Journal of Universal Computer Science, vol. 25, no. 6 (2019), 627-646
submitted: 7/1/19, accepted: 20/5/19, appeared: 28/6/19

Fast Binarization of Unevenly Illuminated Document Images Based on Background Estimation for Optical Character Recognition Purposes

Hubert Michalak
(West Pomeranian University of Technology, Szczecin, Poland
michalak.hubert@zut.edu.pl)

Krzysztof Okarma
(West Pomeranian University of Technology, Szczecin, Poland
okarma@zut.edu.pl)

Abstract: One of the key operations during the image preprocessing step in Optical Character Recognition (OCR) algorithms is image binarization. Although for uniformly illuminated images, obtained typically by flatbed scanners, the use of a single global threshold may be sufficient for further recognition of individual characters, it cannot be applied directly in case of non-uniform lightened document images. Such problem may occur during capturing photos of documents in unknown lighting conditions making a proper text recognition impossible in some parts of the image.

Since the application of popular adaptive thresholding methods, e.g. Niblack, Sauvola and their modifications, based on the analysis of the neighbourhood of each pixel is time consuming, a faster solution might be the division of images into blocks or elimination of non-uniform background. Such an approach can be considered as a balance solution filling the gap between global and local adaptive thresholding. The solution proposed in the paper, useful also for various mobile devices due to limited computational requirements, is based on the approximation of lighting distribution of the background using the reduced resolution images. The proposed method allows to obtain very good OCR results being superior in comparison to typical adaptive binarization algorithms both in terms of the resulting OCR accuracy and computational efficiency.

Key Words: binarization, OCR, document image analysis
Category: I.4.6, I.4.10, I.7.5

1 Introduction

A fast and accurate binarization of natural images acquired in unknown lighting conditions is an important issue in many applications starting from optical text recognition from images acquired by a smartphone camera, or recognition of QR and other 2D binary codes to Visual SLAM (Simultaneous Localization and Mapping) applications for mobile robotics. Apart from the line followers based on image analysis and navigation of autonomous robots in dark corridors similar solutions can be applied in modern "intelligent" vehicles equipped with ADAS (Advanced Driver-Assistance Systems).

The most typical applications, illustrating the importance of the proper image binarization as the preprocessing step, are related to document image analysis. Since this step determines the success of further shape classification and recognition of individual characters, it should be performed as accurately as possible. On the other hand the application of sophisticated algorithms might be troublesome, especially in some real-time video analysis applications. As the main computational resources in the OCR systems are usually devoted to the analysis and recognition of shapes, the initial thresholding should be made in much shorter time. It may be especially important in embedded systems with limited computational power and reduced amount of memory.

Assuming the controlled lighting conditions, which can be ensured in 2D flatbed scanners, the simplified global thresholding can be efficiently applied e.g. using the most popular Otsu's method [Otsu, 1979]. A similar situation may take place using some dedicated illuminators often used in machine vision systems for automation and robotics purposes to ensure the uniform illumination of objects subjected to further shape analysis. Nevertheless, for natural images acquired in unknown lighting conditions some more advanced methods should be applied for compensation of the influence of directional light. Typically used adaptive thresholding algorithms, such as e.g. used in MATLAB's *adapttresh* function [Bradley and Roth, 2007], are computationally demanding due to the necessity of analysis of each pixel's neighbourhood allowing the calculations of the local thresholds.

The simplest adaptive approach proposed by Niblack [Niblack, 1986] is based on the dependence of the local threshold on the mean intensity and variance of the small fragment of the image. A bit more complicated formula proposed by Sauvola [Sauvola and Pietikäinen, 2000] utilizes the local mean value and the local standard deviation in the pixel's neighborhood. Another similar method proposed by Bernsen [Bernsen, 1986] is based on the average of the minimum and maximum values within the window surrounding the analysed pixel whereas Wolf's method [Wolf and Jolion, 2004] assumes the maximization of the local contrast. Some other modifications have been developed by Feng [Feng and Tan, 2004] and Gatos [Gatos et al., 2006].

