Improved Thresholding Method for Enhancing Jawi Binarization Performance

Khairun Saddami*, Khairul Munadi*†, Sayed Muchallil† and Fitri Arnia*†
*Postgraduate Program of Electrical and Computer Engineering
†Department of Electrical and Computer Engineering, Faculty of Engineering
Syiah Kuala University, Banda Aceh, Indonesia
Email: khairun91@mhs.unsyiah.ac.id, khairul.munadi@unsyiah.ac.id, f.arnia@unsyiah.ac.id

***Abstract*—Most local adaptive image binarization techniques have been inspired by Niblack's method and thus use local thresholding. Niblack's and the NICK methods that is based on it introduce a parameter k to determine the object boundaries in a given window, but one of the deficiencies of local thresholding is that the k value is constant. In this paper, we propose a new approach to calculating k values for the NICK binarization method: we set the k value based on the standard deviation of the image. For our experiments, we used the DIBCO 2013 dataset and a privately held ancient Jawi manuscript dataset. The results were evaluated using the F-measure, pseudo F-measure, peak signal-to-noise ratio, misclassication penalty metrics (MPM), distance reciprocal distortion (DRD), and overall rank score. The proposed method achieved result of 91.09% for the DIBCO dataset and 87.07% for the Jawi dataset, which were higher than those with earlier methods that used fixed k values ranging from - 0.2 to - 0.1. These results indicate that the k values produced by the proposed method can adapt to the state of the manuscript and that using them for NICK thresholding can increase binarization performance.**

***Keywords*—adaptive binarization, Jawi historical manuscript, document image binarization, enhancing NICK binarization performance**

Fig. 1. DIBCO 2013 dataset

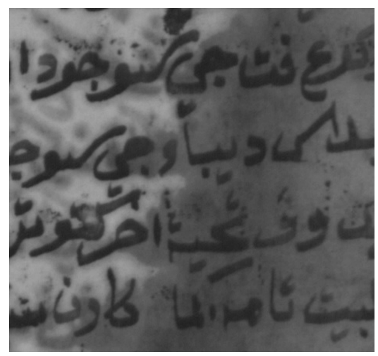

Fig. 2. Degraded Jawi historical manuscript

I. Introduction

Document image binarization is a procedure for transforming grayscale or color images into binary images. "Binarization is one of the pre-preprocessing steps, which include denoising, skew correction, and skeletonization, in optical character recognition (OCR) systems [1], [2]. The aim of binarization is to retain the most significant information in documents. In OCR systems, this mean the text in the document. Most old ancient documents are badly degraded in several ways, such as having uneven brihtness and thin text strokes. This degradation can reduce binarization performance. Figures 1 and 2 show examples of old degraded documents that exhibited thin text strokes and uneven brightness, respectively.

Binarization techniques can be divided into three categories: clustering, thresholding, and hybrid methods [3]. Thresholding-based are simple popular binarization techniques. Thresholding-based methods of binarization are popular as they are simpler than clustering and hybrid techniques. These, in turn, can be divided into two types: global thresholding and local (adaptive) thresholding. The aim of global thresholding is to find a single threshold value for the entire image. Otsu [4], Kittler [5], AbuTaleb [6] have studied global thresholding techniques. This type of thresholding performs very well on good- quality document with uniform brightness and makes it easy to determine the thresholding value quickly. However, it often fails on severely degraded documents images with nonuniform brightness.

Local or adaptive thresholding is therefore preferred for non-uniform, degraded documents due to its flexible selection of the thresholding value. Niblack's method [7] is a benchmark local thresholding method for document image binarization. However, while it can reproduce the text perfectly, the empty areas of the image are often very noisy. Niblack's method defines the threshold value (T) as

$$T = m + k\sigma \tag{1}$$

where m is the mean value over the window and k is used to determine the boundaries of the objects in the window. In Niblack's method, it is set to - 0.2.

2379-2140/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDAR.2017.183

Sauvola [8] developed a new local thresholding method to solve the problem of noise in empty areas that Niblack's suffers from. Wolf [9] and Feng [10] extended Sauvola's and Niblack's methods to produce better binarization results. However, Wolf did not address low contrast images and Feng's method introduces additional parameters as compared to Wolf's method.

