# Combination of Document Image Binarization Techniques

Bolan Su[12*], Shijian Lu[2+] and Chew Lim Tan[1*]
[1]*Department of Computer Science,School of Computing,National University of Singapore*
*Computing 1, 13 Computing Drive, Singapore 117417*
[2]*Department of Computer Vision and Image Understanding,Institute for Infocomm Research*
*1 Fusionopolis Way, #21-01 Connexis, Singapore 138632*
*{subolan,tancl}@comp.nus.edu.sg
+slu@i2r.a-star.edg.sg

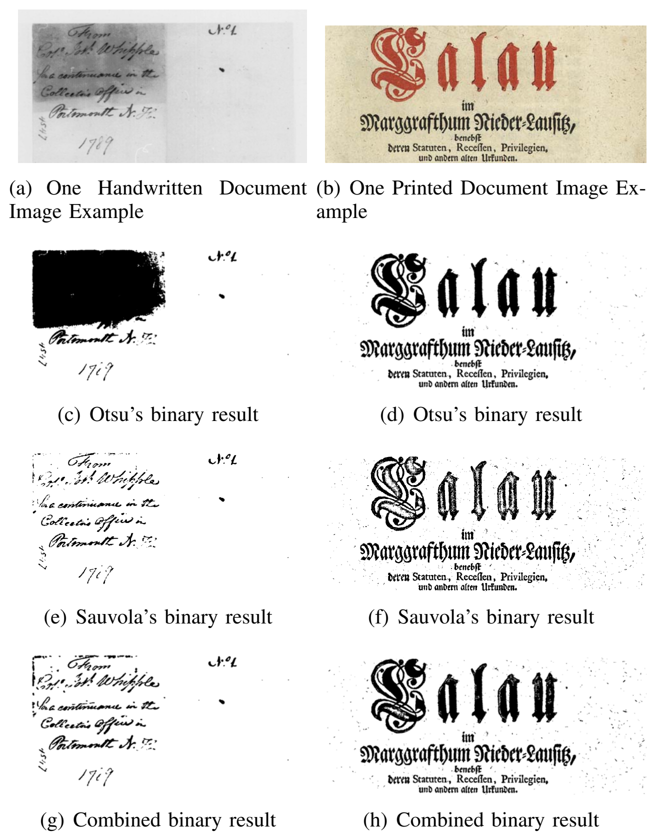
***Abstract*—Document image binarization has been studied for decades, and many practical binarization techniques have been proposed for different kinds of document images. However, many state-of-the-art methods are particularly suitable for the document images that suffer from certain specific type of image degradation or have certain specific type of image characteristics. In this paper, we propose a classification framework to combine different thresholding methods and produce better performance for document image binarization. Given the binarization results of some reported methods, the proposed framework divides the document image pixels into three sets, namely, foreground pixels, background pixels and uncertain pixels. A classifier is then applied to iteratively classify those uncertain pixels into foreground and background, based on the pre-selected froeground and background sets. Extensive experiments over different datasets including the Document Image Binarization Contest(DIBCO)2009 and Handwritten Document Image Binarization Competition(H-DIBCO)2010 show that our proposed framework outperforms most state-of-the-art methods significantly.**

***Keywords*-document image binarization; pixel classification; thresholding technique combination**

## I. Introduction

Document image binarization tries to extract only the text stroke pixels from the gray-scale document images, and is usually performed in the document preprocessing stage. It is an active research area and has been studied for decades because it is important for the ensuing document image processing tasks such as optical character recognition and document layout analysis.

Many document binarization methods [1] have been reported in the literature that can be roughly categorized into two groups: one is global thresholding methods [2], [3] which assign a single threshold for the whole document image, the other is local thresholding methods [4], [5], [6], [7], [8], [9], [10] which assign a threshold for each pixel or a small region of the document images. The global thresholding methods are widely used in many document image analysis applications for their simplicity and efficiency. However, these methods are usually not suitable for degraded document images, because they do not have a clear bimodal pattern that separates foreground text and background. So the local thresholding methods are better approaches for degraded document images with non-uniform background and foreground distribution. The local threshold can be calculated using different information of the document images, such as the mean and standard deviation of pixel values within a local windows [7], [8], water flow model [5], background subtraction [4], [10], illumination model [6] and local image contrast [9]. One drawback of these thresholding approaches is that the thresholding performance depends on the window size and hence the



(a) One Handwritten Document Image Example
(b) One Printed Document Image Example
(c) Otsu's binary result
(d) Otsu's binary result
(e) Sauvola's binary result
(f) Sauvola's binary result
(g) Combined binary result
(h) Combined binary result

Figure 1. Two degraded document image examples and corresponding binarization results produced by Otsu's method, Sauvola's method and our proposed combination framework, respectively.