Some other attempts to adaptive thresholding based on local calculation of Otsu's threshold have been proposed by Moghaddam [Moghaddam and Cheriet, 2012] and Wen [Wen et al., 2013]. The first one, known as AdOtsu, utilizes additional background estimation and analysis of the neighbourhood of each pixel, whereas the second is based on Curvelet transform. An interesting modification proposed by Chou [Chou et al., 2010] considers the local Otsu's thresholding with additional use of Support Vector Machines for background regions. Another algorithm has been proposed by Su [Su et al., 2013] where the local adaptive contrast is used with computationally expensive edge filtering and post-processing.

A method of colour image binarization has been proposed by Mysore [Mysore et al., 2016], whereas methods based on morphological image processing have been investigated by some other researchers [Erol et al., 2008; Lu et al., 2010; Okamoto et al., 2013], as well as the application of local features and Gaussian mixture modelling [Mitianoudis and Papamarkos, 2015]. Some recent papers employ time-consuming median filtering for additional background estimation [Khitas et al., 2018], non-local means [Chen and Wang, 2017] and variational model [Feng, 2019] for noisy images, as well as deep convolutional neural networks [Tensmeyer and Martinez, 2017]. Due to high computational demands of many recently proposed methods, they have not been considered in experimental comparisons presented in Section 5.

To reduce resources needed for computations, some simplified binarization methods utilising the Monte Carlo method, based on the approximated histograms, have been proposed recently [Lech and Okarma, 2014].

An overview of state-of-the-art algorithms can be found in some survey papers published by Leedham [Leedham et al., 2003], Khurshid [Khurshid et al., 2009] as well as recently by Samorodova [Samorodova and Samorodov, 2016], Shrivastava [Shrivastava and Srivastava, 2016], Mustafa [Mustafa and Kader, 2018] and Saxena [Saxena, 2017].

A simple method useful for the validation of document binarization algorithms has been proposed by Stathis [Stathis et al., 2008], however some more advanced approaches have also been presented by Lu [Lu et al., 2004] – known as Distance-Reciprocal Distortion (DRD) – as well as Misclassification Penalty Metric (MPM) introduced by Young [Young and Ferryman, 2005]. Nevertheless, due to the assumed application for the OCR systems, in this paper we have focused on the "classical" F-Measure (calculated for characters instead of single pixels) and Levenshtein distance.

Another attempt is the fast region based approach proposed for the use with low quality documents and unevenly illuminated images [Michalak and Okarma, 2018, 2019]. The local value of the threshold is calculated assuming the 64×64 pixels blocks as 95% of the average intensity lowered by 7 for the dynamic range from 0 to 255. Results achieved for widely used DIBCO datasets [Pratikakis et al., 2018] and some additional non-uniformly lightened documents being the OCR input have been encouraging. Nonetheless, DIBCO benchmarking datasets have not been used for comparisons in the paper due to the assumed use of binarization methods as the preprocessing step for text recognition purposes.

The rest of the paper contains the description of the proposed approach presented in Section 2 and its verification (Section 3) followed by the analysis of the impact of parameters discussed in Section 4. The last part of the paper consists of the analysis of results (Section 5) and final conclusions.

2 Description of the Proposed Method

Due to the necessity of time consuming analysis of each pixel's neighbourhood even in the simplest adaptive binarization methods, the development of some more efficient approaches would be desired. Therefore, the method proposed in the paper is based on the assumption of possibly low computational requirements preventing the relatively good OCR accuracy. To overcome the limitations of pixel based local thresholding, the background suppression approach has been investigated.

The first step of the algorithm is related to image downsampling where one of well known interpolation methods may be applied. For this purpose MATLAB's *imresize* function has been used with one bilinear or bicubic interpolation and the simple nearest neighbour method. Additionally two versions of Lanczos kernels have been used, based on *sinc* function with the parameter α equal to 2 or 3.

