

Foreground-Background Regions Guided Binarization of Camera-Captured Document Images

Syed Saqib Bukhari¹, Faisal Shafait², Thomas M. Breuel^{1,2}

¹Technical University of Kaiserslautern, Germany

²German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
bukhari@informatik.uni-kl.de, faisal@iupr.dfki.de, tmb@informatik.uni-kl.de

Abstract

Binarization is an important preprocessing step in several document image processing tasks. Nowadays handheld camera devices are in widespread use, that allow fast and flexible document image capturing. But, they may produce degraded grayscale image, especially due to bad shading or non-uniform illumination. State-of-the-art binarization techniques, which are designed for scanned images, do not perform well on camera-captured documents. Furthermore, local adaptive binarization methods, like Niblack [1], Sauvola [2], etc, are sensitive to free parameter values, which are fixed for whole image. In this paper, we describe a novel binarization technique using ridges-guided local binarization method, in which appropriate free parameter value(s) is(are) selected for each pixel depending on the presence or absence of ridge(s) in the local neighborhood of a pixel. Our method gives a novel way of automatically selecting parameter values for local binarization method, this improves binarization results for both scanned and camera-captured document images relative to previous methods. Experimental results on a subset of CBDAR 2007 document image dewarping contest dataset show a decrease in OCR error rate using reported method with respect to other state-of-the-art binarization methods.

1 Introduction

Most of the state-of-the-art document analysis systems have been designed to work on binary images [3]. Therefore, document image binarization is an important initial step in most of the document image processing tasks, like page segmentation [4], layout analysis [5, 6] or recognition. Performance of these tasks heavily depends on the results of binarization. The main objective of document image binarization is to divide a grayscale or color document into

two groups, that are foreground text/images and clear background.

On one hand, cameras offer fast, easy and non-contact document imaging as compared to scanners and are in more common use nowadays. But on the other hand, the quality of camera-captured documents is worse as compared to scanned documents because of the degradations which are not very common in scanned images, like non-uniform shading, image blurring and lighting variations. Due to this, binarization of camera-captured documents is more challenging than scanned documents.

From decades, many different approaches for the binarization of grayscale [7, 2, 8, 9, 1, 10, 11, 12, 13, 14] and color [15, 16, 17] documents have been proposed in the literature. Additionally, grayscale binarization techniques can be applied by first converting the color documents into grayscale. Grayscale binarization approaches can be classified into two main groups: i) global binarization methods and ii) local binarization methods.

Global binarization methods, like Otsu [7], try to estimate a single threshold value for the binarization of whole document. Then based on the intensity values, each pixel is assigned either to foreground or background. Global binarization methods are computationally inexpensive and perform better for typical scanned document images. However, they produce marginal noise artifacts if grayscale document contains non-uniform illumination, which is usually present in case of scanned thick book, camera-captured document or historical document.

Local binarization methods, like Sauvola [2], try to overcome these problems by calculating threshold values for each pixel differently using local neighborhood information. They perform better on degraded document images but are computationally slow and sensitive to the selection of window size and free parameter values [18]. Some special techniques [11, 12, 13] based on local binarization have been proposed recently for improving the binarization

results of degraded camera-captured and historical documents. In general these methods produce good binarization results under non-uniform illumination as compared to other types of local binarization methods, but are still sensitive to free parameter values.

In this paper, we deal with the binarization of degraded grayscale camera-captured document images. Here, we describe a local binarization method based on Sauvola's binarization method, which is less sensitive to free parameter values. Unlike Sauvola's method, instead of using the same free parameter values for all pixels, we select different values for foreground and background pixels. We use ridges detection technique for finding foreground regions information.

The rest of the paper is organized as follows: Section 2 explains the technical and implementation details of our binarization method. Section 3 deals with experimental results and section 4 describes conclusion.

2 Foreground-Background Guided Binarization

Researchers [19, 20] have evaluated different state-of-the-art global and local binarization methods and reported that Sauvola's binarization method [2] is better than other types of local binarization methods for degraded document images. But the performance of local binarization methods is sensitive to free parameter values [18]. Our binarization method, presented here, is an extension of Sauvola's method. In section 2.1 we discuss about the Sauvola's binarization method and how to improve its performance by selecting different free parameter values for foreground and background pixels. In section 2.2 we describe the method for detecting foreground regions using ridges. In section 2.3 we describe the guided Sauvola's binarization method with respect to foreground/background regions information. .

