits/symbol, the Shannon entropy [10] associated with the source S is defined as in Eq. 7, below.

$$H(S) = -\sum_{i=0}^{L-1} p_i \log(p_i) . \quad (7)$$

In the entropy-based thresholding systems, the input image is considered as a source and the entropy is a summation of H_b and H_w, corresponding to the entropy of black regions (foreground) and the entropy of white ones (background), respectively. H_b and H_w, bounded by the threshold value T, are defined by maximizing the entropy $H^T(S) = H_b + H_w = -\sum_{i=0}^{T} p_i \log(p_i) - \sum_{i=T+1}^{L-1} p_i \log(p_i)$. However, when the entropy is calculated over the input image's entire distribution, then obviously it does not depend upon the threshold value T.

The 1-D Shannon–based image thresholding method was firstly proposed Pun [29] and improved by Kapur [17]. The entropic method was further developed by Renyi [25] and Tsallis [31, 32]. Pun considers the gray-level histogram as an L-symbol source, where all the symbols are statistically independent. The cumulative probability function is defined as $P(T) = \sum_{i=0}^{T} p_i$. The source entropy H is defined as in Eq.8.

$$H = -\sum_{i=0}^{T} p_i \log(p_i) - \sum_{i=T+1}^{L-1} p_i \log(p_i) . \quad (8)$$

The ratio of the a-posteriori entropy $H'(T) = -P(T)\log(P(T)) - [1-P(T)]\log[(1-P(T))]$ and H is lower bounded by

$$\frac{H'(T)}{H} \geq [\frac{\alpha \log P(T)}{\log(\max(p_0, ..., p_T))} + \frac{(1-\alpha)\log(1-P(T))}{\log(\max(p_{T+1}, ...p(L-1)))}] ,$$

where α is the anisotropy parameter and is defined depending on the histogram asymmetry and optimal threshold value. The optimal threshold in Pun's proposed scheme is calculated by solving $T_{opt} = \arg\underset{T}{equal}\,(H_w(T) = \alpha H)$.

In Kapur's method, the black and white regions are considered as two different sources. The source entropy $H(T)$ is a summation of the black entropy $H_b(T)$ and white region entropy $H_w(T)$, where $H_b(T)$ and $H_w(T)$ are defined as in Eqs. 9 and 10.

$$H_b(T) = -\sum_{i=0}^{T} \frac{p_i}{P(T)} \log \frac{p_i}{P(T)} . \quad (9)$$

$$H_w(T) = -\sum_{i=T+1}^{L-1} \frac{p_i}{1-P(T)} \log \frac{p_i}{1-P(T)} . \quad (10)$$

Renyi's entropy is parameterized according to the value of α, where α≥0 and α≠1, and is given by Eq. 11.

$$H(S) = \frac{1}{1-\alpha} \sum_{i=0}^{L-1} p_i^{\alpha} . \quad (11)$$

Tsallis defined his entropy based on two parameters, k and α, as: $H(S) = k\left[\frac{1}{\alpha-1} - \sum_{i=0}^{L-1} \frac{p_i^{\alpha}}{\alpha-1}\right]$. It is easily to prove that Renyi's entropy and Tsallis's entropy tend to resemble the Shannon entropy when α→1, and k = 1 using l'H^opital rule. Depending on different parameter value(s), the algorithms of Renyi and Tsallis provide flexible thresholds. Thus, to make the entropic thresholding scheme more effective, the entropic methods use a training image database. Firstly, the training images are divided into groups, based on statistical information. Secondly, the parameters are chosen for each group. To illustrate Shannon entropy-based thresholding, we take Kapur's entropy as an example, where the thresholding procedure is expressed by **Algorithm 1**.

Algorithm 1: ***1-D histogram Shannon entropy-based thresholding***

Input: A grayscale image
Output: A threshold value T
Procedure:
MAXIMUM=0; T=0;
for each t =0, $t < L$, t ++

$P(t) = \sum_{i=0}^{t} p_i$;

$H(S) = H_b + H_w = -\sum_{i=0}^{t} \frac{p_i}{P(t)} \log\left(\frac{p_i}{P(t)}\right) - \sum_{i=t+1}^{L-1} \frac{p_i}{1-P(t)} \log\left(\frac{p_i}{1-P(t)}\right)$;

if $H(S)$ > MAXIMUM then
MAXIMUM=$H(S)$; T=t;
end if \ end for

Obviously, the 1-D histogram Shannon entropy-based thresholding seems promising, but it does not take the image's spatial correlation into account. It means that two different images with an identical histogram will result in the same threshold value. In order to solve this problem, Abutaleb [4] used the 2-D histogram, which is formed by the original 1-D histogram and its local average.

For an $M \times N$ image with L gray levels (L=256 for a grayscale image), a co-occurrence matrix of an image is an $L \times L$ square matrix. For each image pixel at spatial co-ordinate (m, n) with its gray level specified by $I(m, n)$, it considers two neighbouring pixels at locations of (m+1, n), (m, n+1). Let O_{ij} be the $(i, j)^{\text{th}}$ element of the co-occurrence matrix, $O_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{N} X_{mn}$, where $X_{mn} = 1$ if $I(m,n) = i$ and $I(m+1, n) = j$ and $I(m, n+1) = j$. From the gray level i to the gray level j, the desired transition probability is $p_{ij} = \frac{O_{ij}}{\sum_{l_1=1}^{L} \sum_{l_2=1}^{L} O_{l_1 l_2}}$. Let T be a threshold value used to binarize an image. The threshold partitions a co-occurrence matrix into four quadrants which have a transition entropy called H_{BB}, H_{BF}, H_{FB} and H_{FF}, respectively, and these are defined as follows:

$$H_{BB}(T) = -\sum_{i=0}^{T} \sum_{j=0}^{T} p_{ij|BB} \log(p_{ij|BB}) \text{ and } p_{ij|BB} = \frac{p_{ij}}{\sum_{u=0}^{T} \sum_{v=0}^{T} p_{uv}}$$

$$H_{BF}(T) = -\sum_{i=0}^{T} \sum_{j=T+1}^{L-1} p_{ij|BF} \log(p_{ij|BF}) \text{ and } p_{ij|BF} = \frac{p_{ij}}{\sum_{u=0}^{T} \sum_{v=T+1}^{L-1} p_{uv}}$$

$$H_{FB}(T) = -\sum_{i=T+1}^{L-1} \sum_{j=0}^{T} p_{ij|FB} \log(p_{ij|FB}) \text{ and } p_{ij|FB} = \frac{p_{ij}}{\sum_{u=T+1}^{L-1} \sum_{v=0}^{T} p_{uv}}$$

$$H_{FF}(T) = -\sum_{i=T+1}^{L-1} \sum_{j=T+1}^{L-1} p_{ij|FF} \log(p_{ij|FF}) \text{ and } p_{ij|FF} = \frac{p_{ij}}{\sum_{u=T+1}^{L-1} \sum_{v=T+1}^{L-1} p_{uv}}$$

The 2-D histogram Shannon entropy-based thresholding is divided into three groups: Global Entropy (*GE*), Joint Entropy (*JE*), and Local Entropy (*LE*). Chang et al [9] and Pal et al [23] defined H_{GE}, H_{JE} and H_{LE} as below:

$H_{GE}(T) = H_{BB}(T) + H_{FF}(T) + H_{BF}(T) + H_{FB}(T)$

$H_{JE}(T) = H_{BF}(T) + H_{FB}(T)$

$H_{LE}(T) = H_{BB}(T) + H_{FF}(T)$

The optimal threshold values corresponding to GE, JE, and LE are derived by maximizing Shannon's entropy and are specified as follows.

$T_{GE} = \arg\{\max_{t \in [0,1,\ldots L-1]} H_{GE}(t)\}$;

$T_{JE} = \arg\{\max_{t \in [0,1,\ldots L-1]} H_{JE}(t)\}$;

$T_{LE} = \arg\{\max_{t \in [0,1,\ldots L-1]} H_{LE}(t)\}$.