The result of application of relatively large kernel during downsizing of the image is the loss of details related to shapes of individual characters. Therefore only the low frequency image data is preserved representing the overall distribution of brightness, being in fact mainly the downsampled background information. Nevertheless, a relevant parameter is the scaling factor being the size of the window aggregated into a single pixel with its optimal size determined experimentally as discussed later. After resizing back the downsampled image to the original resolution using the same kernel, the image containing only the low frequency information is obtained, representing the approximated high resolution background.

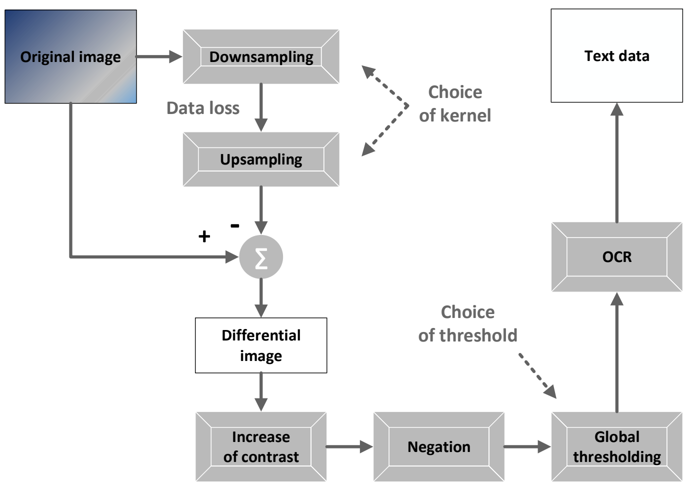


Figure 1: The simplified flowchart of the proposed method

In the next step of the proposed method the subtraction of this image from the original is made to enhance the text data, followed by simple contrast increase and logical negation. Such obtained image can be successfully subjected to fast global thresholding e.g. using Otsu's method. Nevertheless, to speed up the computations the fixed threshold might be used, determined experimentally to maximize the final OCR accuracy as shown in our experimental results. For some of the images even the simplified choice of threshold 0.5 has led to the results very similar to the application of Otsu's thresholding in terms of the OCR accuracy.

The illustration of the simplified flowchart of the proposed approach is illustrated in Fig. 1 whereas the results of consecutive steps of the algorithm for an exemplary image are presented in Fig. 2.

3 Experimental Verification

For the verification of the proposed approach, as well as its comparison to some other popular image binarization methods, a database of unevenly illuminated images has been prepared. This dataset contains 140 document images captured by a standalone camera in various lighting conditions in the presence of some directional light sources located in various positions over the document.

All the images contain the same well known "Lorem ipsum" text generated using a dedicated website[1], consisting of 3788 characters (with spaces), 563 words and 6 paragraphs. Due to the location of the camera directly over the central part of documents and relatively high quality optical system of the camera, the influence of geometrical distortions has been omitted as negligibly small. The documents have been printed using 5 different popular font shapes (Arial, Times New Roman, Calibri, Courier and Verdana) with some typical modifications of attributes (normal, bold and italic versions of all fonts as well as bold italics).

As could be observed during initial experiments, obtained results strongly depend on the kernel size and type used during resizing. Since the correct result of text recognition is known in advance it can be considered as "ground truth" text data. Therefore the verification of the binarization method's influence on the OCR accuracy can be performed calculating the F-Measure considering the number of correctly and incorrectly recognized characters as well as Levenshtein distance defined as the minimum number of text edits (insertions, deleting or replacements of individual characters) required to convert the recognized text into the "ground truth". The F-Measure, known also as F1-score, used often for classification evaluation, is defined as:

$$\mathrm{FM} = 2 \cdot \frac{PR \cdot RC}{PR + RC} , \tag{1}$$

[1] pl.lipsum.com

where Precision (PR) is computed as the ratio of true positives to the sum of all positives, and Recall (RC) is defined as the ratio of true positives to the sum of true positives and false negatives.

Although the most typical application of F-Measure for the binarization quality assessment is related to the comparison of single pixels of binary images [Ntirogiannis et al., 2013], it may also be considered as general classification quality metric, useful also for text recognition purposes at the characters level instead of pixels.