2.1 Local Binarization using Sauvola's method

Grayscale document images contain intensity values in between 0 to 255. Unlike global binarization, local binarization methods calculate a threshold $t(x, y)$ for each pixel such that

$$b(x, y) = \begin{cases} 0 & \text{if } g(x, y) \leq t(x, y) \\ 255 & \text{otherwise} \end{cases} \quad (1)$$

The threshold $t(x, y)$ is computed using the mean $\mu(x, y)$ and standard deviation $\sigma(x, y)$ of the pixel intensities in a $w \times w$ window centered around the pixel (x, y) in

Sauvola's binarization method:

$$t(x, y) = \mu(x, y) \left[1 + k \left(\frac{\sigma(x, y)}{R} - 1 \right) \right] \quad (2)$$

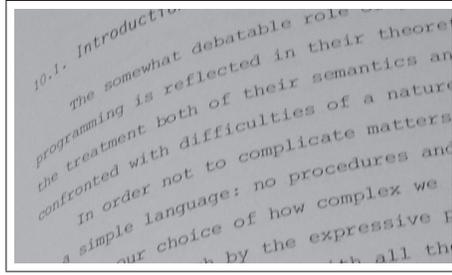
where R is the maximum value of the standard deviation ($R = 128$ for a grayscale document), and k is a parameter which takes positive values. The formula (Equation 2) has been designed in such a way that, the value of the threshold is adapted according to the contrast in the local neighborhood of the pixel using the local mean $\mu(x, y)$ and local standard deviation $\sigma(x, y)$. Because of this, it tries to estimate appropriate threshold $t(x, y)$ for each pixel under both possible conditions: high and low contrast. In case of local high contrast region ($\sigma(x, y) \approx R$), the threshold $t(x, y)$ is nearly equal to $\mu(x, y)$. Under quite low contrast region ($\sigma \ll R$), the threshold goes below the mean value thereby successfully removing the relatively dark regions of the background. The parameter k controls the value of the threshold in the local window such that the higher the value of k , the lower the threshold from the local mean $m(x, y)$.

The statistical constraint in Equation 2 gives acceptable results even for degraded documents. But, there is a contradiction regarding the appropriate value of k in research community. Badekas et al. [20] experimented with different values and found that $k = 0.34$ gives the best results, but Sauvola[2] and Sezgin[19] used $k = 0.5$.

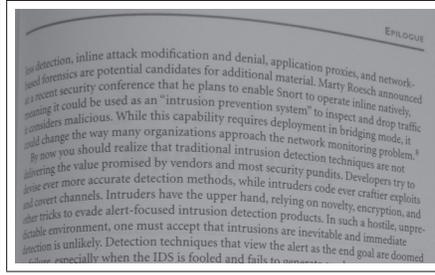
We have analyzed Sauvola's binarization method with different values of k for degraded camera-captured document images. Some of the experimental results are shown in the Figure 1. These results clearly show the sensitivity of Sauvola's binarization on the value of k . Additionally, already reported values of k , i.e $k = 0.5$ [2, 19] and $k = 0.34$ [20], do not give acceptable result under blurring or non-uniform illuminations, as shown in Figure 1.

However, we have noticed that, $k = 0.2$ gives low noise in the background but produces broken characters, shown in Figures 1(g) and 1(h). On the other hand, $k = 0.05$ gives good results for foreground text/images pixels with unbroken characters but with some noise in the background, as shown in Figures 1(i) and 1(j). These experiments allow us to claim that, Sauvola's method can perform better on degraded documents, if we use different value of k for each pixel depending upon its association with foreground or background region.

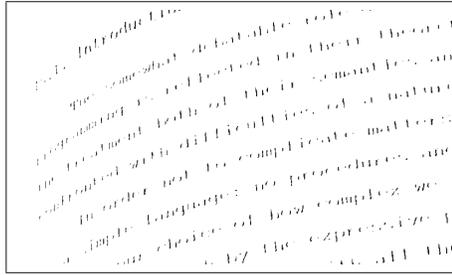
In next section we describe the method of estimating foreground region using ridges detection and in section 2.3 we describe the adaptation of Sauvola's method using foreground/background region information.



