Contrast Enhanced Niblack Binarization of Document Images

Sheshera Mysore
Undergraduate Student
Electronics and Telecommunication Department
D.Y. Patil College of Engineering, Akurdi, Pune-44

Prashant Sharma
Undergraduate Student
Electronics and Telecommunication Department
D.Y. Patil College of Engineering, Akurdi, Pune-44

Vivek Rai
Undergraduate Student
Electronics and Telecommunication Department
D.Y. Patil College of Engineering, Akurdi, Pune-44

Priya Charles Assistant Professor Electronics and Telecommunication Department
D.Y. Patil College of Engineering, Akurdi, Pune-44

ABSTRACT
In this paper we propose a method of document image binarization that performs well on grayscale images with complex backgrounds, maintains good text extraction abilities and retains the graphic features that might be present in the image. The proposed method employs a coarse thresholding step that uses a contrast feature for classification of pixels into foreground and background followed by Niblack thresholding for finer classification of the pixels. The proposed method was found to perform better or at-par with four other popular thresholding methods that it was compared against.

General Terms
Image Processing, OCR Preprocessing

Keywords
Document image processing, Binarization, Niblack Thresholding

1. INTRODUCTION
Document image binarization is one of the preliminary steps involved in numerous document image analysis tasks. Binarization refers to the task of converting a given grayscale or color image to a binary image. The binarization of the document image simplifies further tasks such as character recognition, layout analysis, skew estimation and other associated tasks. However, the task of image binarization is a complex task with no one method performing well for every type of image. Work presented in [1] showed that almost all methods perform poorly in images of at-least one kind. This might be attributed to the wide array of document image types as well all the possible ways in which the document image might be degraded.

The image binarization methods might be classified in many ways. We first classify them by the nature of the image that they work on i.e. methods that work directly on full-color images and methods that work on grayscale images. Working with full color images is inherently more complex than working with a grayscale image. We now present some of the methods that work on full color images, followed by methods that work on grayscale images.

A rise in the use of color images has lead to a need for methods that effectively binarize color documents while making use of all of the information that is present in the form of colors in the input image. The conversion of a color image to grayscale inherently loses some of the information present in the full color image. One of the early methods suggested for color document binarization was the one by Tsai and Lee in [2] where they make use of the luminance and saturation features. They employ a decision tree based approach to decide on selection of an appropriate feature for a given region of the image and its subsequent binarization. Badekas *et al.* suggest a method in [3] for binarization that initially reduces the color of the input full color image by use of a color reduction method (as suggested in [4] by Nikolaou and Papamarkos) followed by a connected component analysis and connected component filtering step. This step is followed by a text block formation and finally the text block binarization. In another approach Wang *et al*. [5] suggest a method that uses a color quantization process followed by a robust clustering method that was proposed in [6]. Following this clustering they consider combinations of the reduced color planes as independent binary images and select the optimally combined image as the final output based on the results from a LDA classifier that uses texture features to select the optimal binary image.

Methods that work on grayscale images are better investigated and several well known surveys exist that present and compare different grayscale image binarization methods. A few of these surveys are those of [7] [8] and [1]. Grayscale methods may be classified as has been done in [1] on the basis of the image information that they exploit. This results in methods being classified as histogram based methods, clustering methods, entropy based methods, object attribute based methods, spatial methods using higher order probabilities and methods that are locally adaptive. The thresholding methods are more often classified simply as global thresholding methods and locally adaptive thresholding methods. Global thresholding methods are ones such as those suggested by Otsu [9], Kittler [10] and Kapur *et al*. [11]. These methods tend to perform well only in the cases of simple images which have a clear bimodal histogram. The other class of thresholding methods are the local thresholding methods which are better suited to the binarization of complex images. Local methods of thresholding work to estimate threshold values for local regions based on local image characteristics. Some of these methods were suggested by Niblack [12], Wolf [13] and Sauvola [14].