For a reliable verification of the proposed method and optimization of parameters both metrics have been calculated for different types and size of interpolation kernels during the preprocessing step, preceding binarization and launching the OCR software. All the experiments have been made separately for different fonts and aggregated for optimization purposes. The last stage of the algorithm related to the text recognition has been implemented by the use of the free Tesseract OCR engine initially invented by Hewlett Packard and University of Nevada and further developed by Google. Calculation times presented in respective tables and figures have been calculated as the average values obtained by several independent executions of consecutive in the same environment.

4 Influence of Parameters

Since the proposed approach may be strongly dependent on some parameters such as the choice interpolation method, kernel size and the method of global binarization, several experiments have been made for the whole prepared dataset to determine the most suitable set of them preserving the balance between relatively low computation time and high final OCR accuracy.

The first experiments have been made for different interpolation methods as well as for various downsampling/upsampling factors determining the kernel size. Comparing the results obtained for five interpolation methods: nearest neighbour, bilinear, bicubic and two Lanczos kernels (Lanczos2 denotes the kernel obtained for $\alpha = 2$ and Lanczos3 stands for $\alpha = 3$ respectively), the best results have been achieved for Lanczos and bicubic interpolations. Nevertheless, the use of bilinear interpolation leads to only slightly worse results with noticeably faster computations and therefore it has been chosen as the basic one for further experiments. The comparison of F-Measure, Levenshtein distance and average computation time obtained for the whole dataset, assuming the same kernel size (obtained using the scale factor 32 during downsampling and upsampling operations) and fixed global threshold (0.75), is shown in Fig. 3.

The next experiments have been related to the determination of the appropriate kernel size assuming the use of bilinear interpolation and the fixed binarization threshold 0.75. Analysing the results shown in Fig. 4 it can be noted that