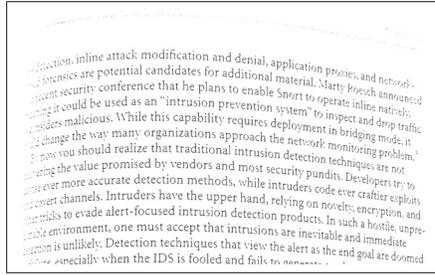
(a) Degraded camera-captured image with blurring.



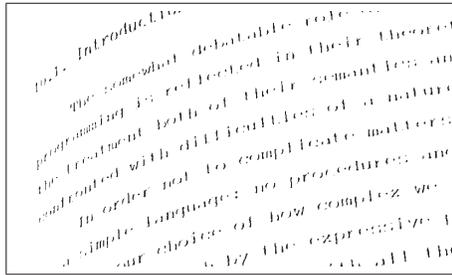
(b) Degraded camera-captured image with non-uniform illumination.



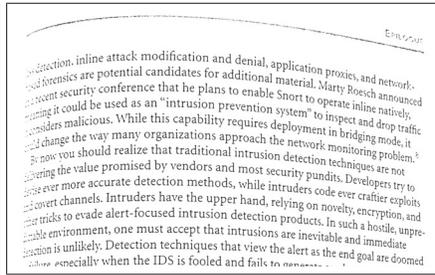
(c) $k = 0.5$.



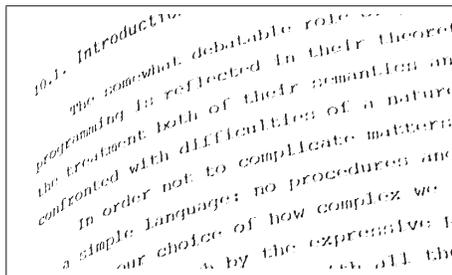
(d) $k = 0.5$.



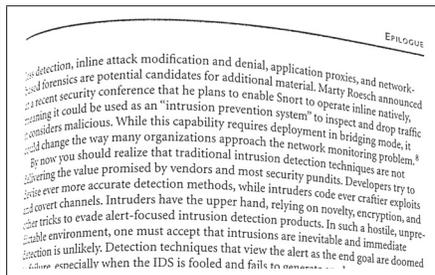
(e) $k = 0.34$.



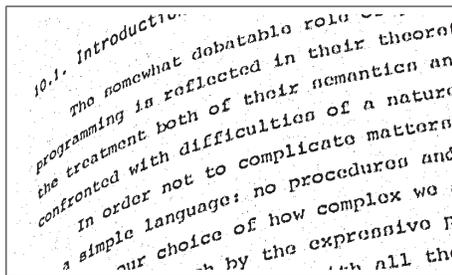
(f) $k = 0.34$.



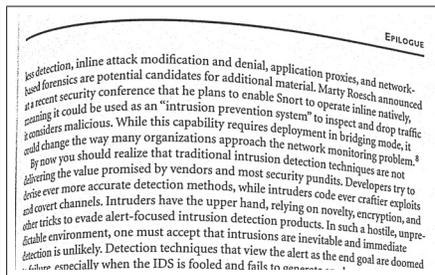
(g) $k = 0.2$.



(h) $k = 0.2$.



(i) $k = 0.05$.



(j) $k = 0.05$.

Figure 1. Sauvola’s binarization results for different values of k . $k = 0.5$ is reported by Sauvola[2] and Sezgin[19]. $k = 0.34$ is used by Badekas et al. [20]. We have selected $k = 0.2$ and $k = 0.05$. With $k = 0.2$, results have cleaned-background and broken-foreground-characters. And with $k = 0.05$ results have uncleaned-background and unbroken-foreground-characters.

2.2 Foreground Regions detection using Ridges

We have already described textline detection techniques for handwritten and camera-captured documents using ridges in [21, 22]. In this paper, we use this technique for finding foreground regions, that are central lines structure of textlines and drawings. Detection of foreground regions using ridges is divided into two sub steps: (i) image smoothing and (ii) ridges detection. Following sections discuss these steps in detail.