Aside from these methods there are several methods that have been proposed that are combinations of existing methods. These methods aim to combine the strengths of different methods. The necessity to do this arises, as was mentioned earlier, due to the fact that all of the binarization methods fail to binarize effectively in at-least one type of input image [1]. Su *et al*. [15] presented a framework for the combination of binarization methods where binarization is achieved by an iterative classification of the pixels of the image based on the combined results of the different binarization methods. Kuo *et al.* [16] present a method by which they combine the results of Otsu's thresholding method with the results of Niblack's method and Sauvola's method. Gatos *et al.* [17] present a method in which they combine different binarization methods as well as make use of edge information in conjunction with the run-length smoothing algorithm in the generation of the final binarized image.

In this paper we propose a method for the binarization of scanned grayscale document images that consist of textual information as well as graphic content. In the proposed method we combine Niblack's method [12] along with a contrast feature proposed in [15]. The proposed method works by first performing a coarse thresholding using the contrast feature and then uses Niblack's method for a finer thresholding. Hence the proposed method can handle a fair amount of complex background or a smoothly varying noise in the background while maintaining excellent text extraction. The proposed method also retains the details of the graphic content of the document image well. At this point we justify our decision to use Niblack's thresholding method; It was found in prominent surveys [8] [18] that the method that performs the best in the extraction of text from an image is that suggested by Niblack. However, one of the drawbacks of Niblack's methods is its introduction of noise in areas of relatively constant intensity; the proposed method solves this problem as well. The image resulting from the proposed method is one that contains purely the extracted text regions and the details of the graphic content of the input image.

The remainder of the paper is organized as follows: in section 2 we present our method, in section 3 results and experiments are presented, here we also highlight the strengths as well as the drawbacks of the proposed method. In section 4 we present our conclusion.

2. PROPOSED METHOD

In the proposed method we perform an initial classification of pixels into foreground and background by making use of a contrast feature proposed in [15] and given by (2). Following the coarse thresholding we make use of the Niblack thresholding method [12] for finer thresholding. The contrast feature value of the pixel under consideration is taken as an initial estimate of the class that the pixel belongs to (foreground or background). This hypothesis is justified since majority of the areas of the image that are of interest have a high contrast value as in the case of the text in the image (Figure 1). In the coarse thresholding step which uses the contrast feature we mark those pixels of extremely low contrast as belonging to the background. Following this we perform Niblack thresholding on only those pixels that correspond to the foreground as marked by the coarse thresholding step.

The Niblack thresholding method is a relatively simple local thresholding method that estimates the local threshold $T(x,y)$ in a neighbourhood of size $M\ X\ M$ by use of the following expression:

$$T(x,y) = m + k * s \qquad (1)$$

Here m refers to the local mean and s refers to the local standard deviation in the $M\,X\,M$ neighbourhood and the constant k takes on negative values. In our implementation we use a neighbourhood of $25\ X\ 25$ and the constant k is set to -0.5. A particular drawback that the Niblack's method suffers from is that of a large amount of noise in the binarized image in those areas of the image which correspond to areas of relatively constant intensity. We effectively solve this problem by use of our initial coarse thresholding step. The contrast feature is such that it suppresses background variation while enhancing the contrast values for the foreground document pixels. This feature $C(x,y)$ is given by:

$$C(x,y) = \frac{f_{max}(x,y) - I(x,y)}{f_{max}(x,y) + \varepsilon} \qquad (2)$$

Here, the term $f_{max}(x,y)$ refers to the maximum graylevel value in a neighbourhood of $N\ X\ N$ and the term $I(x,y)$ refers to the graylevel of the pixel on which the neighbourhood is centered. We determine these contrast values over a neighbourhood of size $10\ X\ 10$. In our implementation of the method we set this threshold for the coarse thresholding step at 10% of the maximum contrast value of the image therefore pixels with contrast values lesser than 10% are marked as background pixels. Following this we proceed with the Niblack thresholding for those pixels that have a contrast value greater than 10% of the maximum contrast value. This overall procedure is given by:

$$B(x,y) = \begin{cases} 0, C(x,y) < 0.1C_{max} \\ 1, C(x,y) > 0.1C_{max}\ and\ I(x,y) > T(x,y) \\ 0, C(x,y) > 0.1C_{max}\ and\ I(x,y) < T(x,y) \end{cases} \qquad (3)$$

Here C_{max} refers to the maximum value of the contrast feature C for a given image and the values 0 and 1 correspond to the background and foreground respectively. $B(x,y)$ refers to the output binary image.