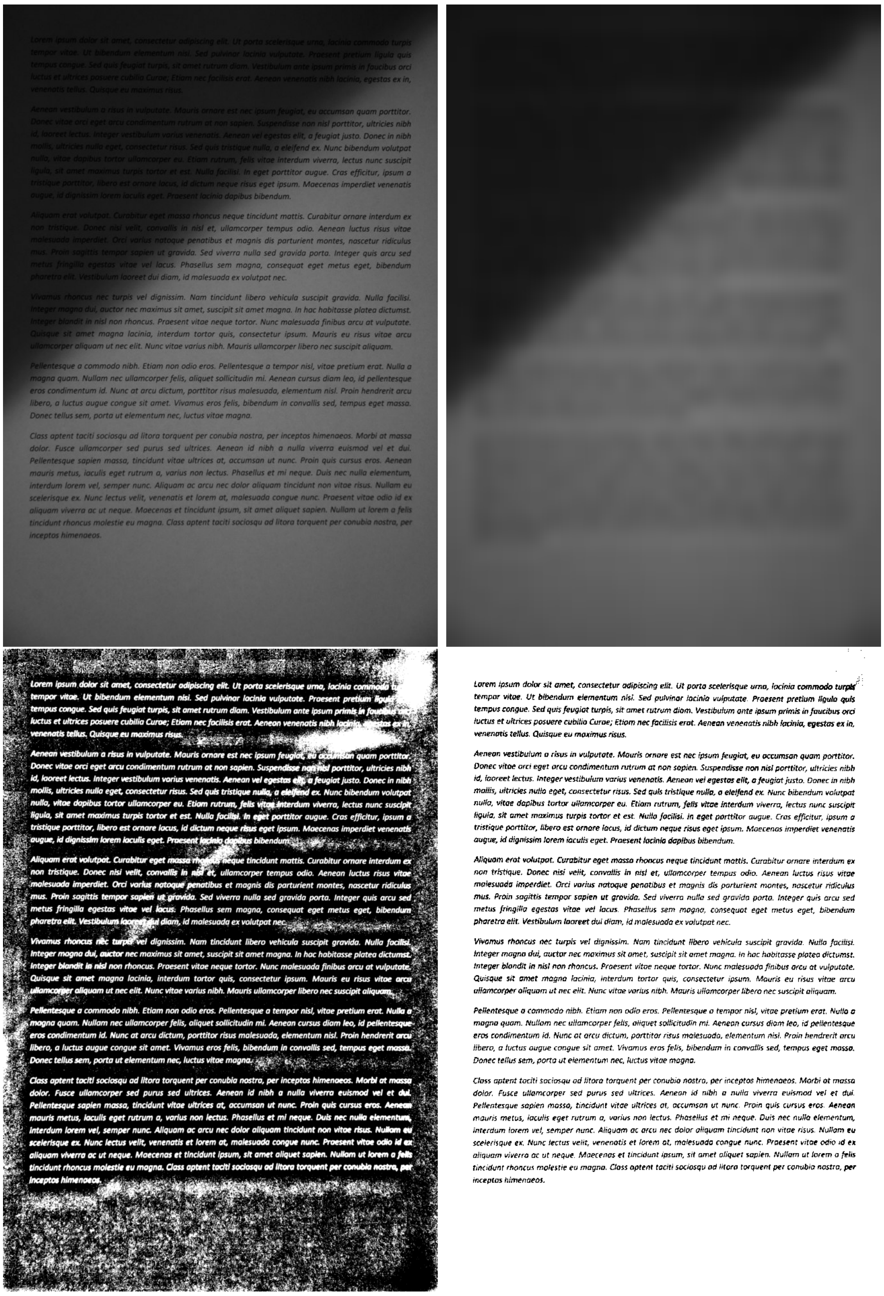


Figure 2: Results obtained for consecutive stages of the proposed algorithm - from left: input image and approximated background (top), result of subtraction before contrast increasing and final binary image (bottom)

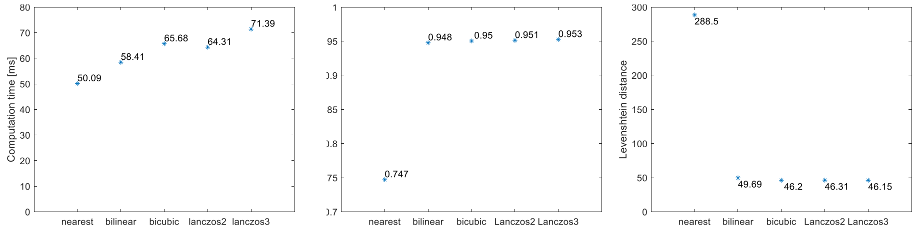


Figure 3: Comparison of the OCR accuracy and computation time obtained using five different interpolation kernels

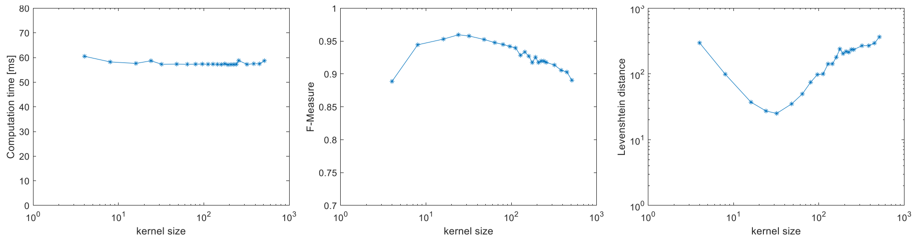


Figure 4: Comparison of the OCR accuracy and computation time obtained using various kernel sizes/scale factors for the fixed threshold of 0.75