In 2009, Khurshid et al., developed an improved version of Niblack's binarization method that resolves the issues with Niblack's and Sauvola's methods. This is known as the NICK method [11], and the NICK thresholding is defined as:

$$T = m + k\sqrt{\frac{\sum P_i^2 - m^2}{NP}} \tag{2}$$

where m is mean value over the window, k is in the range - 0.1 to - 0.2, P_i is the pixel value at i, and NP is the total number of pixels in the window.

Although NICK Binarization deals with the deficiencies Niblack's and Sauvola's methods, the k value introduced in Niblack's method and still used in NICK is a constant. The optimal k value range for the NICK method was recently determined to be from 0.1 to 0.2. According to the authors, a k value of -0.2 removes all noise in the document but causes the text to deteriorate, while a k value of - 0.1 allows the text to be extracted the precisely at the cost of some noises.

According to Bataineh [12], the main issue with the Niblack, NICK, and Sauvola methods is that they, require parameters that are set to constant values and used for all windows during the binarization process. Bataineh's method, which was inspired by NICK, does not use the parameter k and instead calculates adaptive standard deviation value. Bataineh's method is defined as

$$T = m_W - \frac{m_W^2 - \sigma_W}{(m_g + \sigma_W) \times (\sigma_{adaptive} + \sigma_W)} \tag{3}$$

where m_W is the mean value over local window, σ_W is the standard deviation of the local window, m_g is the mean over entire image, and $\sigma_{adaptive}$ refers to the adaptive standard deviation which is defined as

$$\sigma_{adaptive} = \frac{\sigma_W - \sigma_{min}}{\sigma_{max} - \sigma_{min}} \tag{4}$$

Here σ_W refers to the standard deviation over the local window, and σ_{min} and σ_{max} are the minimum and maximum standard deviations respectively, across all windows.

Bataineh's method demonstratess that using adaptive standard deviation can increase local thresholding performance. However, there is a need to calculate the standard deviations for all the windows before calculating the threshold values which raises computing time.

In this paper, we propose a new approach to determine suitable k values for NICK binarization based on the standard deviation over the whole image. The advantages of this new approach are as follows: (1) The k values are adaptive, unlike the NICK method where they are assigned between - 0.2 and - 0.1. (2) Even though Bataineh's method does not use k values, it calculates the standard deviation over all the image windows and than recalculates standard deviations for each window during the binarization process. In contrast, the proposed method only calculates a single standard deviation known as the global standard deviation, thus making the threshold calculation faster.

II. Proposed Method

In this section, we describe the proposed method. As discussed in the introduction, although NICK thresholding is one of the best local adaptive thresholding methods, its k parameter value is a constant. Using a constant k value decreases NICK's binarization performance on severely degraded documents. To resolve this weakness, we propose a new approach to calculate the k value where it is determined by the contrast of the image to be binarized.

Standard deviation measures the contrast level of an image: a higher standard deviation means higher contrast. It is denoted as sigma (σ) and defined in [13], [14]

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}(I_{i,j} - \bar{I})^2} \tag{5}$$

where M and N are the number of columns and rows of the image in pixels, respectively, $\bar{I}$ is the average pixel intensity, and $I_{i,j}$ is the pixel intensity at column i and row j.

Based on this, our new approach calculates the k value as

$$k_f = -\frac{\sigma}{(255 - f \times \sigma)} \tag{6}$$

where σ refers to the standard deviation of the local window and f is a factor that is set to either 1, 2, or $\frac{3}{2}$.

The k factor value is adaptively calculated from the standard deviation for each local window, not manually fixed in the range proposed by Khurram (- 0.1 to - 0.2). Using this k value, the threshold calculated by the proposed method is

$$T = m + k_f\sqrt{\frac{\sum P_i^2 - m^2}{NP}} \tag{7}$$

where m is the mean pixel value over the window, k_f is the proposed k value given in eq. 6, P_i is the pixel value at position i, and NP is total number of pixels in the window.

Based on ig. 3 which is a histogram calculated from Fig. 1, the k values that are generated for binarizing the image in Fig. 1 are $k_{f=1}$ = - 0.049, $k_{f=2}$ = - 0.051 and $k_{f=3/2}$ = - 0.050, respectively. To binarize Fig. 2, based on the histogram in Fig. 4, the k values that are generated are $k_{f=1}$ = - 0.137, $k_{f=2}$ = - 0.158 and $k_{f=3/2}$ = - 0.137, respectively.