By this method we effectively disallow Niblack's method to work on any pixel that belongs to a region of uniform intensity. This in effect brings down the noise that Niblack thresholding would otherwise have added in these regions of uniform intensity. In addition, since we do not alter Niblack's method in any way, the extraction of text and other details of the image continue to be at-least as robust as in the original Niblack's method.

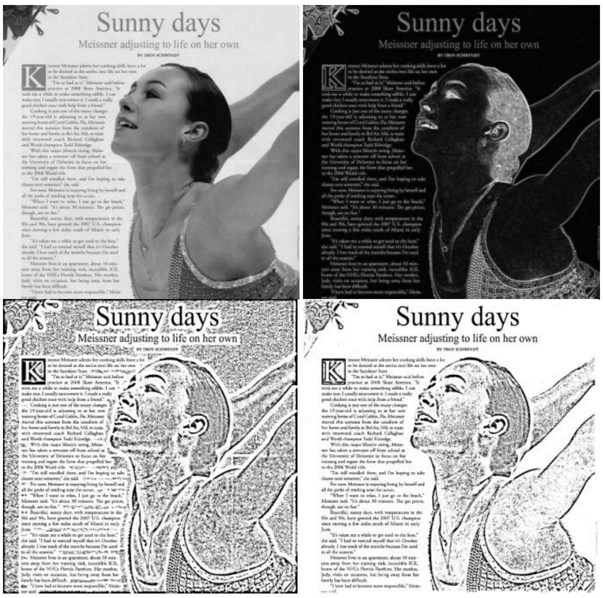


Figure 1- (Clockwise from top) - Input grayscale image, Contrast map of the input image, Output of the proposed method, Output of Niblack's method.

However, we are aware that our initial hypothesis used in the coarse thresholding step might not always be one that is desirable since it is possible that an application demands that a region of uniform intensity (low contrast) be classified as belonging to the foreground. A sample result of the proposed method is presented in Figure 1.

3. EXPERIMENTS AND RESULTS

The proposed method was tested and experimented on further on a dataset of approximately 120 images. These images were chosen to be representative of a wide range of types of input images. The images consisted of the DIBCO 2009 dataset [19], a large number of scanned magazine pages, scanned document images and a few scene images. The input images were binarized with 4 other widely used binarization methods alongside the proposed method. The methods that the proposed method was compared against are those proposed by Otsu [9], Niblack [12], Wolf [13] and Sauvola [14]. The results of these comparisons are presented in this section along with the experiments that were performed. Through the account of the comparison and our experiments we illustrate the drawbacks and the strengths of the proposed method.

In runs of the proposed method on our test dataset it was consistently found to be the most effective in the extraction of text, even in the images which had a very low contrast between the background and the foreground areas (Figure 3 & 4). In addition the proposed method was also consistently found to retain the graphic features the most effectively (Figure 2). The method was also found to effectively handle complexly varying backgrounds as is often found in the pages of a magazine (Figure 4) but this held true only in the case of an un-textured background.

3.1 Comparison between methods

Next we present a comparison between the four methods mentioned previously and the proposed method.