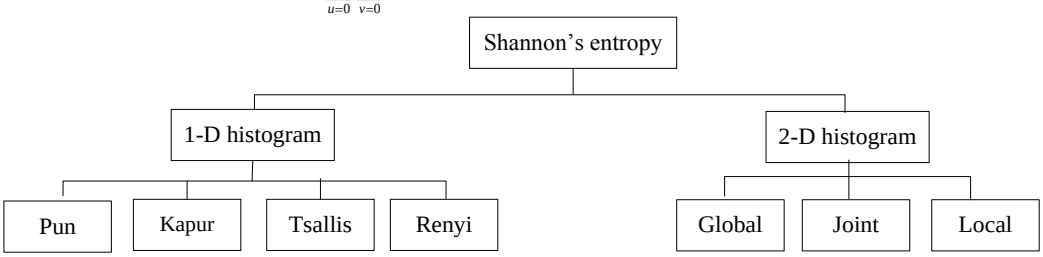


Figure 7. Shannon entropy-based image thresholding categories

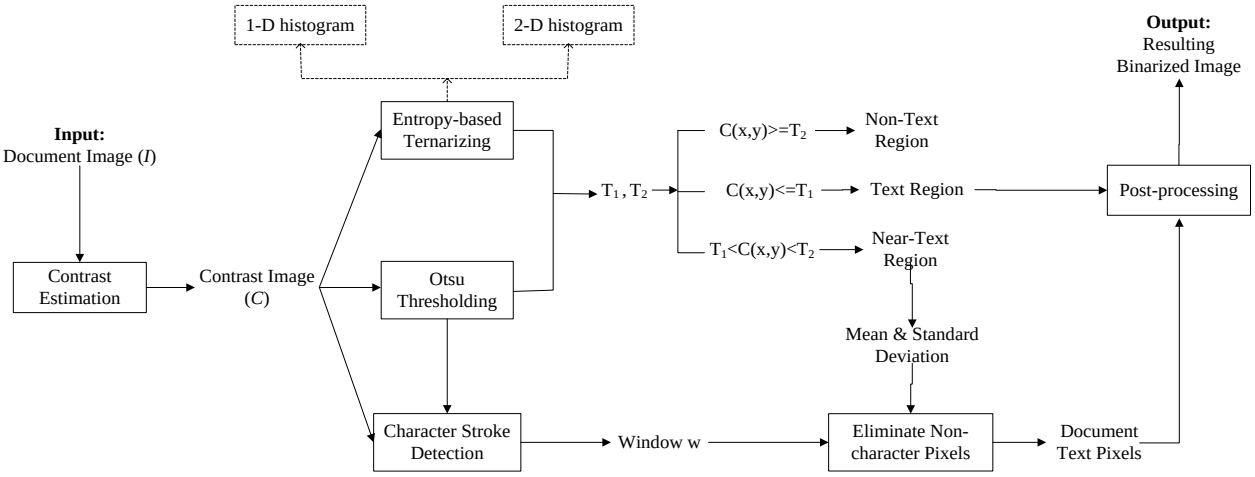
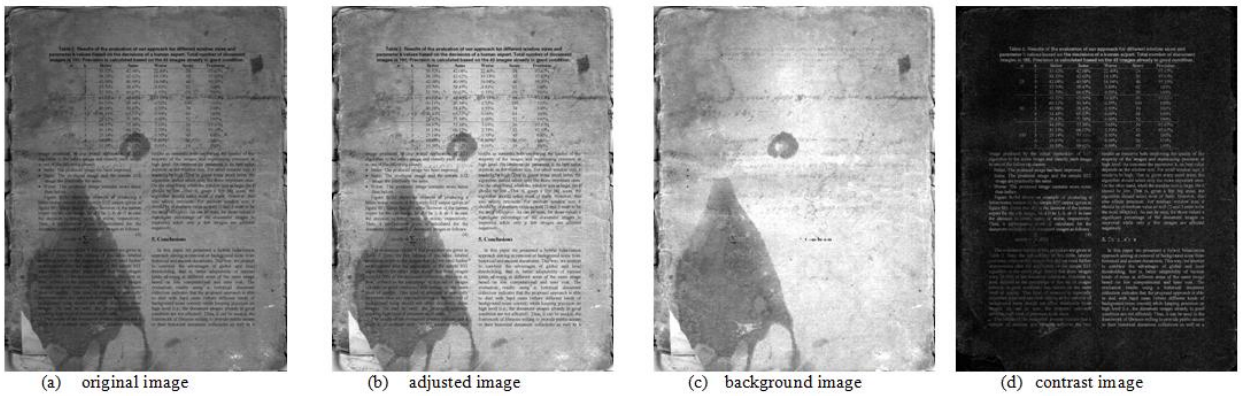


Figure 8. Flowchart of our proposed scheme



(a) original image (b) adjusted image (c) background image (d) contrast image

Figure 9. An example of contrast image estimation

In general, the Shannon entropy-based image thresholding methods can be categorized and are expressed in Fig. 7.

IV. PROPOSED SCHEME

This section presents the proposed binarization scheme using with the Shannon entropy-based ternarization, under the 1-D histogram and the 2-D histogram entropic thresholding. In particular, we divide this section into four subsections, dealing with contrast image estimation, character stroke width detection, double threshold-based binarization, and post-processing, respectively. The flowchart of our system can be seen in Fig. 8.

A. Contrast Estimation

Based on the gray-level morphological operators in Section 2.2, the morphological closing function is used to extract the contrast intensity. To ensure that the contrast image contains all pixels that lie within the text area, the pixel intensity values in the document image are adjusted. Simply, all gray-levels are mapped to a new range from 0 to 255. Fig. 9 (a) shows an unevenly illuminated, bleeding-through document image and its adjusted one is presented in Fig. 9 (b). Fig. 9 (c) illustrates the background image extracted by the gray-level morphological closing operation. By subtracting the original image from the background one, the contrast image is generated, as in Fig. 9 (d).

B. Character Stroke Width Detection

We detect the character stroke width based on the lengths and the frequencies of connected components. For the elimination of noisy areas and smoothing of the background area, the input image is first filtered by the mean and Gaussian filters. In addition to edge-preserving, the mean filter is particularly useful in reducing noise, e.g., salt and pepper noise. To obtain a smoother result, the image is convolved with a Gaussian function. To roughly estimate the foreground regions, Otsu's global thresholding is used to binarize the filtered image. Connected components are computed in every row. Let f_i be the frequency of occurrence of i^{th} connected components with length l_i. Character stroke width is approximated by averaging the lengths of the connected components, $w = \frac{\sum_{i=1} f_i \times l_i}{\sum_i f_i}$. Obviously, the stroke width is affected by the maximum of f_i.

C. Double Threshold-based Binarization

This section describes our methods of generating the binarized image using the ternarization thresholding technique in combination with the local mean and the standard deviation. Let I and C be the input degraded document image and its contrast image, respectively, as described in Section 4.1. The binarizing procedure of our proposed scheme is illustrated in two phases. The first phase classifies an input image into three regions corresponding to text, near-text, and non-text regions. In this phase, two thresholds, T_1 and T_2, are generated by our Shannon entropy-based ternarization technique, corresponding to either the 1-D histogram-based or the 2-D histogram-based. To relabel the pixels in the second (near-text) region, a comparison is made between the two threshold values in Phase 1, and the local mean and standard deviation strategy is applied in Phase 2. The two phases are described below.

Phase 1: Shannon entropy-based ternarization

In our 1-D histogram Shannon entropy-based ternarization, two threshold values are selected to maximize the entropy $H(S) = H_b + H_{bw} + H_w$. In this system, H_b, H_{bw}, and H_w are equivalent to the entropy of text, near-text, and non-text regions, respectively, and are defined as:

$$H_b = \frac{-1}{\sum_{i=0}^{t_1} p_i} \sum_{i=0}^{t_1} p_i \left(\log(p_i) - \log(\sum_{i=0}^{t_1} p_i) \right);$$

$$H_{bw} = \frac{-1}{\sum_{i=t_1+1}^{t_2} p_i} \sum_{i=t_1+1}^{t_2} p_i \left(\log(p_i) - \log(\sum_{i=t_1+1}^{t_2} p_i) \right);$$

$$H_w = \frac{-1}{\sum_{i=t_2+1}^{L-1} p_i} \sum_{i=t_2+1}^{L-1} p_i \left(\log(p_i) - \log(\sum_{i=t_2+1}^{L-1} p_i) \right).$$

The 1-D histogram Shannon entropy-based ternarization procedure has been constructed, as in **Algorithm 2.**

Algorithm 2 *Ternarization system using 1-D histogram Shannon based entropy*

Input: A grayscale image
Output: Two threshold values, T_1, and T_2
Procedure:
MAXIMUM_K = 0; T_1 = 0; T_2 = 0;
for each $t_1 = 0$, $t_1 < L\text{-}1$, t_1++
for each $t_2 = t_1+1$, $t_2 \le L\text{-}1$, t_2++

$$P_b = \sum_{i=0}^{t_1} p_i \;;\; P_{bw} = \sum_{i=t_1+1}^{t_2} p_i \;;\; P_w = \sum_{i=t_2+1}^{L-1} p_i$$

$$H(S) = -\sum_{i=0}^{t_1} \frac{p_i}{P_b}\log(\frac{p_i}{P_b}) - \sum_{i=t_1+1}^{t_2} \frac{p_i}{P_{bw}}\log(\frac{p_i}{P_{bw}}) - \sum_{i=t_2+1}^{L-1} \frac{p_i}{P_w}\log(\frac{p_i}{P_w});$$

if $H(S)$ > MAXIMUM_K then
MAXIMUM_K = $H(S)$; $T_1 = t_1$; $T_2 = t_2$;
end if \ end for \ end for

As for the 2-D histogram Shannon entropy-based ternarization, two thresholding values, T_1 and T_2, partition the co-occurrence matrix into nine equivalent parts which is shown in Fig. 10, as follows:

	0	T_1	T_2	L
T_1	A	AB	B	
T_2	AC	ABCD	BD	
L	C	CD	D	

Figure 10. Nine equivalent parts of a co-occurrence matrix

The transition entropy of each part is defined in Table 1.