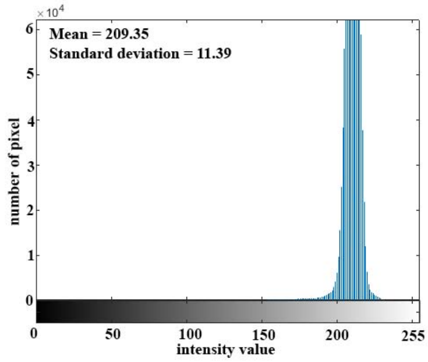

Fig. 3. Histogram showing the mean and standard deviation values for Fig. 1

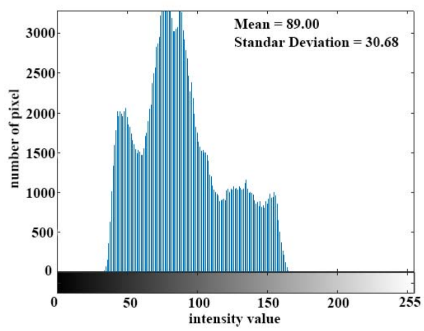

Fig. 4. Histogram showing the mean and standard deviation values for Fig. 2

III. Experiment Condition

The proposed method was tested on the Document Image Binarization Competition (DIBCO) 2013 dataset and a private collection of Jawi historical manuscripts. The DIBCO database consists of printed and handwritten documents, while the Jawi dataset only consists of handwritten documents. In the experiment, we used three values for f value in equation (6), namely 1, 2, and $\frac{3}{2}$, and refer to these as the Proposed I, Proposed II, and Proposed III method, respectively. The window size used was 25 x 25 pixels.

The proposed method was compared with the NICK method for manually-assigned k values of 0.2, -0.15, and -0.1. For this method the window size was set to 19x19 pixels, as recommended in the original paper [11]. The binarization results were evaluated using the DIBCO performance evaluation metrics, namely, F-measure (FM), pseudo F-measure (FMps), peak signal-to-noise ratio (PSNR), misclassification penalty metrics (MPM), distance reciprocal distortion (DRD), and rank score [15], [16]. In addition, we compared the best k value results to those of the Wolf, Feng, and Bataineh methods for binarizing the Jawi ancient documents.

The F-measure is a metric for evaluating whether the binarization process has produced the desired pixel values, and is defined as follows

$$F-measure = 2 \times \frac{RC \times PR}{RC + PR} \tag{8}$$

where RC refers to recall (the fraction of black pixel values produced by the method that are correct) and PR refers to precision (the fraction of black image pixels that have been correctly identified as black by the method).

The PSNR measures the similarity between the ground-truth (GT) image and the result image, and PSNR is defined as

$$PSNR = 10\log(\frac{C^2}{MSE}) \tag{9}$$

where C is the difference between the intensity values of foreground and background pixels and MSE is

$$MSE = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} I'(x,y) - I(x,y) \tag{10}$$

Here, I(x,y) and I'(x,y) are the ground truth and predicted pixel intensities, respectively, at position (x,y).

The pseudo F-measure (FM_{ps}) is a metric for evaluating the binarization of an image based on a GT and is defined as.

$$F_{ps} = 2 \times \frac{RC_{ps} \times PR}{RC_{ps} + PR} \tag{11}$$

To evaluate the contour representation of the resultant image, MPMs were calculated. These are defined as.

$$MPM = \frac{\sum_{i=1}^{F_N} d_{FN}^i + \sum_{i=1}^{F_P} d_{FP}^i}{2D} \tag{12}$$

In addition, DRD metrics were calculated as follows:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \tag{13}$$

The last evaluation technique used was rank score, which sums up the rankings achieved by the method for each of the above performance metrics. Here, lower rankings mean better results.

IV. Performance Analysis

In this section, we analyze the performance of the proposed methods, compared primarily with the NICK method.

A. Result

The experimental results are shown in Table I for the DIBCO 2013 dataset and in Table II for the Jawi historical manuscript dataset. As shown in Table I, Proposed II gave the best results for every evaluation metric except MPM, particularly for handwritten documents, achieving values of 85.12%, 91.06%, 19.03, 3.05, and 5.40 for the FM, FMps, PSNR, MPM, and DRD values, respectively. It produced better binarization results than NICK with k = - 0.1, k = - 0.15, and - 0.2, as well as Proposed I and Proposed III, based on rank score. Overall, Proposed II ranked first for the handwritten DIBCO 2013 dataset, while Proposed III, Proposed I, and the NICK method with k = - 0.2, - 0.15, and - 0.1 ranked second, third, fourth, fifth, and sixth, respectively.