2.2.1 Image Smoothing

Camera-captured document images contain variety of curled textlines and drawings structure with respect to size and orientation angle. Match filter bank approach has been used for enhancing the structure of multi-oriented blood vessels [23] and finger prints [24]. In [21, 22], we have described multi-oriented multi-scale anisotropic Gaussian smoothing, based on matched filter bank approach, for enhancing textlines structure. In this paper, we use multi-oriented multi-scale anisotropic Gaussian smoothing for enhancing curled textlines and drawings structure. A single range is selected for both σ_x and σ_y , which is the function of the height of the document image (H), that is aH to bH with $a < b$. The suitable range for θ is from -45 to 45 degrees. From these ranges, a set of filters is generated for different combinations of σ_x , σ_y and θ . This set of filters is applied to each pixel of grayscale image and the maximum resulting value is selected. Figures 2(a) and 2(b) show the input and smoothed images respectively.

2.2.2 Ridges Detection

Multi-oriented multi-scale anisotropic Gaussian smoothing enhances the foreground structure well, which is clearly visible in Figure 2(b). Now the task is to find the foreground regions information. Since decades, ridges detection has been popularly used for producing rich description of significant features from smoothed grayscale images [25] and speech-energy representation in time-frequency domain [26]. Ridges detection over smoothed image can produce unbroken central lines structure of foreground textlines/drawings. In this paper, Horn-Riley [25, 26] based ridges detection approach is used. This approach is based on the informations of local direction of gradient and second derivatives as the measure of curvature. From these informations, which are calculated by Hessian matrix, ridges are detected by finding the zero-crossing of the appropriate directional derivatives of smoothed image. Detected Ridges

over the smoothed image of Figure 2(b) are shown in Figure 2(c) and Figure 2(d). It is clearly visible in the Figure 2(c) that ridges are present where the foreground data are present and each ridge covers the complete central line structure of a foreground object.

2.3 Foreground-Background Guided Sauvola's Binarization

We have already discussed in section 2.1 that no single value of parameter k in Sauvola's method is suitable for different types of degraded camera-captured documents. But $k = 0.05$ gives good results for foreground textlines/drawings with some background noise and $k = 0.2$ gives noise free background with broken characters, as shown in Figure 1. Ridges have been detected in section 2.2, that give information about foreground data. Therefore, instead of using fixed value of k for all pixels, we use different values of k for foreground and background pixels to improve the binarization result. We redefine Sauvola's binarization method, such that:

$$t(x, y) = \mu(x, y) \left[1 + \mathbf{k}(\mathbf{x}, \mathbf{y}) \left(\frac{\sigma(x, y)}{R} - 1 \right) \right] \quad (3)$$

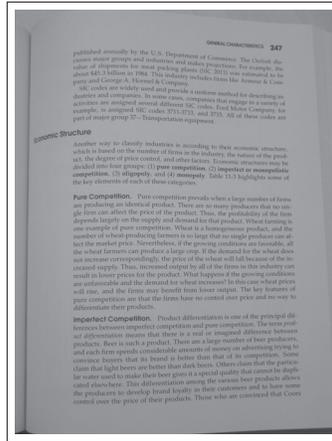
where $\mathbf{k}(\mathbf{x}, \mathbf{y})$ is equal to 0.05 if ridge(s) is(are) present in the local neighborhood window $w \times w$ window centered around the pixel (x, y) , otherwise equal to 0.2. After thresholding, median filter is applied to remove the salt and pepper noise. Binarization results based on foreground/background guided Sauvola's method are shown in Figures 2(e) and 2(f).

3 Experiments and Results

We evaluate our binarization approach on the hand-held camera-captured document images dataset used in CBDAR 2007 for document image dewarping contest [27]. For this purpose, we have selected 10 degraded documents from the dataset. State-of-the-art Otsu's [7] and Sauvola's [2] binarization methods are used for comparative evaluation. The results of Otsu's, Sauvola's and foreground-background guided Sauvola's binarization methods on some example documents are shown in Figure 3.

We compare the OCR error rate of all three binarization methods for 10 selected documents. These documents have non-planar shape, therefore we apply dewarping algorithm¹ on the results of all three binarization methods. Then dewarped documents of all methods are processed through a commercial OCR system **ABBYY Fine Reader 9.0**. After

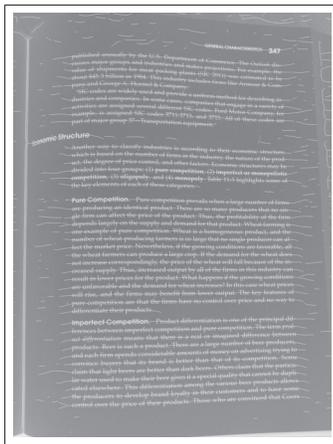
¹We have described dewarping method using ridges based coupled-snakes model, which is currently in review phase of CBDAR 2009.