The 2-D histogram Shannon-based entropy can be either global, joint or local. Take the global one for example; it is expressed in **Algorithm 3.**

Algorithm 3 *Ternarization system using 2-D histogram Shannon based entropy*

Input: A grayscale image
Output: Two threshold values, T_1, and T_2
Procedure:
MAXIMUM_G=0; T_1=0; T_2=0;
for each $t_1 = 0$, $t_1 < L\text{-}1$, t_1++
for each $t_2 = t_1+1$, $t_2 \le L\text{-}1$, t_2++

$$H_A = -\sum_{i=0}^{t_1}\sum_{j=0}^{t_1} p_{ij|A}\log(p_{ij|A}) \quad H_{AB} = -\sum_{i=0}^{t_1}\sum_{j=t_1+1}^{t_2} p_{ij|AB}\log(p_{ij|AB})$$

$$H_B = -\sum_{i=0}^{t_1}\sum_{j=t_1+1}^{L-1} p_{ij|B}\log(p_{ij|B}) \quad H_{BD} = -\sum_{i=t_1+1}^{t_2}\sum_{j=t_1+1}^{L-1} p_{ij|BD}\log(p_{ij|BD})$$

$$H_C = -\sum_{i=t_1+1}^{L-1}\sum_{j=0}^{t_1} p_{ij|C}\log(p_{ij|C}) \quad H_{AC} = -\sum_{i=t_1+1}^{t_2}\sum_{j=0}^{t_1} p_{ij|AC}\log(p_{ij|AC})$$

$$H_D = -\sum_{i=t_1+1}^{L-1}\sum_{j=t_1+1}^{L-1} p_{ij|D}\log(p_{ij|D}) \quad H_{CD} = -\sum_{i=t_1+1}^{L-1}\sum_{j=t_1+1}^{t_2} p_{ij|CD}\log(p_{ij|CD})$$

$$H_{ABCD} = -\sum_{i=t_1+1}^{t_2}\sum_{j=t_1+1}^{t_2} p_{ij|ABCD}\log(p_{ij|ABCD})$$

$H(S) = H_A + H_{AB} + H_B + H_{AD} + H_{ABCD} + H_{BD} + H_C + H_{CD} + H_D$
if $H(S)$> MAXIMUM_G then
MAXIMUM_G=$H(S)$; $T_1=t_1$; $T_2=t_2$;
end if \ end for \ end for

In the Shannon entropy based 2-D histogram, let T_1 and T_2 be the two threshold values, which are adjusted by Otsu's threshold T_0 as follows:

$$T_1 = \frac{T_1 + \max\left((T_1+T_0)/2, T_0\right)}{2}$$

$$T_2 = \frac{T_2 + \min\left((T_2+T_0)/2, T_0\right)}{2},$$

The classification is defined by the following rules:

$$C(x,y) \in \begin{cases} H_b & if \quad C(x,y) \le T_1 \\ H_{bw} & if \quad T_1 < C(x,y) < T_2 \\ H_w & if \quad C(x,y) \ge T_2 \end{cases}$$

$$I(x,y) = \begin{cases} 1 & if \quad C(x,y) \in H_b \\ 0 & if \quad C(x,y) \in H_w \end{cases}.$$

The experimental results have shown that in document binarization applications, GE outperforms JE and LE. Thus, GE is applied to generate our 2-D histogram Shannon-based entropy.

TABLE I. THE TRANSITION ENTROPY OF EACH PART IN THE CO-OCCURRENCE MATRIX

$H_A = -\sum_{i=0}^{T_1}\sum_{j=0}^{T_1} p_{ij|A}\log(p_{ij|A}),\ p_{ij|A} = \frac{p_{ij}}{\sum_{i=0}^{T_1}\sum_{j=0}^{T_1} p_{ij}}$	$H_{AB} = -\sum_{i=0}^{T_1}\sum_{j=T_1+1}^{T_2} p_{ij|AB}\log(p_{ij|AB}),\ p_{ij|AB} = \frac{p_{ij}}{\sum_{i=0}^{T_1}\sum_{j=T_1+1}^{T_2} p_{ij}}$	$H_B = -\sum_{i=0}^{T_1}\sum_{j=T_2+1}^{L-1} p_{ij|B}\log(p_{ij|B}),\ p_{ij|B} = \frac{p_{ij}}{\sum_{i=0}^{T_1}\sum_{j=T_2+1}^{L-1} p_{ij}}$
$H_{AC} = -\sum_{i=T_1+1}^{T_2}\sum_{j=0}^{T_1} p_{ij|AC}\log(p_{ij|AC}),\ p_{ij|AC} = \frac{p_{ij}}{\sum_{i=T_1+1}^{T_2}\sum_{j=0}^{T_1} p_{ij}}$	$H_{ABCD} = -\sum_{i=T_1+1}^{T_2}\sum_{j=T_1+1}^{T_2} p_{ij|ABCD}\log(p_{ij|ABCD}),\ p_{ij|ABCD} = \frac{p_{ij}}{\sum_{i=T_1+1}^{T_2}\sum_{j=T_1+1}^{T_2} p_{ij}}$	$H_{BD} = -\sum_{i=T_1+1}^{T_2}\sum_{j=T_2+1}^{L-1} p_{ij|BD}\log(p_{ij|BD}),\ p_{ij|BD} = \frac{p_{ij}}{\sum_{i=T_1+1}^{T_2}\sum_{j=T_2+1}^{L-1} p_{ij}}$
$H_C = -\sum_{i=T_2+1}^{L-1}\sum_{j=0}^{T_1} p_{ij|C}\log(p_{ij|C}),\ p_{ij|C} = \frac{p_{ij}}{\sum_{i=T_2+1}^{L-1}\sum_{j=0}^{T_1} p_{ij}}$	$H_{CD} = -\sum_{i=T_2+1}^{L-1}\sum_{j=T_1+1}^{T_2} p_{ij|CD}\log(p_{ij|CD}),\ p_{ij|CD} = \frac{p_{ij}}{\sum_{i=T_2+1}^{L-1}\sum_{j=T_1+1}^{T_2} p_{ij}}$	$H_D = -\sum_{i=T_2+1}^{L-1}\sum_{j=T_2+1}^{L-1} p_{ij|D}\log(p_{ij|D}),\ p_{ij|D} = \frac{p_{ij}}{\sum_{i=T_2+1}^{L-1}\sum_{j=T_2+1}^{L-1} p_{ij}}$

Phase 2: Relabel the misclassified regions

Let $M \times N$ and w be the size of the image and the character stroke width, respectively. The near-text pixels are relabeled by the function ***Relabel_Img***. In the $(2w+1) \times (2w+1)$ window, which is around the near-text pixel, *Num* is the number of high contrast pixels, namely, the pixel intensity is higher than threshold T_1. The mean (I_{Mean}) and the standard deviation (I_{Std}) of the detected high contrast pixels in the original image I are determined by the function ***Relabel_Reg***.

function *Relabel_Img*

Input: the original image I, the contrast image C
Output: Relabel the near-text pixel in the image I
for $i = 1 : M$
  for $j = 1 : N$
    if $(C(i,j) \geq T_1)$
      $[Num, I_{Mean}, I_{Std}] =$ ***Relabel_Reg***(I, C, i, j);
      if ($I(i,j) < \min (I_{Mean} + I_{Std}, T_2)$ and $(Num > 0)$ then
        $I(i,j) = 1$;
      else
        $I(i,j) = 0$;
      end if
      $i = i + w$;
      $j = j + w$
    end if \ end for \ end for

function $[Num, I_{Mean}, I_{Std}] =$ *Relabel_Reg*(C, i, j)

Consider the $(2w+1) \times (2w+1)$ window of the contrast image $C_w = C(i - w: i + w, j - w: j + w)$
Consider the $(2w+1) \times (2w+1)$ window of the original image $I_w = I(i - w: i + w, j - w: j + w)$
$Num = 0$; $I_{Mean} = 0$; $I_{Std} = 0$;
for $x = 1 : 2w+1$
  for $y = 1 : 2w+1$
    if $(C_w(x,y) \geq$ T1) then
      $C_w(x,y) = 1$;
      Num ++;
    else
      $C_w(x,y) = 0$;
    end if \ end for \ end for
  if $(Num > 0)$ then

$$I_{Mean} = \frac{I_w \times C_w}{Num} \text{ and } I_{Std} = \sqrt{\frac{C_w \times (I_w - I_{Mean})^2}{Num}}$$

  end if

D. Post-processing

The post-processing operations are applied on the binarized image in order to eliminate noise, fill the breaks, gaps or holes, and preserve stroke connectivity. Two kinds of noise are considered in our work. The first one is "salt and pepper" noise which is defined as (1) white regions that are smaller than the stroke width and that usually stay inside the character stroke and (2) black regions that are smaller than the stroke width and that settle around the document text or that scatter over the image. Fig. 11 (a) shows some examples of "salt and pepper" noise as defined in our scheme. Fig. 11 (b) illustrated some cases of another kind of noise, called "block" noise, which consists of the black regions with a window size bigger than the stroke width. According to the experiments, the "block" noise was often located along the border of the images from the ICHFR 2010 database [34]. These images contained most kinds of problems that could be met in old documents: presence of stains and strains, backgrounds with big variations and uneven illumination, ink seepage, etc. To deal with the two different kinds of noise, two strategies have been applied in two phases and are described as follows:

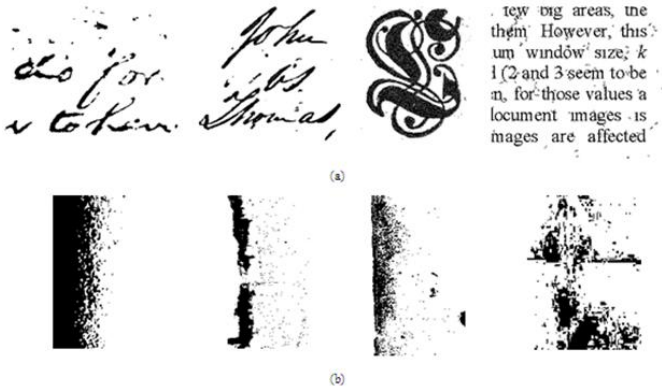


Figure 11. Representative examples of noise: (a) "salt and pepper" noise and (b) "block" noise

Strategy 1: To deal with the first kind of noise – "salt and pepper" noise.