1. The method suggested by Otsu [9] due to being a global thresholding method failed to handle varying backgrounds and degraded images; this was not the case with the proposed method as long as the degradation or background was un-textured.
2. Niblack's method [12] despite being effective in extracting the text and image features also tended to add a lot of noise in those areas of the image which corresponded to areas of constant intensity; the proposed method trumped Niblack's method in this regard. Our implementation of Niblack's method made use of a value of -0.5 for k and a local neighbourhood of $25\ X\ 25$.
3. The method suggested by Wolf [13] though found to be successful in binarizing images with simple backgrounds and images with no graphic content, was not effective in the images which contained more complex backgrounds and in images containing graphic content. Wolfs method was also found to be ineffective in extracting text embedded in darker backgrounds. The proposed method trumped that suggested by Wolf in all of these regards. Our implementation of Wolf's method made use of a value of 0.2 for k and a local neighbourhood of $25\ X\ 25$.
4. In comparison to the proposed method, Sauvolas method caused a loss of the features of the graphic content. In addition Sauvolas method lagged behind the of the proposed method as well as Niblack's method in its effectiveness in extracting text areas, this second result has also been illustrated elsewhere in [20]. Our implementation of Sauvola's method made use of a value of 0.5 for k and a local neighbourhood of $25\ X\ 25$.

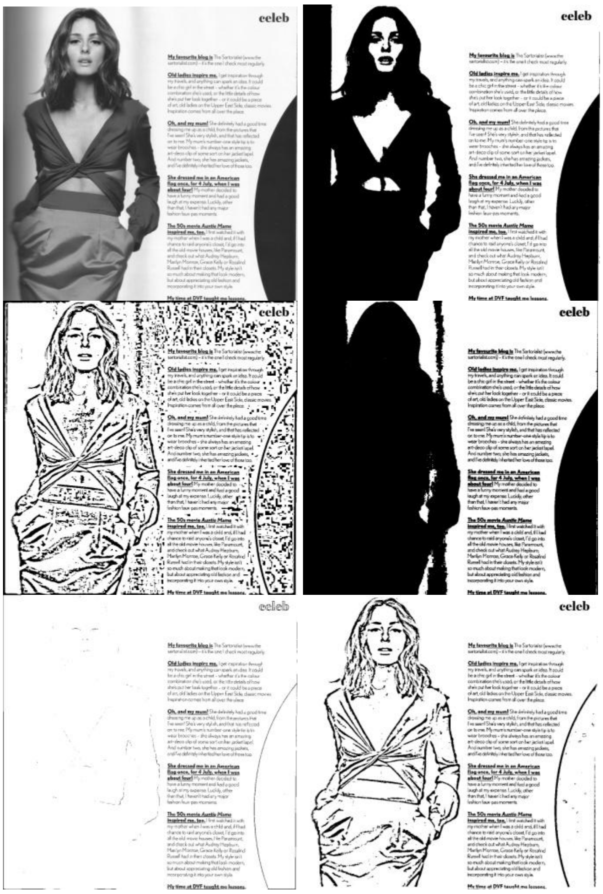


Figure 2- (Clockwise from top) - Input grayscale image, Otsu's output, Wolf's output, Output of the proposed method, Sauvola's output, Niblack's output.

3.2 Drawbacks of the proposed method

Next we elaborate on some of the drawbacks of the proposed method.

1. As was mentioned previously, the proposed method was found to perform poorly in images that consisted of coarsely textured backgrounds. We attribute this shortcoming to the use of the contrast values of the image which are non-zero or are sufficiently high in the background due to the presence of the texture. This drawback could be overcome to an extent by a smoothing (averaging) of the contrast values in a local neighbourhood prior to the coarse thresholding step, but performing this smoothing beyond an extent also caused a loss of the image features in the binarized image. This loss of image features was found to be even more prominent in the graphic content of the image. This smoothing operation on the contrast values was also attempted on images with a non textured background so as to attempt to improve upon the results of the proposed method further. But this lead to a loss of the image features in the graphic portions of the binarized image when done in excess. Due to this loss of image features we do not endorse the process of smoothing the contrast image unless an application demands it and tolerates a loss of graphic features in the output image.

2. The proposed method was also found to be sensitive to filter size if the input image consisted of a large range of character sizes and stroke widths (Figure 4 (c)). The proposed method inherits this drawback from Niblack's method. In addition the contrast features are also

dependent on the neighbourhood size. In the case of a smaller range of character sizes and stroke widths setting the filter size of the contrast estimation neighbourhood as well as the threshold estimation neighbourhood to be large enough to encompass about 2 to 3 characters was found to help in alleviating this problem. But this problem was found to be harder to deal with when the input image consisted of numerous different character sizes and stroke widths as is often the case for pages of a magazine.