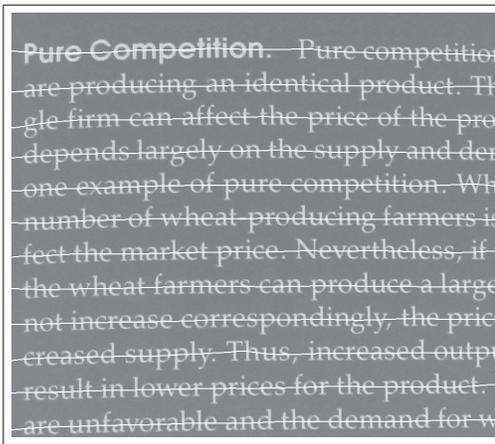
(a) Input Image.



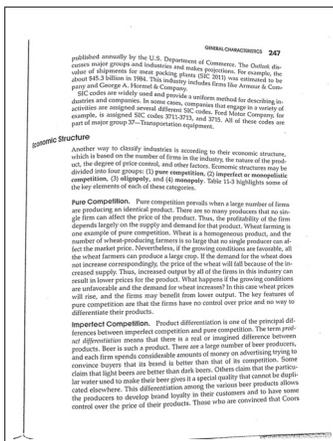
(b) Smoothed Image generated by using match filter bank approach.



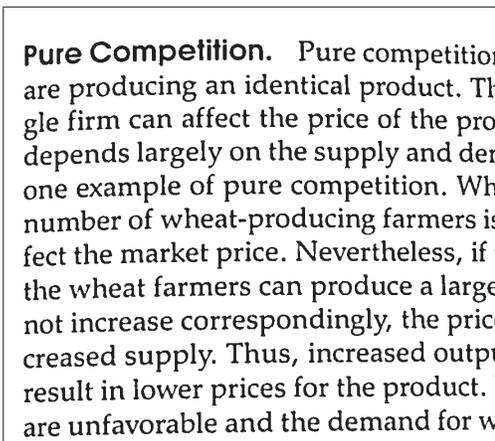
(c) Horn-Riley method [25, 26] is used for detecting ridges.



(d) Closeup portion of detected ridges.

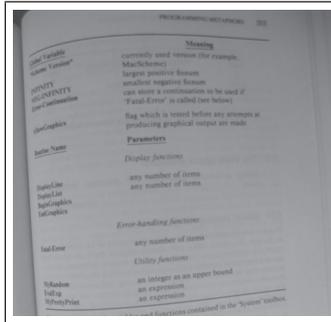


(e) Result of foreground/background guided Sauvola's binarization.

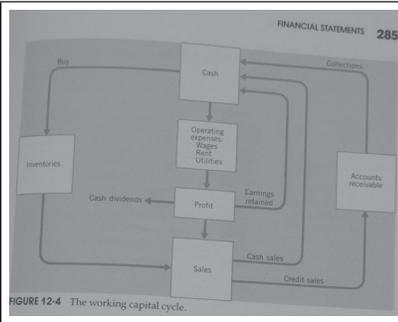


(f) Closeup portion of binarized result.

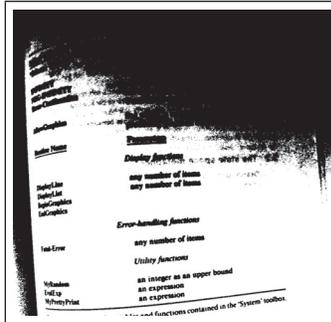
Figure 2. Binarization algorithm snapshots.