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.14

character stroke width.

The thresholding of document images is still an unsolved problem due to different types of document degradations, such as uneven illumination, image contrast variation, bleeding-through, and smear. The latest Document Image Binarization Contest (DIBCO) [11] held under the framework of the International Conference on Document Analysis and Recognition (ICDAR) 2009 and Handwritten Document Image Binarization Competition(H-DIBCO)2010 [12] held under the framework of International Conference on Frontiers in Handwriting Recognition (ICFHR)2010 also shows recent efforts on this issue.

The high intensity variation within both the document background and foreground caused by degradations makes it difficult to design an uniform classification method that correctly separates text and background for all kinds of degraded document images. Figure 1(a),(b) shows two examples of the degraded document images. There exists severe intensity variation within the document background in Figure 1(a), which makes Otsu's method generates a bad result as shown in Figure 1(c). And Sauvola's method fails to produce a good result for Figure 1(b) due to the variation of the text stroke width, as illustrated in Figure 1(f).

Instead of designing a new document thresholding technique, we present a learning framework in this paper that combines different existing document threhsolding methods for the purpose of a better thresholding result. B. Gatos et al. [**?**] use a simple voting strategy to combine different binarization methods, which just relabels the pixels with most frequency label assigned by given methods. The combined image can be used as a preliminary binarization result for further analysis. E.Badekas and N. Papamarkos [13] make use of neural network to learn from the binarization results produced by different techniques, this method can work for documents with complex background and images, but it maybe time comsuming. Su et. al. [14] proposed a self-training learning document binarization framework to refine the binarization results of a given document binarization method. However, it depends on the historical binarization records of the examining document binarization method. And it applies the nearest neighbor classification algorithm to the all the document image pixels, which may not be suitable for the high variation characteristics of degraded document images. So we propose a technique to combine existing binarization methods to acquire better binarization performance.

For a given document image, different binarization methods may create different corresponding binary image. Some binarization methods perform superior on certain kinds of document image, while others create better results for other kinds of document images. By combining different binarization techniques, better performance can be achieved with carefully analysis. Those pixels that are labeled the same by different methods are usually correctly classified, and those pixels which are classified as text by some methods and labeled as background by other methods have higher possibility to be misclassified than others. Based on such observation, we divide all the image pixels into three sets: foreground set, where those pixels are classified into foreground by all the examining binarization methods; background set, where those pixels are classified into background by all the examining binarization methods; and uncertain set, where the rest pixels belong to, which is defined as follows:

$$P(x) = \begin{cases} foreground, & \sum_{i=1}^{n} B_i(x) = 0 \\ background, & \sum_{i=1}^{n} B_i(x) = n \\ uncertain, & \text{otherwise} \end{cases} \tag{1}$$

where $P(x)$ denotes one image pixel, $B_i(x)$, which is either 0(foreground) or 1(background), denotes the corresponding binarization result of pixels $P(x)$ generated by the $i_{th}$ binarization methods. The pixels are then projected into a feature space. Those pixels in foreground and background sets can be viewed as correctly labeled samples, and used to determine the label of those uncertain pixels. A classifier is then applied to iteratively classify those uncertain pixels into foreground and background.

Figure 1(g),(h) show the combined results of Otsu and Sauvola's methods. As shown in Figure 1, our proposed framework can produce reasonable results for both the two example document images. The Datasets of the recent DIBCO 2009 [11] and H-DIBCO 2010 [12] are used in our experiments, and experimental results show that our proposed framework significantly improves the accuracy of reported document binarization techniques.

The rest of this paper is arranged as follows. Section 2 first describes our proposed document binarization combination framework in detail. Then experimental results are represented in Section 3 to demonstrate the superior performance of our framework. Finally, conclusions are discussed in Section 4.

## II. Proposed Framework

The overall flowchart of our proposed document binarization combination framework is illustrated in Figure 2. For a given document image, the binarization results are first generated using existing binarization methods, two of those binarization results are combined first using extracted image features of the document image. Then the combined binarization result is used as input of the next round combination procedure along with the next binarization result.

The later part of this section is divided into two subsections, which explain the details of feature extraction and combination of binarization results, respectively.