Firstly, the binary morphological operators are applied to bridge the unconnected pixels and to remove the isolated or spurious pixels. Fig. 12 gives some representative examples of binary morphological operations in which the second row has the corrected version of the first row.

Secondly, we correct the areas that are larger than one pixel and smaller than the character stroke width by using shrink and swell filters. These filters can remove the region noise from the background, especially around the text, as well as fill the foreground region. Take an $R_1 \times R_2$ region as an example: Let *Num*_1 and *Num*_2 be the number of white

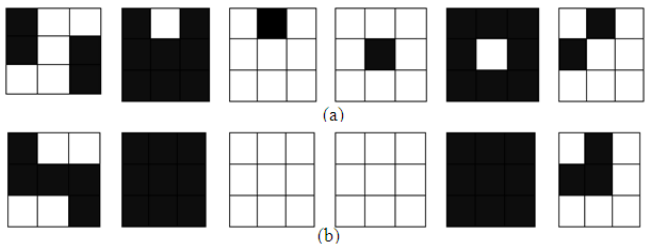


Figure 12. Examples of (a) unconnected, isolated, spurious pixels, (b) corrected ones by the binary morphological operators

pixels inside $(R_1 \times R_2)$ and $(R_1 - 1) \times (R_2 - 1)$ windows, respectively. Simply, if *Num*_1 equals *Num*_2, then window $(R_1 - 1) \times (R_2 - 1)$ becomes black. If *Num*_1 equals [*Num*_2+2($R_1 + R_2 - 2$)] then window $(R_1 - 1) \times (R_2 - 1)$ is filled in white.

Strategy 2: To deal with the second kind of noise – "block" noise.

As defined, the block noise consists of black regions that are bigger than the stroke width; therefore, a filter cannot work well because of the unwanted effects on document text areas. To deal with such kind of noise, a graph searching strategy is utilized. A graph (also called a tree) is defined as follows: (1) ***Root***: a complete black block, (2) ***Node*** (internal/external): a block in which the number of black pixels is larger than $2\times w$, and (3) ***Edge***: two adjacent black pixels that belong to two different and adjacent nodes. Each node belongs to only one tree (graph).

The graph searching strategy is performed as in **Algorithm 4**

Algorithm 4: ***The graph searching strategy***

Input: A binarized image

Output: A set of trees. Each tree is considered to be a block noise.

Procedure:

- The binarized image (as in Section 4.3) is divided into non-overlapping blocks with the window size bigger than the character stroke width w.
- Scan the image from left to right, top to bottom. For each block X:
  - If X is completely black and X does not belong to any tree, then
    - Set X as root of a new tree
    - For each node inside the tree, called A_i, consider its 4-neighbouring blocks, denoted by A_t, (for $t = 1, 2, .., 4$). If A_t is a node and there is an edge between A and A_t then add A_t to the tree X.

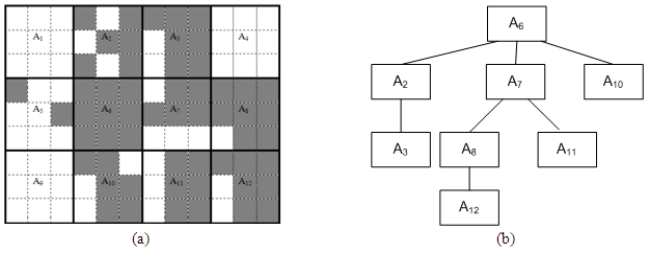


Figure 13. Examples of graph searching strategy where (a): binarized image and (b): tree

Take Fig. 13 as an example. Fig. 13 (a) presents the binarized image. The character stroke width is assumed to be $w = 2$. The image is divided into 12 non-overlapping 3x3 blocks, called A_i (for i=1, 2, .., 12). Scan the image from left to right, top to bottom, A_6 is the complete black block and does not belong to any tree. Set A_6 as a root of a tree.

For each node of the tree = $\{A_6\}$, consider the 4-neighbouring blocks of A_6, $\{A_2, A_7, A_{10}\}$ are nodes and A_5 is not a node by definition. Thus, the tree becomes $\{A_6, A_2, A_7, A_{10}\}$.

Next, consider the 4-neighbouring blocks of A_2, A_3 is a node and A_1 is not a node. Thus, the tree becomes: tree = $\{A_6, A_2, A_7, A_{10}, A_3\}$.

Similarly, we fill the tree with the root A_6. The tree is shown in Fig. 13 (b).

After the graph searching algorithm is applied to the image, all the trees (graphs), which have been detected, are removed from the image. In order to avoid the possibility of removing text, we have used the window size bigger two times the average stroke width.

V. Experimental Results

In this section, four testing datasets that have been used in the DIBCO 2009 [6], H-DIBCO 2010 [12], ICHFR 2010 [34] and Carlos et al.'s [2] databases are firstly introduced. The performance of our proposed schemes has been tested on the four databases. In addition, we also compare our methods with two other state-of-the-art document binarization methods including Tan et al.'s scheme [7, 8] and Carlos et al.'s scheme [2]. In order to make an entirely non-subjective comparison, nine measures [6, 12, 21] were used to do the evaluation.

A. Testing datasets

The performance of our scheme as well as the comparison will be evaluated qualitatively. We used four testing datasets consisting of historical documents, respectively obtained from DIBCO 2009 [6], H-DIBCO 2010 [12], and ICHFR 2010 [34] contests, and Carlos et al.'s database. The DIBCO 2009 testing database contains five handwritten document images and five printed document images that are associated with ground truth images for the evaluation and they were shown in Fig.1 The testing images of H-DIBCO 2010 were shown in Fig. 2 which includes ten handwritten document images, with the associated ground truth images from the collections of the Library of Congress [33]. The ICHFR 2010 contest is related to quantitative assessment of historical documents image binarization algorithms. Most kinds of problems are taken into account, such as presence of stains and strains, background of large variations and uneven illumination, ink seepage, etc. The ICHFR 2010 contest provides a database, consisting of 60 images for training and 210 images for testing. In the training set, there are 30 blended averaging intensity images and 30 blended maximum intensity ones corresponding to three ground truth images. In the training set, there are 105 blended averaging intensity images and 105 blended maximum intensity ones corresponding to seven ground truth images. In total, there are 10 ground truth images and 270 degraded document images in the database of the ICHFR 2010 contest. Representative samples of the ICHFR 2010 database were shown in Fig. 3. Carlos et al.'s database (Fig. 4) contains more than 30,000 images and documents from the end of the 19th century onwards.

B. Evaluation Measures

In this section, criteria that were used to measure and evaluate a binarization algorithm in the last two contests, namely, DIBCO 2009, H-DIBCO 2010, are firstly described. These measures consist of: (a) F-Measure (**FM**); (b) pseudo F-Measure (**p_FM**); (c) Peak Signal to Noise Ratio (**PSNR**); (d) Negative Rate Metric (**NRM**); and (e) Misclassification Penalty Metric (**MPM**). Moreover, to make our comparison and evaluation more general, we consider binarization as a bi-classification task, as detailed by Sokolova and Lapalme [21]. In their research, they present a systematic analysis of twenty-four performance measures, used in Machine

Learning classification tasks for binary, multi-class, multi-labelled, and hierarchical classes. Sokolova and Lapalme showed that the correctness of a classification can be evaluated by: (1) computing the number of correctly recognized class samples (true positives - TP), (2) calculating the number of correctly recognized samples that do not belong to the class (true negatives - TN), (3) and finding the samples that either are incorrectly assigned to the class (false positives - FP) or that are not recognized as class samples (false negatives - FN). To deal with the binary classification, in addition to the five above measures, four additional measures are considered: (f) Sensitivity (**Sens**), (g) Specificity (**Spec**), (h) Balanced Classification Rate (**BCR**), (i) $F_{ß}$- measure (**ß_FM**). These criteria, from (a) to (i), are briefly described as follows:

(a) F-measure (FM): evaluates how well an algorithm can retrieve the desired pixels and is defined as $\mathrm{FM} = \frac{2 \times precision \times recall}{precision + recall}$, where the precision, is also called a positive predictive value, and where the proportion of predicted positives is actually a positive value and is calculated as $precision = \frac{TP}{TP + FP}$, and the recall is the proportion of actual positives that are predicted to be positive and it is determined as $recall = \frac{TP}{TP + FN}$.