3. We also found that although the proposed method was highly effective in extracting text, it was not as effective if the polarity of the text and background were reversed (Figure 4 (b)). This is due to the lack of a step to invert these kinds of text regions in the proposed method. We refer readers to the methods suggested by [3] [21] as methods that can handle a reversal of polarity between foreground and the background.

For an image binarization method to be able to handle textured backgrounds, to be invariant to the size of characters and to be invariant to foreground-background polarity in the input image we suspect that more sophisticated methods will need to be used. In addition a conversion to grayscale might not be advisable. Possible methods which effectively handle the cases stated above more elegantly are the ones suggested in [3] [5] [20] [21].

Also, in the spirit of reproducible research [22] we make the C implementation, which uses the OpenCV library, for the proposed method available at [23].

4. CONCLUSIONS

In this paper we have suggested a simple binarization scheme that works primarily on grayscale document images with complex backgrounds, moderate amounts of degradation and graphic content. The proposed method makes use of a coarse thresholding step by use of a contrast feature followed by a finer thresholding by use of Niblack's method. The method effectively binarizes document images with complex backgrounds, as is usually found in the pages of a magazine. The proposed method also copes with moderate amounts of degradation. In addition it retains features of the graphic content in the image well while having excellent text extraction characteristics. The results of the proposed method were compared against results of four other popular methods and the proposed method was found to perform better or at par with the methods it was compared against.

A few of the drawbacks of the method are, an inability to produce binary images of uniform polarity if the input consists of regions where the foreground-background polarity is inverted, a dependence on local neighbourhood size and a failure to perform well if the document image was textured.

Possible future areas of work include development of methods that effectively handle the drawbacks mentioned above. In addition future areas of work will also involve work on methods that binarize full color images rather than grayscale images as has been proposed here.

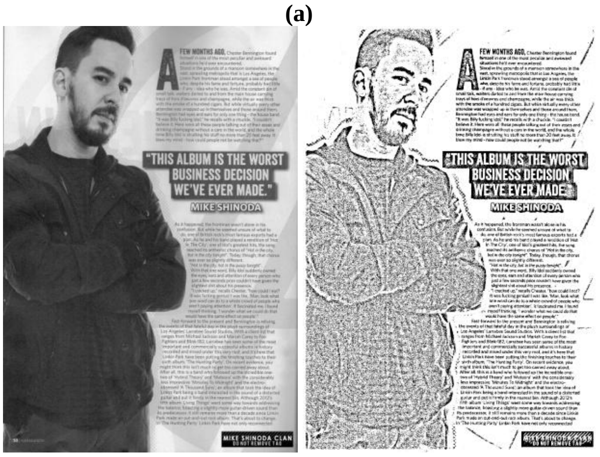
5. ACKNOWLEDGEMENTS

Work presented here was performed while we worked on our undergraduate research project at the Centre for Development of Advanced Computing- Graphics and Intelligence based Script Technology (CDAC-GIST), Pune-411007. We would like to acknowledge the support and guidance provided by Mr. Manish Kumar Gupta at CDAC-GIST in the development

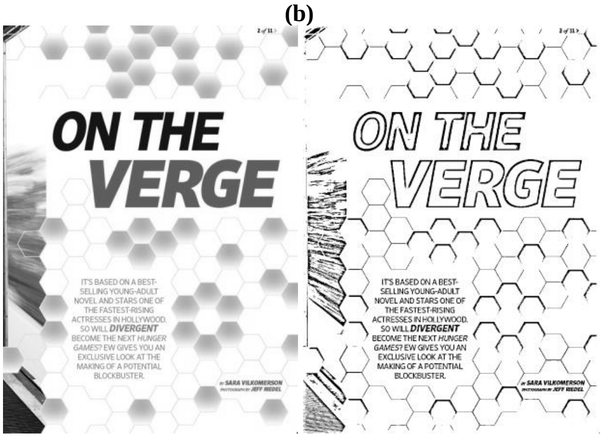
Figure 3- (Clockwise from top) - Input grayscale image, Otsu's output, Wolf's output, Output of the proposed method, Sauvola's output, Niblack's output.