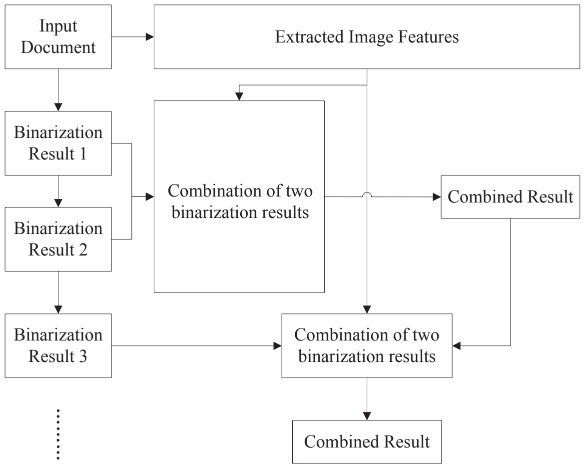



Figure 2. The overall flowchart of our proposed document binarization combination framework.

## A. Feature Extraction

Feature extraction is one of the most important steps in classification. Projecting the image pixels into an appropriate feature space makes the foreground text and document background easier to separate. For document image binarization, the two most frequently used features are intensity and contrast, which is also used in M. Valizadeh and E. Kabir's paper [15]. Because there must be a significant intensity change at the boundary between the foreground text and the document background, or else human being cannot recognize those characters.

Su et al. [9] presented a contrast feature defined as follows:

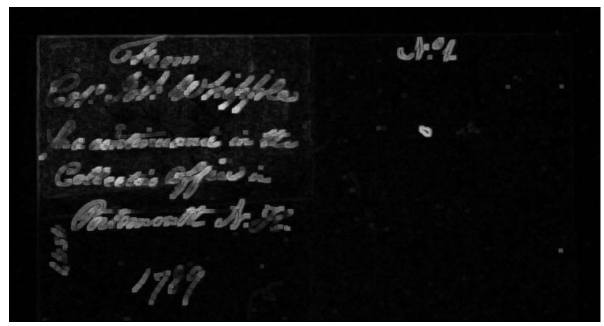
$$Con(x,y) = \frac{f_{max}(x,y) - f_{min}(x,y)}{f_{max}(x,y) + f_{min}(x,y) + \epsilon} \tag{2}$$

where $f_{max}(x,y)$ and $f_{min}(x,y)$ refer to the maximum and the minimum image intensities within a local neighborhood window. The term $\epsilon$ is a positive but infinitely small number, which is added in case the local maximum is equal to 0. And $Con(x,y)$ denotes the contrast value of the estimating pixel $(x,y)$. The numerator $f_{max}(x,y) - f_{min}(x,y)$ captures the local image difference that is similar to the traditional image gradient, the denominator acts as a normalization factor that lowers the effect of the image contrast and brightness variation. So there should be high contrast responses at the area near the boundaries between the text strokes and document background.

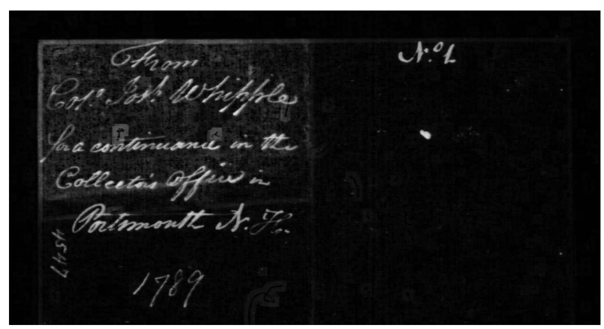
However, the contrast feature defined by Figure 2 leads to high contrast response at either side of text stroke edges. So in our proposed framework, we modify the contrast presentation as follows:

$$Con(x,y) = \frac{f_{max}(x,y) - I(x,y)}{f_{max}(x,y) + \epsilon} \tag{3}$$

where $I(x,y)$ denotes the intensity of pixel $(x,y)$. $f_{max}(x,y)$ refers to the maximum image intensities within a local neighbor window. The term $\epsilon$ is a positive but infinitely small number, which is added in case the local maximum is equal to 0. And $Con(x,y)$ denotes the contrast value of the estimating pixel $(x,y)$ as in Equation 2. In our implementation, the local neighbor window is set to $10 \times 10$. The contrast defined in Equation 3 preserves the ability to suppress the background variation while assigns a more accurate contrast value to document pixels. Figure 3 shows two contrast images generated by Equation 2 and Equation 3, respectively. The contrast map created by Equation 3 makes the text and background more separable than the contrast map created by Equation 2, as shown in Figure 3.



(a) Contrase Map Created by Equation 2



(b) Contrase Map Created by Equation 3

Figure 3. Two contrast map examples of Figure 1(a) generated by Equation 2 and Equation 3, respectively.