(b) pseudo F-Measure (p_FM): is based on the skeletonized ground truth image, in which the skeletonized ground truth image is defined as $SG(x,y) = \begin{cases} 0 & if \quad background \\ 1 & if \quad document\,text \end{cases}$. The pseudo F-Measure is calculated as $p\_FM = \frac{2 \times precision \times p\_recall}{precision + p\_recall}$, where p-recall is defined as the percentage of the skeletonized ground truth image, *SG*, that is detected in the resulting binary image *B*, and is defined as $p\_recall = \frac{1}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} SG(i,j)} \left(\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} SG(i,j) \times B(i,j) \right)$.

(c) Peak Signal to Noise Ratio (PSNR): is a measure of how close an *M*×*N* image (*I*) is to another (*I'*) and is defined as $PSNR = 10 \log\left(\frac{C^2}{MSE} \right)$, where *C* is a constant that denotes the difference between the foreground and the background and is set equal to 1 in our case study, and *MSE* is the mean square error, defined as $MSE = \frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} (I(i,j) - I'(i,j))^2}{MN}$.

(d) Negative Rate Metric (NRM): is also called the Balanced Error Rate, based on the pixel-wise mismatches between the ground truth image and the original one. It is defined as: $\mathrm{NRM} = \frac{1}{2}\left(FN/(TP + FN) + FP/(TN + FP) \right)$.

(e) Misclassification Penalty Metric (MPM): is a measure of how well the resulting image represents the contour of the ground truth image and is defined as $MPM = \frac{1}{2D}\left(\sum_{i=1}^{FN} d^{i}_{FN} + \sum_{j=1}^{FP} d^{j}_{FP} \right)$, where d^{i}_{FN} and d^{j}_{FP} respectively denote the distance of the i^{th} false negative and the j^{th} false positive pixel from the contour of the text in the ground truth image. The factor *D* is the sum of all the pixel-to contour distances of the ground truth object. A low MPM indicates that the algorithm is good at detecting the boundary.

(f) Sensitivity (Sen): proportion of actual positives which are predicted positive: $Sens = \frac{TP}{TP + FN}$.

(g) Specificity(Spec): proportion of actual negatives which are predicted negative: $Spec = \frac{TN}{TN + FP}$

(h) Balanced Classification Rate (BCR): gives balanced assessments on the two classes that have to be adopted, such as $BCR = (Sens + Spec)/2$.

(i) $F_{ß}$-Measure (ß_FM): weighted harmonic mean between sensitivity and specificity $\beta\_FM = \frac{2 \times Sens \times Spec}{Sens + Spec}$.

In contrast to FM, p_FM, ß_FM, PSNR, and BCR, the binarization quality is better for lower values of NRM, and MPM.

C. Results and Comparison

Evaluation of DIBCO 2009 and H-DIBCO 2010 showed that Tan et al.'s scheme [7, 8] outperformed 43 document thresholding algorithms that were submitted to DIBCO 2009, and outperformed 17 methods submitted to H-DIBCO 2010. The top results of Tan et al.'s algorithm were directly downloaded from [35] and have been evaluated using the nine distinct measures.

Taking the DIBCO 2009 testing database as an example, the first row in Fig. 14 shows the results of Tan et al.'s algorithm and our results are shown in the last two rows corresponding to our proposed method 1 (1-D histogram based) and our proposed method 1 (2-D histogram based).

Table II shows the detailed performance of each algorithm for eight measures on DIBCO 2009. Each algorithm is considered at three levels of quality: ***Best_FM*** corresponds to the image that obtains the greatest F-Measure; ***Worst_FM*** corresponds to the image that has the smallest F-Measure. ***Average_FM*** corresponds to the average quality over the whole DIBCO 2009 database and is calculated by averaging the measures of binarized images.

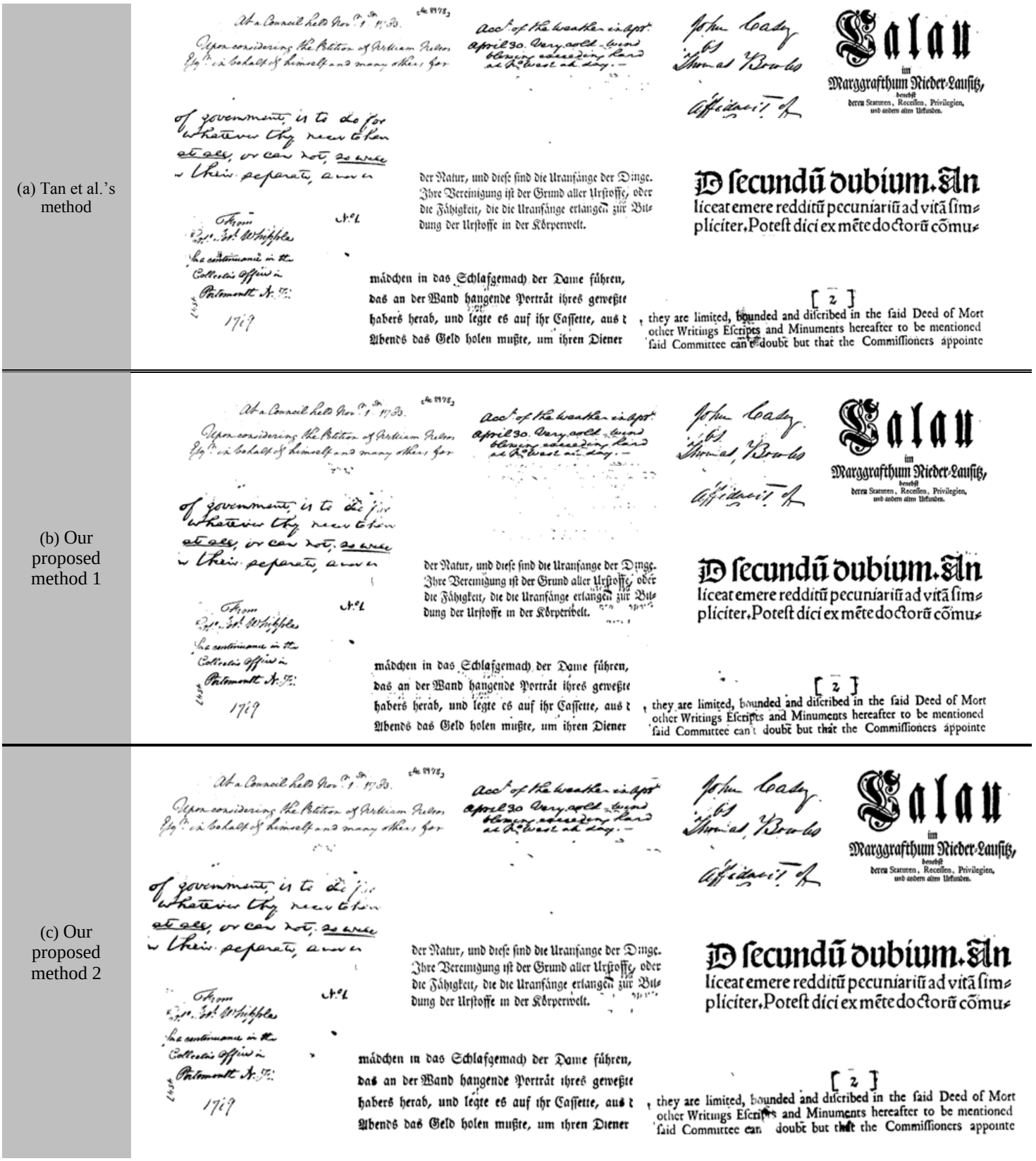
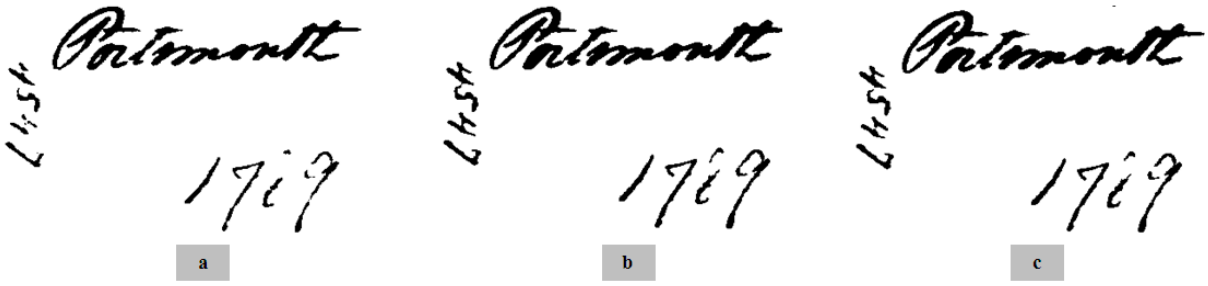