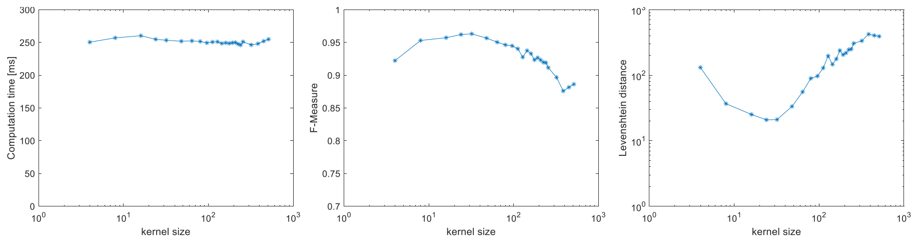


Figure 5: Comparison of the OCR accuracy and computation time obtained using various kernel sizes/scale factors for Otsu thresholding

the computation time is generally independent on the kernel size and therefore its choice should be dependent only on the F-Measure and Levenshtein distance values. As can be observed the best results can be obtained for the scale factors 24 and 32 (FM = 0.9596 and Levenshtein distance = 25.1 respectively).

To examine the influence of different parameters for the other choices of global thresholding the next calculations have been conducted assuming Otsu's binarization. Nevertheless, in all cases the conclusions related to the optimal choice of individual parameters for the other reasonable selections ensuring good OCR accuracy have been the same. Analysing the results provided in Fig. 5 much longer computation time can be easily noticed in comparison for fixed threshold (about 250 ms for Otsu against the average of 57 ms as shown in Fig. 4). Significantly shorter processing time can also be noticed in Fig. 3. Regardless of computational resources, the best OCR accuracy has been obtained again for the scale factor 24 and 32 (Levenshtein distance = 20.79 and FM = 0.9631 respectively). In this case slightly better F-Measure results have been obtained for the scale factor 32 and a little smaller Levenshtein distance for the scale coefficient 24 - reversely as in the previous case. However, these results are insignificantly better than obtained for the fixed threshold, hence the application of time consuming Otsu's thresholding in the last stage of the proposed algorithm is not recommended due to the increase of computing requirements.

Since the resolution of document images used in experiments has been set as 1940×2872 pixels and the number of text lines has been changing from about 40 to about 55 depending on the font shapes, it can be noticed that the most appropriate scale factors are always smaller than the number of text lines. However, results obtained for the scale factors smaller than about 50% of the number of text lines are significantly worse.

Having optimized the kernel size, the OCR results obtained using the proposed binarization method with the fixed threshold should be compared with the application of some other global binarization algorithms applied in the last stage. For comparison purposes a simple mean thresholding has been chosen as well as Otsu's method and a similar approach based on the maximization of the entropy of the thresholded image proposed by Kapur [Kapur et al., 1985]. The results obtained for various thresholding methods are shown in Fig. 6. As can be easily observed, the choice of the fixed threshold 0.75 combines the short calculation time with high OCR accuracy, comparable with much more computationally demanding Otsu thresholding.

One of the last questions is the motivation of the choice of the fixed thresh-

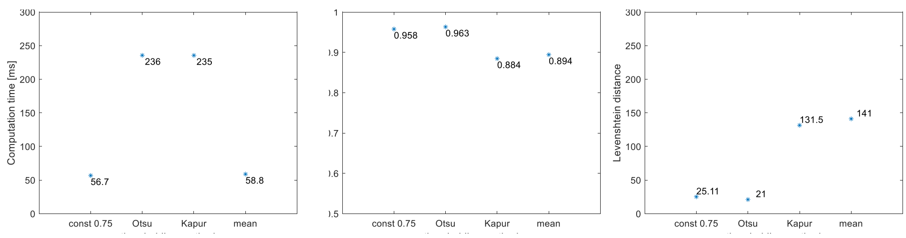


Figure 6: Comparison of the OCR accuracy and computation time obtained using various global thresholding methods