Another feature used in our framework is the pixel intensity $I$ of the document image. Therefore, the pixel $(x,y$ is projected to a 2D feature space $[Con, I]$, where $Con$ denotes the contrast feature, and $I$ denotes the intensity feature.

## B. Combination of Binarization Results

After the image pixels are projected into a feature space, we need to use the pixels in the foreground/background set to determine the categories of the uncertain pixels. It is not suitable to compare the examining uncertain pixel with the whole foreground/background set, due to the high variation within both the foreground and background of the degraded document image. So the uncertain pixel is examined under

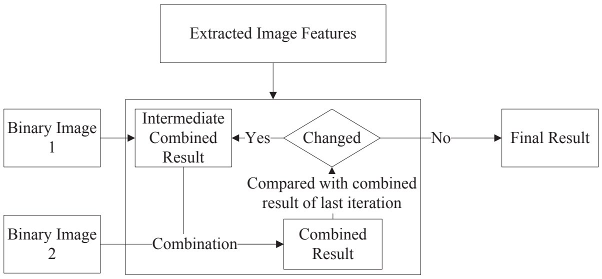



Figure 4. The flowchart of combination of two binarization results.

a local neighborhood window, the pixel is set to background or foreground depending on its distance to local background pixels and foreground pixels, which is defined as follows:

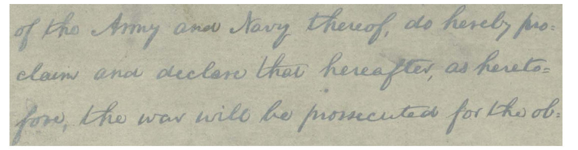
$$P(x) = \begin{cases} foreground, & \frac{Con(x)}{Con_F} > \frac{Con_B}{Con(x)} \| \frac{I_F}{I(x)} > \frac{I(x)}{I_B} \\ background, & \text{otherwise} \end{cases} \quad (4)$$

where $P(x)$ denotes one uncertain pixel, $Con(x), I(x)$ denote the corresponding contrast and intensity features, respectively. $Con_F, I_F$ refer to the mean contrast and intensity feature values of foreground pixels within a local neighborhood window, respectively. And $Con_B, I_B$ refer to the mean contrast and intensity feature values of background pixels within a local neighborhood window, respectively. Since $Con_F > Con_B$ and $I_B > I_F$, $\frac{Con(x)}{Con_F} > \frac{Con_B}{Con(x)}$ and $\frac{I_F}{I(x)} > \frac{I(x)}{I_B}$ mean that distance between local contrast mean value and local intensity mean value of foreground and the examining uncertain pixel is smaller than that of background and the examining uncertain pixel, respectively.

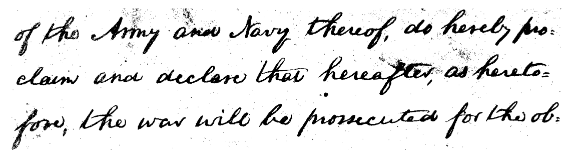
The local neighborhood window is set to $3 \times 3$ in our implementation. There may be no foreground and background pixels within a neighbor window of an uncertain pixel. So we use an iterative strategy to update the foreground/background sets. Only those uncertain pixels that have foreground or background pixels within its neighbor window will be classified into foreground or background in each iteration. The procedure repeats until all the uncertain pixels are classified, which is shown in Figure 4. The input is two binarization resultant images, one is selected as the initial combined result, then the document pixels are divided into three categories using the intermediate combined result and the other binary image. Then some of the uncertain pixels are classified to form the new intermediate combined result, this procedure repeats until the combined result doesn't change, then the final result are produced. It usually takes around 10 iterations to coverage according to experiments.

## III. Experiments and Discussion

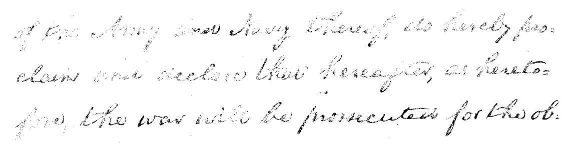
The proposed method has been tested over the dataset recent DIBCO 2009 [11] and H-DIBCO 2010 [12]. The DIBCO 2009 dataset and H-DIBCO 2010 dataset contain



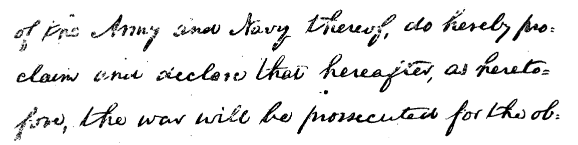
(a) One Handwritten Document Image Example



(b) Otsu's binarization result



(c) Sauvola's binarization result



(d) Combined result

Figure 5. One degraded document image example and corresponding binarization results produced by Otsu's method, Sauvola's method, and combination of the two methods, respectively.