Figure 14. Resulting images of (a) Tan et al.'s algorithm and (b), (c) our proposed schemes

TABLE II. COMPARISON OF METHODS ON DIBCO 2009, BASED UPON EIGHT DISTINCT MEASURES

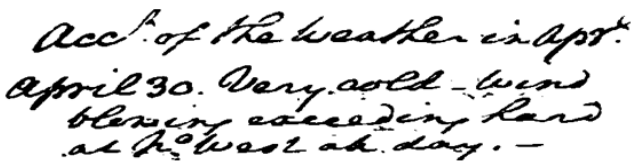
Algorithm		FM	PSNR	NRM ($\times 10^{-2}$)	MPM ($\times 10^{-3}$)	Sens	Spec	BCR	ß_FM
Tan et al's method	Best_FM	96.47303	19.306	3.119	0.317	0.93927	0.9992	0.9689	96.7916
	Worst_FM	84.9803	14.14	12.798	0.32	0.74561	0.9878	0.8720	85.3692
	Average_FM	91.10432	18.615	6.3616	0.51	0.87719	0.9956	0.9364	93.1668
Our proposed method 1	Best_FM	97.23910	20.3287	2.3228	0.094	0.95557	0.998	0.9768	97.6312
	Worst_FM	87.09253	17.424	8.698	0.353	0.8323	0.9937	0.913	90.589
	Average_FM	91.2494	18.6712	5.201	1.223	0.90376	0.9922	0.948	94.5482
Our proposed method 2	Best_FM	97.23558	20.3237	2.33	0.094	0.95541	0.998	0.9767	97.6236
	Worst_FM	86.87789	19.96	6.464	0.342	0.87631	0.9944	0.9354	93.1632
	Average_FM	91.32428	18.6662	4.7297	1.181	0.91353	0.9919	0.9527	95.0714

According to Table II, generally our proposed methods give better results than Tan et al.'s method. In the average (Average_FM) and in the best cases (Best_FM), the F-Measures of our methods vary from 0.15 to 0.8 higher than Tan et al.'s results. In the worst case (Worst_FM), the binarized images generated by our methods obtain much better result than Tan et al.'s with an F-Measure of 1.8 higher. The two proposed schemes correspond to the Shannon-based 1-D histogram and the Shannon-based 2-D histogram, and both give resulting images of equivalent quality, even though the latter obtains slightly better results. Figs. 15 - 18 further compare the binarization results with the original document images in DIBCO 2009 as in Fig.1. (a), (e), (h), (i).

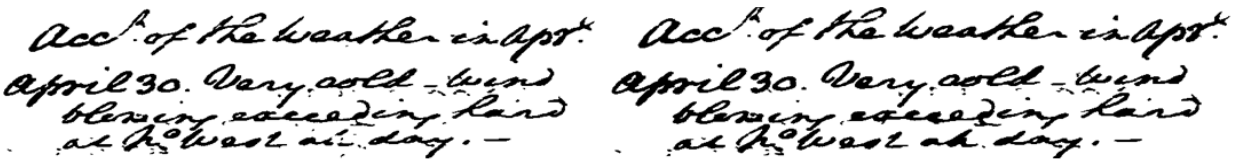


a b c

Figure 15. Binarization results of Fig 1 (a) by (a) Tan et al.'s method, (b) Our method 1, and (c) Our method 2



a



b c

Figure 16. Binarization results of Fig 1 (e) by (a) Tan et al.'s method, (b) Our method 1, and (c) Our method 2

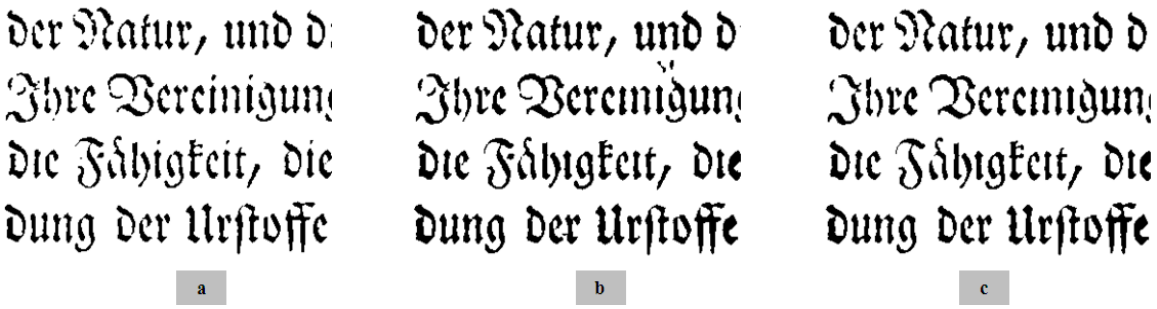


Figure 17. Binarization results of Fig 1 (i) by (a) Tan et al.'s method, (b) Our method 1, and (c) Our method

Figure 18. Binarization Results of Fig 1 (h) by (a) Tan et al.'s method, (b) Our method 1, and (c) Our method 2

Table III shows the evaluation of each algorithm on the H-DIBCO 2010 images. Tan et al.'s F-Measure, in the best case (Best_FM) and average case (Average_FM), is 1.8 smaller than ours. Notably, in the worst case (Worst_FM), their F-Measure is from 1.37 to 4.16 smaller than ours. As for our two proposed schemes, the first one (1-D histogram based) achieves better results in the worst case; however, they perform equally well in general.

Figs. 19 - 21 further compare the binarization results with the original document images in H-DIBCO 2010 as in Fig.2. (a), (c), (e).

TABLE III. EVALUATION OUR SCHEME ON THE H-DIBCO 2010 DATABASE

Algorithm		**FM**	**p_FM**	**PSNR**	**NRM ($\times 10^{-2}$)**	**MPM ($\times 10^{-3}$)**	**Sens**	**Spec**	**BCR**	**ß_FM**
Tan et al's method	Best_FM	91.6992	95.872	17.585	5.319	2.528	0.9013	0.9923	0.9468	94.463
	Worst_FM	78.6669	99.228	16.681	17.34	0.297	0.6537	0.9995	0.8266	79.044
	Average_FM	85.567	97.467	17.829	11.22	0.711	0.7786	0.9971	0.8878	87.218
Our proposed method 1	Best_FM	93.5041	97.201	18.664	4.46	0.812	0.916	0.9948	0.9554	95.376
	Worst_FM	82.824	98.431	17.407	13.9	0.646	0.7233	0.9985	0.8609	83.89
	Average_FM	87.8377	96.07	18.367	8.308	0.672	0.8387	0.9952	0.9169	89.497
Our proposed method 1	Best_FM	93.5658	97.336	18.714	4.5	0.736	0.9149	0.9951	0.9550	95.332
	Worst_FM	80.0398	98.35	16.874	16.07	3.07	0.68	0.9988	0.8394	80.908
	Average_FM	87.2137	96.63	18.298	9.262	0.805	0.8187	0.9961	0.9074	89.616

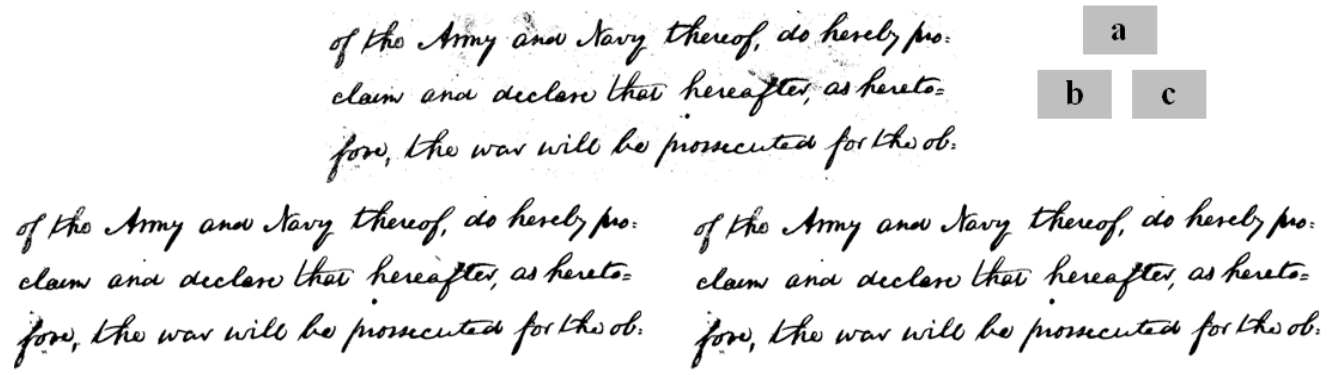


Figure 19. Binarization results of Fig 2 (a) by (a) Tan et al.'s method, ((b) Our method 1, and (c) Our method 2