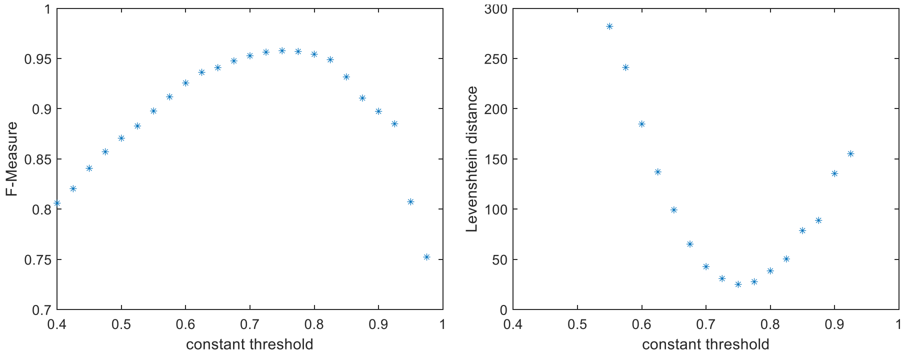


Figure 7: Comparison of the OCR accuracy obtained using different fixed global threshold's values

old's value assumed as equal to 0.75. To verify its appropriateness several experiments have been made for various interpolation kernels. As the optimal threshold's value has turned out to be independent on the chosen kernel and interpolation method, the representative results obtained for the scale factor 64 and bilinear interpolation kernel are presented in Fig. 7. As expected, the influence of the chosen threshold on the computation time is negligible, hence these values are not presented. As can be seen, any changes of the threshold value cause the decrease of the OCR accuracy both in terms of F-Measure and Levenshtein distance (the best values obtained for the threshold 0.75 are FM = 0.9577 and Levenshtein distance equal to 25.11).

5 Comparative Analysis of Results

After the tuning of the parameters of the proposed methods, the comparison of its speed and the final OCR accuracy obtained after the application of Tesseract engine has been made. To illustrate additionally the advantages of the proposed approach, the execution time of each method has also been compared using the same environment (PC with Intel i7 CPU at 2.8 GHz and 16 GB of RAM with MATLAB 2018b for 64-bit Windows 10), although our implementation has not been optimised in terms of execution speed. The region based method indicated in Tables 1 and 2 has been described in [Michalak and Okarma, 2018].

The OCR quality measures (Levenshtein distance and F-Measure), together with the running time of various binarization algorithms, averaged for all 140 images are presented in Table 1, whereas Table 2 shows the results obtained for uniformly illuminated document images. Results better than the best of the previously developed methods are marked with bold fonts. Since the OCR accuracy depends also on the font shapes and their modifications, the OCR

Table 1: Average OCR results and execution times (with standard deviations) obtained using various binarization methods for all 140 images

Binarization method	F-Measure	Levenshtein distance	Calculation time [ms]
None	0.7291 (0.1392)	1299.9 (634.2)	0.15 (0.04)
Otsu (global)	0.7356 (0.1255)	1280.4 (619.7)	8.14 (0.4)
Chou	0.8130 (0.0914)	935.1 (546.0)	30.3 (0.4)
Region based	0.8838 (0.0735)	212.4 (230.2)	110.5 (3.7)
Niblack	0.8760 (0.0681)	284.5 (277.1)	446.4 (4.0)
Sauvola	0.9387 (0.0736)	109.0 (242.7)	438.1 (9.0)
Wolf	0.9298 (0.0819)	174.3 (340.1)	452.7 (5.8)
Bradley (mean)	0.8925 (0.0807)	296.2 (420.4)	112.1 (1.5)
Bradley (Gaussian)	0.8384 (0.0917)	644.4 (503.9)	1195.4 (11.0)
Feng	0.7283 (0.1815)	1057.7 (817.6)	1340.8 (95.3)
Bernsen	0.7646 (0.1183)	773.5 (651.8)	1231.4 (114.5)
Mean thresh	0.8091 (0.1033)	511.0 (354.0)	245.4 (16.6)
Proposed			
scale=24, bilinear, 0.75	**0.9596** (0.0305)	**27.36** (40.35)	72.7 (1.1)
scale=24, bilinear, Otsu	**0.9620** (0.0280)	**20.79** (28.04)	278.4 (14.8)
scale=24, Lanczos3, 0.75	**0.9567** (0.0343)	**30.09** (49.44)	87.7 (1.8)
scale=24, Lanczos3, Otsu	**0.9593** (0.0255)	**21.55** (19.08)	298.7 (13.9)
scale=32, bilinear, 0.75	**0.9578** (0.0297)	**25.11** (29.61)	70.6 (0.9)
scale=32, bilinear, Otsu	**0.9631** (0.0260)	**21.00** (27.62)	273.6 (14.3)
scale=32, Lanczos3, 0.75	**0.9594** (0.0284)	**24.52** (34.04)	87.2 (1.3)
scale=32, Lanczos3, Otsu	**0.9643** (0.0224)	**20.31** (31.37)	295.2 (15.7)