For the printed DIBCO 2013 dataset, Table I shows that the NICK method with k = - 0.15 obtained 84.09%, 92.87%, 15.18, and 5.55 as FM, FMps, PSNR, and DRD values, respectively. On the other hand, NICK method with k = - 0.2 achieved the best MPM result. Overall, the NICK method with k = - 0.15 ranked first for the printed DIBCO 2013 dataset, while Proposed I, Proposed III, NICK method with k = - 0.1, Proposed II, and NICK method with k = - 0.2 ranked second, third, fourth, fifth, and sixth, respectively.

Table II shows the results for the privately held Jawi historical manuscripts, demonstrating that Proposed III has better performance on the FM metric for the Jawi historical documents, achieving 84.13% compared with 84.01% and 83.93% for Proposed I and Proposed II, respectively. In terms of FMps, Proposed II performed better than Proposed I or Proposed III, achieving 87.07% as compared with 86.45% and 86.91% for Proposed I and Proposed II, respectively. This table also shows that Proposed II had better MPM performance for the FMps, and MPM parameters, while the Proposed III had better DRD performance. Proposed II and Proposed III achieved similar scores of 13.64 for PSNR. Overall, Proposed III ranked first for the Jawi historical manuscript dataset, while Proposed II, Proposed I, and the NICK method with k = - 0.15, k = - 0.1, and k = - 0.2 ranked in second, third, fourth, fifth, and sixth, respectively.

B. Discussion

As shown in Fig. 5, the k values calculated by the proposed method produced better results than the other methods, but at the cost of some noise in empty areas. Otherwise, the NICK method with k = - 0.2 produced less noise than with the other k values, but failed to extract thin strokes from the images. Compared to the NICK method, the k values from the proposed method produced better binary images, particularly in extracting thin strokes for text. Based on the experimental results, Proposed III extracted fewer false foreground pixels than the other methods.

Based on the results for the Jawi historical manuscripts, as shown in Fig. 6, Proposed III achieved better performance than the other methods. In particular, the NICK method with k = - 0.2 failed to extract the text of Jawi documents with severe degradation and nonuniform illumination. Overall, Proposed III performed best on the ancient Jawi documents, while Proposed II and Proposed I came second and third, respectively. While these results demonstrate that Proposed III performed best overall, Proposed II achieved higher FMps and MPM results.

The handwritten and printed DIBCO 2013 datasets were slightly different, in that the handwritten dataset contained brighter documents than the printed one as well as more thin text strokes. Therefore, we can conclude that Proposed II gives superior performance on show-through degraded documents with bright backgrounds, as in most of the handwritten DIBCO 2013 documents. Furthermore, Proposed II produced fewer false negatives than Proposed I and Proposed III. On the other hand, Proposed III gave excellent performance for noisy documents with dark (lower brightness) and very dark backgrounds.

As discussed, in the section II, the k values for Fig. 1 are $k_{f=1}$ = - 0.049, $k_{f=2}$ = - 0.051 and $k_{f=3/2}$ = - 0.050. Based on Fig. 5, we can see that these proposed k values extracted text with thin strokes better than NICK. Likewise, the proposed k values for Fig. 2 are f_1 = - 0.137, f_2 = - 0.158 and $f_{3/2}$ = - 0.137, which again produced better binarization results than NICK. When using the previous NICK Thresholding method, selecting suitable k values for specific images is difficult, while the proposed method was able to generate better k values automatically based on image contrast.

Compared with NICK Thresholding with k values in the range - 0.2 to - 0.1, the proposed method produced more adaptive k values that varied with image contrast. With previous methods, it was difficult estimate the right k value for a given image. On the other hand, the standard deviation used by the proposed method can handle image with different contrast levels, as shown in equation 6.

Finally, Table III confirms that the proposed method gives better results than Bataineh thresholding which replaces the k value with an adaptive standard deviation.

V. Conclusion

In this work, we have proposed a new approach to calculate k values for the NICK binarization method, enhancing its performance on degraded ancient Jawi documents. We have introduced an adaptive method of calculating k values based on the images standard deviation. Previously, the k values for NICK were constant, reducing its performance. The experimental results demonstrate that the proposed adaptive k values were able to improve binarization performance for degraded documents, particularly for the ancient Jawi documents. Overall, Proposed III, which use an f factor of $\frac{3}{2}$ performed better than the other methods, particularly for binarizing the ancient Jawi document. For optimal results in practice, we suggest using Proposed III (i.e. $\frac{3}{2}$ as the f factor value). In future work, we will extend this method and integrate additional preprocessing and post-processing stages to enhance the binarization results. In addition, we will test our method against a wider range of datasets than the DIBCO and Jawi datasets used here.