of this work. In addition discussions and feedback from the engineering staff at CDAC-GIST proved valuable as well. We would also like to extend our gratitude towards D.Y.Patil College of Engineering and the Electronics and Telecommunication Department for giving us the opportunity to work on a research project at CDAC-GIST.

(a)



(b)



(c)

Figure 4- (Left to right) - Input grayscale image, Output of the proposed method

6. REFERENCES

[1] M. Sezgin, "Survey over image thresholding techniques and quantitative performance evaluation," *J. Electron. Imaging*, vol. 13, no. January, pp. 146–165, 2004.

[2] C.-M. Tsai and H.-J. Lee, "Binarization of color document images via luminance and saturation color features.," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 434–51, Jan. 2002.

[3] E. Badekas, N. Nikolaou, and N. Papamarkos, "Text binarization in color documents," *Int. J. Imaging Syst. Technol.*, vol. 16, no. 6, pp. 262–274, 2006.

[4] N. Nikolaou and N. Papamarkos, "Color reduction for complex document images," *Int. J. Imaging Syst. Technol.*, vol. 19, no. 1, pp. 14–26, Mar. 2009.

[5] B. Wang, X. Li, F. Liu, and F. Hu, "Color text image binarization based on binary texture analysis," *Pattern Recognit. Lett.*, 2005.

[6] J. Matas and J. Kittler, "Spatial and feature space clustering: Applications in image analysis," *Comput. Anal. Images Patterns*, 1995.

[7] P. Sahoo, S. Soltani, and A. Wong, "A survey of thresholding techniques," *Comput. vision, Graph. image*, pp. 233–260, 1988.

[8] O. Trier and A. Jain, "Goal-directed evaluation of binarization methods," *Pattern Anal. Mach. Intell.* 1995.

[9] N. Otsu, "A threshold selection method from gray level histograms," IEEE Trans. Systems, Man and Cybernetics, vol. 9, pp. 62–66, 1979

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

[11] J. Kapur, P. Sahoo, and A. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Comput. vision, Graph. image*, 1985.

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

[13] C. Wolf, J. Jolion, and F. Chassaing, "Text localization, enhancement and binarization in multimedia documents," *Pattern Recognition, 2002*, pp. 2–5, 2002.

[14] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognit.*, vol. 33, pp. 225–236, 2000.

[15] B. Su, S. Lu, and C. L. Tan, "Combination of Document Image Binarization Techniques," *2011 Int. Conf. Doc. Anal. Recognit.*, pp. 22–26, Sep. 2011.

[16] T. Kuo, Y. Lai, and Y. Lo, "A novel image binarization method using hybrid thresholding," *Multimed. Expo (ICME), 2010*, pp. 608–612, 2010.

[17] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information," *2008 19th Int. Conf. Pattern Recognit.*, pp. 1–4, Dec. 2008.

[18] M. J. Taylor and C. R. Dance, "Enhancement of Document Images from Cameras", Proc. of IS&T/SPIE EIDR V, pp. 230-241, 1998.

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

[20] S. Milyaev and O. Barinova, "Image binarization for end-to-end text understanding in natural images," *(ICDAR), 2013 12th*, 2013.

[21] T. Kasar, J. Kumar, and A. G. Ramakrishnan, "Font and Background Color Independent Text Binarization," vol. 2, no. 1, pp. 3–9, 2007.

[22] S. Fomel and J. F. Claerbout, "Guest editors' introduction: Reproducible research," Computing in Science and Engineering, vol. 11, no. 1, pp. 5–7, 2009

[23] Bitbucket URL: http://bit.ly/1Hi9H1