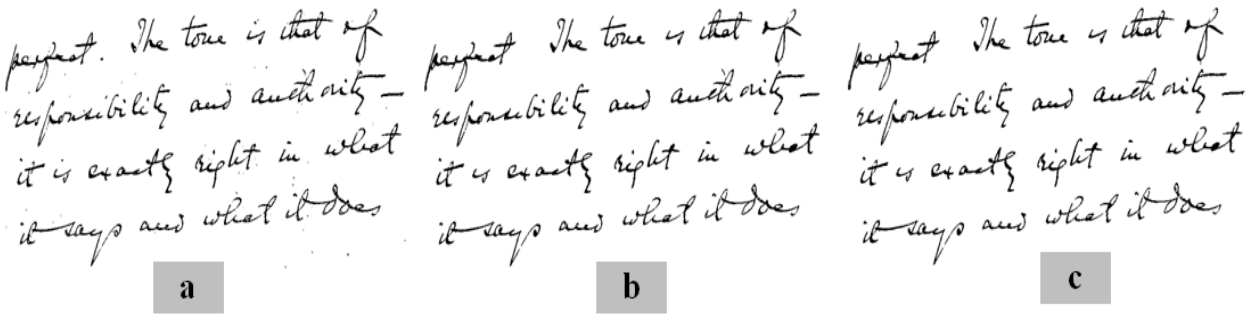


Figure 20. Binarization results of Fig 2 (e) by (a) Tan et al.'s method, (b) Our method 1, and (c) Our method 2

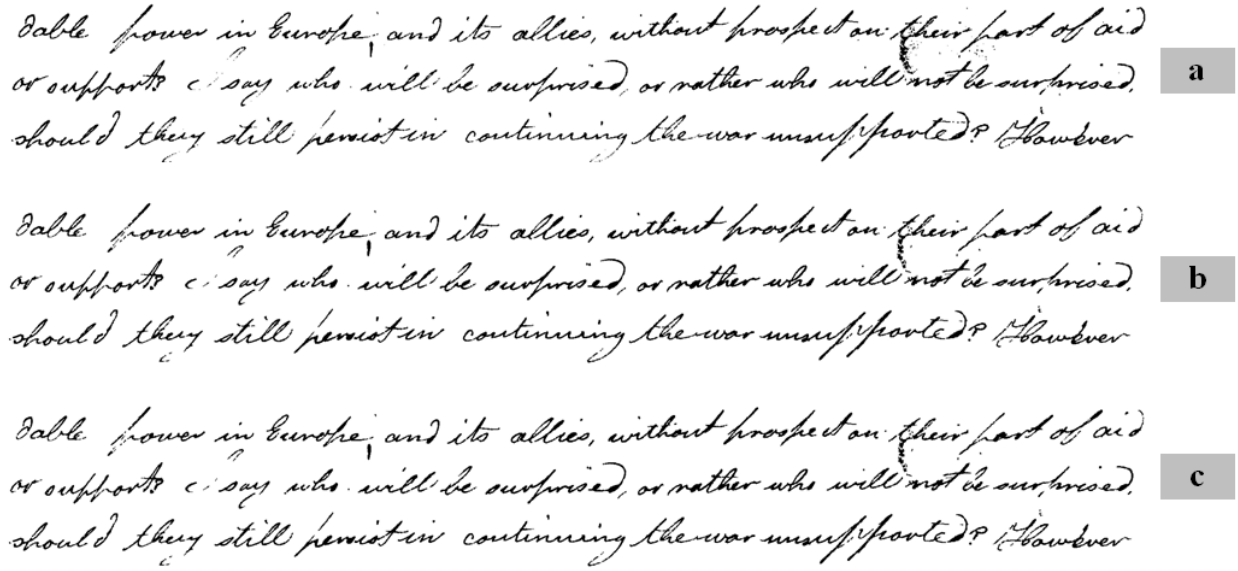


Figure 21. Binarization results of Fig 2 (c) by (a) Tan et al.'s method, (b) Our method 1, and (c) Our method 2

As the seven figures (Fig. 15 - 21) have shown, our proposed method extracts the text properly from the images that suffer from different types of degradation. The performance of our methods is either comparable or better when compared with the proposed method by Tan et al.

In comparison with the testing databases of DIBCO 2009 and H-DIBCO 2010, the database given by the ICFHR 2010 contest is more practical and difficult because of the various problems in the historical documents. The performances of Tan et al.'s algorithm and ours on the images in Fig. 3(a) from the ICFHR 2010 database are shown in Fig. 22 where the top row shows the whole page and the row below gives some parts of the top row's image in detail. The resulting binarized images from our method are illustrated in Figs. 22 (b) and (c) in comparison with that obtained by Tan et al in Fig. 22 (a).

The performances of Carlos et al.'s method, Tan et al.'s algorithm, and ours on the image in Fig. 4(e) from Carlos et

al.'s database are shown in Fig. 23. The text regions in Figs. 23 (a), (b) (c), and (d) are quite similar. However, the black border cannot be removed in Carlos et al.'s resulting image (Fig. 23 (a)). Tan et al.'s method (Fig. 23 (b)) takes away almost the black border except the border's edge. Figs. 23 (c), and (d) illustrate that our methods are able to clear the black border with better results.

The experiments show that our two proposed methods are quite comparable and they both achieve better results than Tan et al.'s method on all DBICO 2009, H-DIBCO 2010, ICFHR 2010 and Carlos et al.'s databases.

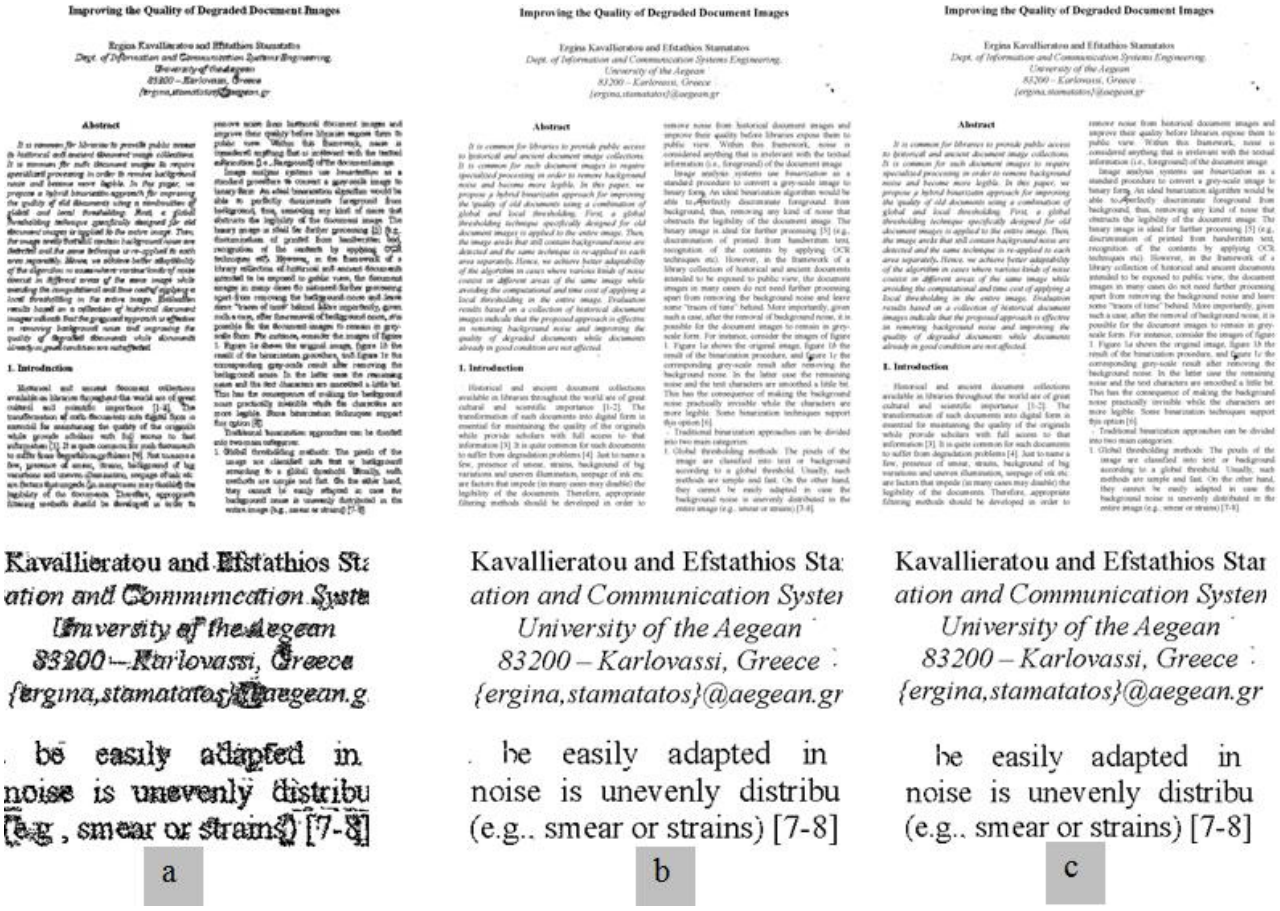


Figure 22. Binarization results Fig 3 (a) by (a) Tan et al.'s method, (b) Our method 1, and (c) Our method 2