Acknowledgment

This research was supported by the Ministry of Research, Technology, and Higher Education, under the PMDSU scheme .

References

[1] F. Arnia, K. Munadi, S. Muchallil, and F. Fardian, "Improvement of binarization performance by applying dct as pre-processing procedure," in *Communications, Control and Signal Processing (ISCCSP), 2014 6th International Symposium on*. IEEE, 2014, pp. 128–132.

TABLE I
AVERAGE RESULTS FOR DIBCO 2013 DATASET

Handwritten dataset

No.	Method	F-measure (%)	Pseudo F-measure (%)	PSNR	MPM ($\times 10^{-3}$)	DRD	Rank score (sum)	Rank
1	NICK (k = 0.2)	74.96	83.78	17.67	**2.24**	6.97	21	4th
2	NICK (k = 0.15)	77.94	85.00	17.69	4.74	8.72	22	5th
3	NICK (k = 0.1)	76.12	81.74	16.72	11.23	14.67	29	6th
4	Proposed I	83.94	88.89	18.64	4.66	6.80	16	3rd
5	Proposed II	**85.12**	**91.06**	**19.03**	3.05	**5.40**	**6**	**1st**
6	Proposed III	84.61	90.00	18.85	3.82	6.05	11	2nd

Printed dataset

No.	Method	F-measure (%)	Pseudo F-measure (%)	PSNR	MPM ($\times 10^{-3}$)	DRD	Rank score (sum)	Rank
1	NICK (k = - 0.2)	77.34	90.99	14.10	**2.73**	8.23	24	6th
2	NICK (k = - 0.15)	**84.09**	**92.87**	**15.18**	5.85	**5.55**	9	**1st**
3	NICK (k = - 0.1)	83.90	89.76	15.03	9.49	7.62	20	4th
4	Proposed I	82.76	91.67	14.84	5.76	7.19	15	2nd
5	Proposed II	79.89	91.41	14.41	4.49	7.79	21	5th
6	Proposed III	81.71	91.79	14.66	5.09	7.37	16	3rd

TABLE II
AVERAGE RESULTS FOR THE PRIVATE JAWI HISTORICAL MANUSCRIPT DATASET

Handwritten dataset

No.	Method	F-measure (%)	Pseudo F-measure (%)	PSNR	MPM ($\times 10^{-3}$)	DRD	Rank score (sum)	Rank
1	NICK (k = - 0.2)	66.35	69.55	12.10	7.43	12.94	29	6th
2	NICK (k = - 0.15)	80.26	83.47	12.82	7.05	8.12	22	4th
3	NICK (k = - 0.1)	80.63	82.48	12.54	10.46	7.99	24	5th
4	Proposed I	84.01	86.45	13.59	6.63	6.38	13	3rd
5	Proposed II	83.93	**87.07**	**13.64**	**5,67**	6.44	9	2nd
6	Proposed III	**84.13**	86.91	**13.64**	6,09	**6,35**	**7**	**1st**