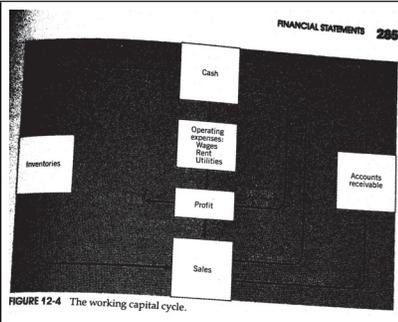
(a) Input Image



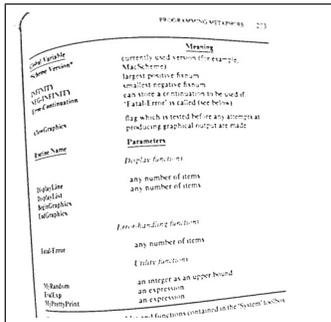
(b) Input Image



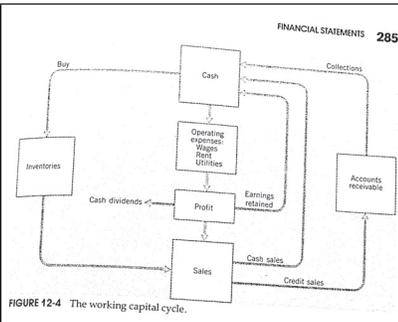
(c) Otsu's result



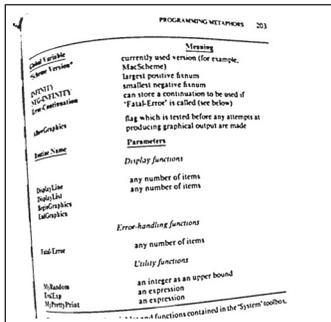
(d) Otsu's result



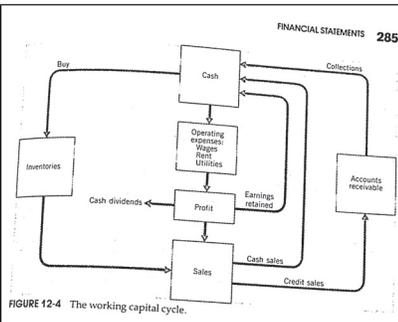
(e) Sauvola's result



(f) Sauvola's result



(g) Guided-Binarization's result



(h) Guided-Binarization's result

Figure 3. Binarization results of Otsu [7], Sauvola [2] and our Guided-Binarization. Note that Otsu's results have large amount of noise. For Sauvola's binarization we have manually selected the appropriate parameter values $w = 15$ and $k = 0.15$ for given dataset. Sauvola's results ($w = 15, k = 0.15$) have broken-characters for blurred images. Our proposed guided binarization method shows better results for both text and drawing regions, even in the presence of blurring.

obtaining text from the OCR software, the block edit distance² with the ASCII ground-truth has been used as the error measure. Table 1 shows the comparative results of all methods with respect to mean edit distance, median edit distance and the number of documents for each algorithm on which it has the lowest edit distance (in case of tie, all algorithms having the lowest edit distance are scored for that document).

4 Conclusion

In this paper we presented a novel way of automatically selecting free parameter values for locally adaptive binarization methods. Local binarization methods, like Niblack's [1] and Sauvola's [2] binarization, use constant values of free parameter for all pixels in the image and are sensitive to these values. We overcome this sensitivity by not using constant values of free parameters for all pixels. We used different free parameter values in Sauvola's methods for foreground and background pixels and achieved promising results for degraded camera-captured documents having blurring and non-uniform illumination. We have also described the simple and efficient way of finding foreground regions of document image using ridges detection.

Comparative results in Figure 3 and Table 1 show that, our guided Sauvola's method outperforms other state-of-the-art global and local binarization methods for degraded documents. Furthermore, our method of selecting free parameter values can also be used with other types of local binarization techniques.