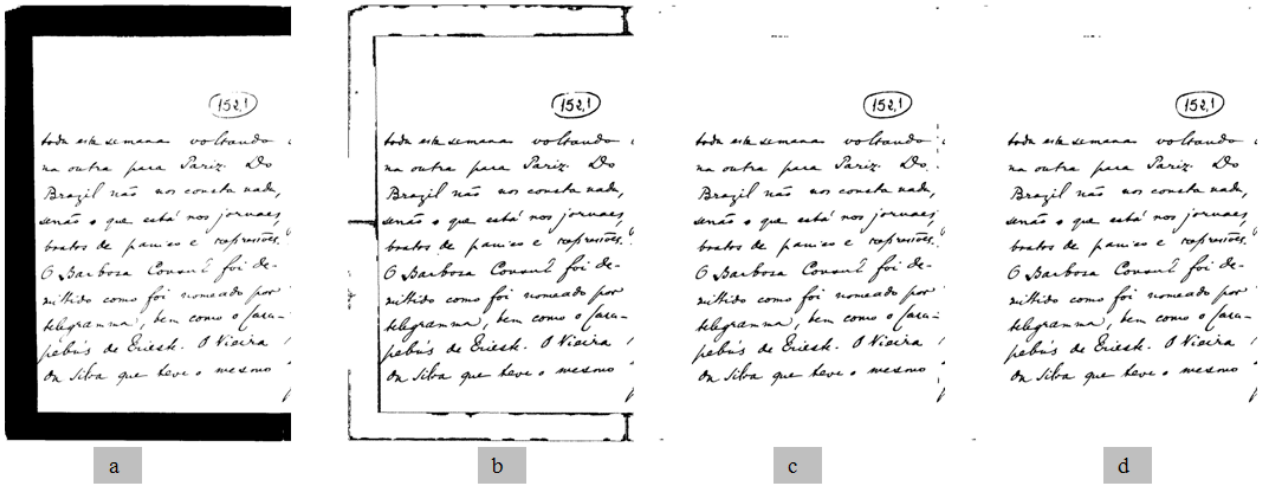


Figure 23. Binarization results of Fig 4 (e) by (a) Carlos et al.'s method, (b) Tan et al.'s method, Our method 1, and (d) Our method 2

VI. CONCLUSION

This paper introduces a novel document binarization technique that makes use of a contrast image, based on the morphological operators. In our proposed approach, the 1-D histogram and the 2-D histogram Shannon entropy-based ternarization methods are suggested to classify the content of an input image into three categories, namely, complete-text, near-text, and non-text. To relabel the near-text pixels, a local threshold is estimated by a summation of the mean and the standard deviation. The methods are compared with others on the databases of the three recent contests: DIBCO 2009, H-DIBCO 2010, ICFHR 2010 and PROHIST Project of Carlos et al. The performance is evaluated by the nine distinct measures. The experimental results show that both of our binarization methods are better than the top-ranked ones [7, 8], and [2]. Besides that, our binarization system performs well on Carlos et al.'s database.

ACKNOWLEDGEMENTS

This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and Concordia University. The authors would like to thank Carlos A. B. Mello and Adriano L. I. Oliveira for providing their database and their results. The authors are also grateful to C. L. Tan and his co-researchers for providing information on their algorithm. We would like to thank Ms. Shira Katz for editorial assistance.

REFERENCES

[1] A. Rosenfeld and P. D. L. Torre, "Histogram Concavity Analysis as an Aid in Threshold Selection," *IEEE Transactions on System Man Cybernetics*, Vol. 13, pp. 231-235, 1983.

[2] A. B. Mello Carlos, L. I. Oliveira Adriano, and S. Ángel, "Historical Document Image Binarization," *Proceeding International Conference on Computer Vision Theory and Applications*, Vol. 1, pp.108-113, Madeira, 2008.

[3] A. Al-Hussain and A. El-Zaart, "Moment-Preserving Thresholding Using Gamma Distribution," *Proceeding International Conference on Computer Engineering and Technology* (ICCET), Vol. 6, pp. 323-325, Chengdu, 2010.

[4] A. S. Abutaleb, "Automatic Thresholding of Grey-level Pictures Using Two-Dimensional Entropy," *Computer Graphics and Image Processing*, Vol. 47, pp. 22-32, 1989.

[5] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Adaptive Degraded Document Image Binarization," *Pattern Recognition*, Vol. 39, pp. 317-327, March 2006.

[6] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO 2009)," *Proceeding International Conference on Document Analysis and Recognition*, pp. 1375-1382, Barcelona, 2009.

[7] B. Su, S. Lu, and C. L. Tan, "Binarization of Historical Document Images Using The Local Maximum and Minimum," *Proceeding International Workshop on Document Analysis Systems*, pp. 159-165, Boston, 2010.

[8] B. Su, S. Lu, 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, October 2010.

[9] C. I. Chang, K. Chen, J. Wang, and M. L. G. Althouse, "A Relative Entropy-based Approach to Image Thresholding," *Pattern Recognition*, Vol. 27, pp. 1275-1289, 1994.

[10] C. Shannon, "A Mathematical Theory of Communication," *Bell System Technology Journal*, Vol. 27, pp. 370-423, 623-656, 1948.

[11] 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, pp. 141-150, October 2001.

[12] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "H-DIBCO 2010 - Handwritten Document Image Binarization Competition," *Proceeding International Conference on Frontiers in Handwriting Recognition*, pp. 727-732, Kolkata, 2010.

[13] I. P. Chen, M. Y. Lai, and L. L. Wang, "An Improvement in the Moment-Preserving Thresholding Method," *International Journal of Imaging Systems and Technology*, Vol. 18, pp. 365-370, July 2008.

[14] J. R. Parker, C. Jennings, and A. G. Salkauskas, "Thresholding Using an Illumination Model," *Proceeding International Conference on Document Analysis and Recognition*, pp. 270-273, 1993.

[15] J. Sauvola and M. Pietikainen, "Adaptive Document Image Binarization," *Pattern Recognition*, Vol. 33, pp. 225-236, December 1999.

[16] J. Bernsen, "Dynamic Thresholding of Gray-level Images," *International Conference on Pattern Recognition*, pp. 1251-1255, Paris, 1986.

[17] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram," *Computer Vision, Graphics, and Image Processing*, Vol. 29, pp. 273-285, 1985.

[18] J. Kittler and J. Illingworth, "Minimum Error Thresholding," *Pattern Recognition*, Vol. 19, pp. 41-47, 1986.

[19] J. M. M. Silva and R. D. Lins, "A Fast Algorithm to Binarize and Filter Documents with Back-to-Front Interference," *ACM Symposium on Applied Computing*, pp. 639-640, Seoul, 2007.

[20] M. Sezgin and B. Sankur, "Survey over Image Thresholding Techniques and Quantitative Performance Evaluation," *Journal on Electronic Imaging*, Vol. 13, pp. 146-168, January 2004.

[21] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Information Processing and Management*, Vol. 45, pp. 427-437, May 2009.

[22] N. Otsu, "A Threshold Selection Method from Gray-level Histogram," *IEEE Transactions on Systems Man Cybernet*, Vol. 9, pp. 62-66, 1979.

[23] N. R. Pal and S. K. Pal, "Entropic Thresholding," *Signal Processing*, Vol. 16, pp. 97-108, 1989.

[24] O. H. Hwa, L. T. Kil, and C. I. Sung, "An Improved Binarization Algorithm Based on a Waterflow Model for Document Image with Inhomogeneous Backgrounds," *Pattern Recognition*, Vol. 38, pp. 2612-2625, September 2005.

[25] P. Sahoo, C. Wilkins,and J. Yeager, "Threshold selection using Renyi's entropy," *Pattern Recognition*, Vol. 30, pp. 71-84, 1997.

[26] P. Soille, "Morphological Image Analysis—Principles and Applications," the 2nd Edition, Springer, Berlin, 2003.

[27] R. C. Gonzalez and R. E. Woods, Digital Image Processing. New Jersey: Prentice-Hall, 2007.

[28] R. F. Moghaddam and M. Cheriet, "A Variational Approach to Degraded Document Enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1347-1361, August 2010.

[29] T. Pun, "Entropic Thresholding: A New Approach," *Computer Graphics and Image Processing*, Vol. 16, pp. 210-239, 1981.

[30] W. H. Tsai, "Moment-Preserving Thresholding: A New Approach," *Computer Vision, Graphics, and Image Processing*, Vol. 19, pp. 377-379, 1985.

[31] X. Tian and X. Hou, "A Tsallis-Entropy Image Thresholding Method Based on Two-Dimensional Histogram Obique Segmentation,"

International Conference on Information Engineering WASE, Vol. 1, pp.164-168, Taiyuan, 2009.

[32] X. Zhang and H. Zhang, "Improved Image Thresholding Based on 2-D Tsallis Entropy," *International Conference on Environmental Science and Information Application Technology*, Vol. 1, pp. 363-366, Chania, 2009.

[33] Library of Congress (http://memory.loc.gov/ammem).

[34] ICFHR 2010 Contest (http://users.dsic.upv.es/~iaprtc5/icfhr2010contest/index.php)

[35] Document Image Binarization (http://www.comp.nus.edu.sg/~tancl/binarization.html)