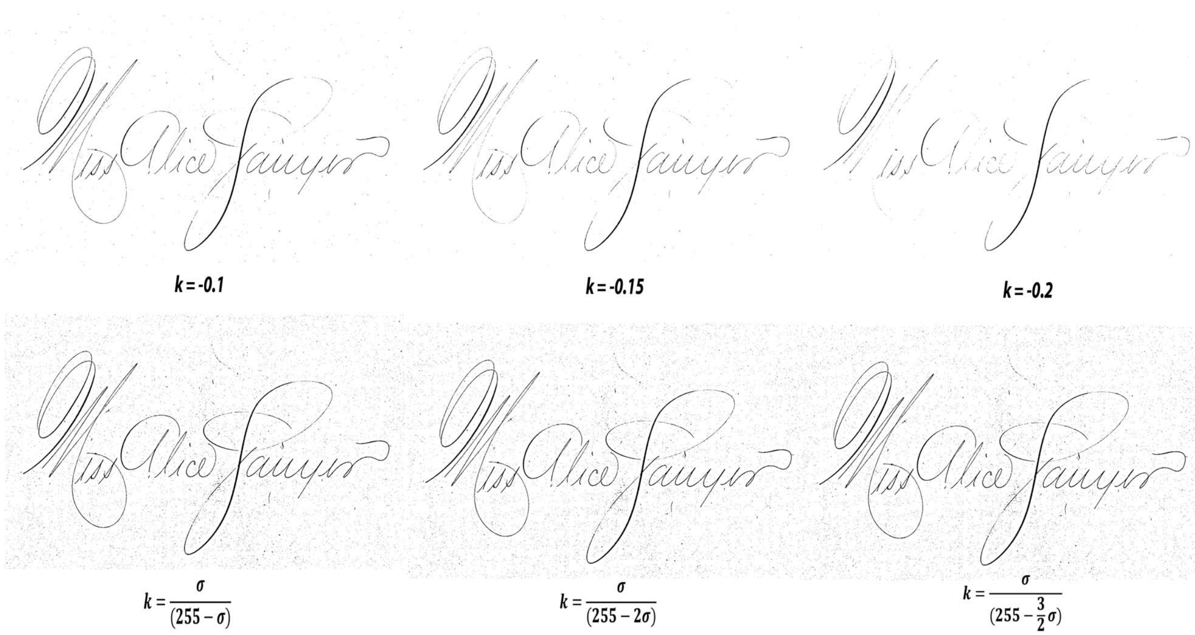

Fig. 5. Sample binarization results for the DIBCO dataset with different k values

Fig. 6. Binarization results for a degraded ancient Jawi document with different k values

TABLE III
AVERAGE PERFORMANCE OF THE THRESHOLDING METHODS FOR THE JAWI DOCUMENT DATASET

Ancient Jawi dataset						
No.	**Method**	**F-measure (%)**	**Pseudo F-measure (%)**	**PSNR**	**MPM ($\times 10^{-3}$)**	**DRD**
1	Wolf	20.72	18.62	9.24	8.24	19.26
2	Feng	37.53	36.27	9.65	9.19	8.73
3	Bataineh	67.46	68.47	9.14	41.89	19.15
4	proposed III	**84.13**	**86.91**	**13.64**	**6,09**	**6,35**

[2] F. Arnia, S. Muchallil, K. Munadi, and Fardian, "Noise characterization in ancient document images based on dct coefficient distribution," in *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*. IEEE, 2015, pp. 971–975.

[3] J. Wen, S. Li, and J. Sun, "A new binarization method for non-uniform illuminated document images," *Pattern recognition*, vol. 46, no. 6, pp. 1670–1690, 2013.

[4] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[5] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern recognition*, vol. 19, no. 1, pp. 41–47, 1986.

[6] A. S. Abutaleb, "Automatic thresholding of gray-level pictures using two-dimensional entropy," *Computer vision, graphics, and image processing*, vol. 47, no. 1, pp. 22–32, 1989.

[7] W. Niblack, *An introduction to digital image processing*. Strandberg Publishing Company, 1985.

[8] J. Sauvola, T. Seppanen, S. Haapakoski, and M. Pietikainen, "Adaptive document binarization," in *Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on*, vol. 1. IEEE, 1997, pp. 147–152.

[9] C. Wolf and J.-M. Jolion, "Extraction and recognition of artificial text in multimedia documents," *Pattern Analysis & Applications*, vol. 6, no. 4, pp. 309–326, 2004.

[10] M.-L. Feng and Y.-P. Tan, "Contrast adaptive binarization of low quality document images," *IEICE Electronics Express*, vol. 1, no. 16, pp. 501–506, 2004.

[11] K. Khurshid, I. Siddiqi, C. Faure, and N. Vincent, "Comparison of niblack inspired binarization methods for ancient documents," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009, pp. 72 470U–72 470U.

[12] B. Bataineh, S. N. Abdullah, K. Omar, and M. Faidzul, "Adaptive thresholding methods for documents image binarization," in *Mexican Conference on Pattern Recognition*. Springer, 2011, pp. 230–239.

[13] I. T. Young, J. J. Gerbrands, and L. J. Van Vliet, *Fundamentals of image processing*. Delft University of Technology Delft, 1998.

[14] E. Peli, "Contrast in complex images," *JOSA A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[15] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "Icdar 2009 document image binarization contest (dibco 2009)," in *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*. IEEE, 2009, pp. 1375–1382.

[16] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "H-dibco 2010-handwritten document image binarization competition," in *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*. IEEE, 2010, pp. 727–732.