20 historical document images suffering from different kinds of degradations in total. We apply our framework to different well-known document binarization methods, including Otsu's, Sauvola's, Gatos's, Lu's and Su's method [2], [4], [7], [9], [16]. And the four evaluation measures(F-Measure, PSNR, NRM, MPM) adapted from DIBCO's report [11] are used to compare the performance of the testing methods and proposed framework.

The evaluation results are shown in Table I. As shown in Table I, our proposed framework can produce better results than other methods by combining Su's method and Lu's method, which are the best performance method in the DIBCO 2009 contest. And the combined results can perform better in terms of F-Measure, PSNR, NRM than the two origin methods separately. This means a higher precision and better text stroke contour can be obtained after combination.

We also test our framework on the H-DIBCO dataset, Figure 5 shows a degraded document image example taken from H-DIBCO dataset. There are a few noises remained in Otsu's binarization result, as shown in Figure 5(b). And Sauvola's method fails because the contrast of Figure 5(a) is very low, many text stroke pixels are lost as shown in Figure 5(c). After combination, the binarization result preserves most of the text strokes while removing most of the noise pixels as shown in Figure 5(d).

Table I
EVALUATION RESULTS OF THE DATASET OF DIBCO 2009

| Method | F-Measure(%) | PSNR | NRM($\times 10^{-2}$) | MPM($\times 10^{-3}$) |
|---|---|---|---|---|
| Otsu's method | 78.72 | 15.34 | 5.77 | 13.3 |
| Sauvola's method | 85.41 | 16.39 | 6.94 | 3.2 |
| Combined results of Otsu's and Sauvola's method | 86.62 | 16.76 | 3.99 | 4.1 |
| Gatos's method | 85.25 | 16.50 | 10 | 0.7 |
| Su's method | 91.06 | 18.50 | 7 | 0.3 |
| Combined results of Gatos's and Su's method | 91.86 | 18.72 | 3.97 | 0.4 |
| Lu's method | 91.24 | 18.66 | 4.31 | 0.55 |
| Combined results of Lu's method and Su's method | 93.18 | 19.60 | 3.34 | 0.31 |

Our main contribution is to propose a framework that can be used to combine different binarization methods to produce better results. Instead of designing a new binarization method, we try to apply the self-training strategy on existing binarization methods, which improves not only the performance of existing binarization methods, but also the robustness on different kinds of degraded document images. Better performance may be achieved by more sophisticated learning and classification methods. This issue will be investigated in our future work.

## IV. CONCLUSIONS

This paper presents a novel document image binarization combination framework that improves the performance of reported document image binarization methods. The proposed framework divides the image pixels into three categories based on the binary results of given document binarization methods. All the pixels are then projected into a feature space. The pixels in foreground and background sets can be viewed as correctly labeled samples, and used to determine the label of those uncertain pixels. A classifier is then applied to iteratively classify those uncertain pixels into foreground and background. Experiments over the dataset of recent DIBCO 2009 [11] and H-DIBCO [12] demonstrate superior performance of our proposed framework. Experimental results show that the proposed framework can improve the reported binarization methods significantly.

## REFERENCES


[1] 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.

[2] 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.

[3] Y. Solihin and C. Leedham, "Integral ratio: A new class of global thresholding techniques for handwriting images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 761–768, August 1999.

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

[5] 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, 2002.

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

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

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

[9] 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. 303–314, June 2010.

[10] S. J. Lu and C. L. Tan, "Binarization of badly illuminated document images through shading estimation and compensation," *International Conference on Document Analysis and Recognition*, pp. 312–316, 2007.

[11] 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.

[12] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "**H-DIBCO 2010**–handwritten document image binarization competition," *International Conference on Frontiers in Handwriting Recognition*, pp. 727–732, November 2010.

[13] 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.

[14] 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.

[15] M. Valizadeh and E. Kabir, "Binarization of degraded document image based on feature space partitioning and classification." *International Journal on Document Analysis and Recognition*, pp. 1–13, December 2010.

[16] S. Lu, B. Su, and C. L. Tan, "Document image binarization using background estimation and stroke edges," *International Journal on Document Analysis and Recogniton*, vol. 13, December 2010.