References

- [1] W. Niblack. *An Introduction to Image Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1986.
- [2] J. Sauvola and M. Pietikainen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.
- [3] R. Cattoni, T. Coianiz, S. Messelodi, and C. M. Modena. Geometric layout analysis techniques for document image understanding: a review. Technical report, IRST, Trento, Italy, 1998.
- [4] F. Shafait, D. Keysers, and T. M. Breuel. Performance evaluation and benchmarking of six page segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):941–954, Jun 2008.
- [5] F. Shafait, J. V. Beusekom, D. Keysers, and T. M. Breuel. Structural mixtures for statistical layout analysis. In *Proceedings 8th International Workshop on Document Analysis Systems*, pages 415–422, Nara, Japan, 2008.
- [6] F. Shafait, J. V. Beusekom, D. Keysers, and T. M. Breuel. Background variability modeling for statistical layout analysis. In *Proc. 19th International Conference on Pattern Recognition (ICPR)*, 2008. Accepted for publication.
- [7] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions Systems, Man and Cybernetics*, 9(1):62–66, 1979.
- [8] J. M. White and G. D. Rohrer. Image thresholding for optical character recognition and other applications requiring character image extraction. *IBM Journal of Research and Development*, 27(4):400–411, July 1983.
- [9] J. Bernsen. Dynamic thresholding of gray level images. In *Proceedings 8th International Conference on Pattern Recognition*, pages 1251–1255, 1986.
- [10] L. O’Gorman. Binarization and multithresholding of document images using connectivity. *Graphical Model and Image Processing*, 56(6):494–506, Nov. 1994.
- [11] In-Jung Kim. Multi-window binarization of camera image for document recognition. In *Proceedings 9th International Workshop on Frontiers in Handwriting Recognition*, pages 323–327, Washington, DC, USA, 2004.
- [12] B. Gatos, I. Pratikakis, and S. J. Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39(3):317–327, 2006.
- [13] S. Lu and C. L. Tan. Thresholding of badly illuminated document images through photometric correction. In *Proceedings 2007 ACM symposium on Document engineering*, pages 3–8, Winnipeg, Manitoba, Canada, 2007.
- [14] F. Shafait, D. Keysers, and T. M. Breuel. Efficient implementation of local adaptive thresholding techniques using integral images. In *Proceedings 15th International Conference on Document Recognition and Retrieval*, volume 6815, page 81510, San Jose, CA, USA, 2008.

²<http://sites.google.com/site/ocropus/release-notes>

Table 1. OCR error rates of different binarization algorithms on subset of dataset of CBDAR 2007 Document Image Dewarping Contest using ABBYY Fine Reader 9.0.

| Algorithm | Mean Edit Distance % | Number of documents ^a |
|-------------------------------------|----------------------|----------------------------------|
| Otsu's Binarization | 6.96 | 2 |
| Sauvola's Binarization ^b | 4.92 | 3 |
| Guided-Binarization | 4.62 | 5 |

^aNumber of documents for each algorithm on which it has the lowest edit distance.

^bmanually selected: ($w = 15, k = 0.15$); tested different values for k in between 0.1 to 0.5 and found 0.15 is the best for the given dataset.

- [15] K. Sobottka, H. Kronenberg, T. Perroud, and H. Bunke. Text extraction from colored book and journal covers. *International Journal on Document Analysis and Recognition*, 2(4):163–176, June 2000.
- [16] C.M. Tsai and H.J. Lee. Binarization of color document images via luminance and saturation color features. *IEEE Transactions on Image Processing*, 11(4):434–451, April 2002.
- [17] E. Badeskas, N. Nikolaou, and N. Papamarkos. Text binarization in color documents. *International Journal of Imaging Systems and Technology*, 16(6):262–274, 2006.
- [18] Y. Rangoni, F. Shafait, and T. M. Breuel. Ocr based thresholding. In *Proceedings IAPR Conference on Machine Vision Applications*, Yokohama, Japan, 2009.
- [19] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–165, 2004.
- [20] E. Badeskas and N. Papamarkos. Automatic evaluation of document binarization results. In *Proceedings 10th Iberoamerican Congress on Pattern Recognition*, pages 1005–1014, Havana, Cuba, 2005.
- [21] S. S. Bukhari, F. Shafait, and T. M. Breuel. Script-independent handwritten textlines segmentation using active contours. In *Proceedings 10th International Conference on Document Analysis and Recognition*, Barcelona, Spain, 2009.
- [22] S. S. Bukhari, F. Shafait, and T. M. Breuel. Ridges based curled textline region detection from grayscale camera-captured document images. In *Proc. The 13th International Conference on Computer Analysis of Images and Patterns*, Mnster, Germany, 2009.
- [23] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transaction on Medical Imaging*, 8(3):263–269, 1989.
- [24] L. O. Gorman. Matched filter design for fingerprint image enhancement. In *Proceedings International Conference on Acoustics, Speech, and Signal Processing*, pages 916–919, New York, NY, USA, 1988.
- [25] B. K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. *PhD Thesis, MIT*, 1970.
- [26] M. D. Riley. Time-frequency representation for speech signals. *PhD Thesis, MIT*, 1987.
- [27] F. Shafait and T. M. Breuel. Document image dewarping contest. In *Proceedings 2nd International Workshop on Camera Based Document Analysis and Recognition*, pages 181–188, Curitiba, Brazil, 2007.