accuracy metrics obtained for them are shown separately in Table 3 where much worse results for Courier font can be easily noticed. The best OCR accuracy have been obtained for simple rounded font shapes such as Arial and Verdana. Serif fonts, such as Times New Roman, are usually harder for text recognition but surprisingly worse results have been obtained for Calibri fonts. According to expectations, higher recognition accuracy has been achieved for bold fonts.

The results presented in Table 3 are the average values of the OCR metrics obtained for Sauvola binarization, selected as the best from previously proposed algorithms, according to Tables 1 and 2, and the proposed methods. The columns marked as "proposed" contain the average values obtained after application of 8 variants of the proposed methods listed in the bottom parts of Tables 1 and 2.

Table 2: Average OCR results and execution times (with standard deviations) obtained using various binarization methods for 20 uniformly lightened images

Binarization method	F-Measure	Levenshtein distance	Calculation time [ms]
None	0.9638 (0.0295)	56.4 (150.2)	0.19 (0.10)
Otsu (global)	0.9614 (0.0393)	62.7 (152.2)	8.3 (0.8)
Chou	0.9635 (0.0224)	21.2 (35.3)	30.5 (0.6)
Region based	0.9670 (0.0203)	18.6 (10.1)	111.2 (3.8)
Niblack	0.9614 (0.0174)	30.5 (30.3)	446.0 (5.0)
Sauvola	0.9709 (0.0094)	20.7 (9.2)	442.3 (22.8)
Wolf	0.9661 (0.0128)	21.5 (15.3)	453.6 (12.0)
Bradley (mean)	0.9665 (0.0145)	26.6 (14.9)	121.6 (1.4)
Bradley (Gaussian)	0.9663 (0.0124)	27.3 (13.0)	1198.3 (8.0)
Feng	0.9109 (0.0329)	66.4 (41.9)	1292.6 (16.9)
Bernsen	0.8473 (0.0043)	191.8 (126.7)	1191.6 (23.4)
Mean thresh	0.9595 (0.0235)	24.9 (23.3)	236.8 (3.4)
Proposed			
scale=24, bilinear, 0.75	0.9592 (0.0390)	22.15 (29.15)	71.4 (1.4)
scale=24, bilinear, Otsu	**0.9732** (0.0148)	**13.15** (8.82)	297.5 (10.9)
scale=24, Lanczos3, 0.75	0.9552 (0.0384)	36.35 (52.21)	85.9 (1.1)
scale=24, Lanczos3, Otsu	0.9621 (0.0194)	22.85 (13.39)	314.8 (8.1)
scale=32, bilinear, 0.75	0.9665 (0.0290)	**18.5** (21.19)	69.3 (0.6)
scale=32, bilinear, Otsu	**0.9738** (0.0092)	**13.45** (6.94)	291.1 (11.9)
scale=32, Lanczos3, 0.75	0.9626 (0.0329)	21.00 (25.68)	85.9 (0.9)
scale=32, Lanczos3, Otsu	**0.9729** (0.0126)	**14.8** (9.26)	315.2 (12.2)

As can be easily observed, the proposed approach is much faster than the popular adaptive methods and comparable to recently proposed region based approach [Michalak and Okarma, 2018], leading also to better OCR results. Analysing the relatively low values of standard deviations presented in parentheses in Tables 1 and 2, the results obtained for the proposed methods may be considered as more stable, particularly for unevenly illuminated images.

A visual comparison of binarization results achieved after the application of various methods for an exemplary non-uniformly illuminated image is shown in Figures 8 and 9. A similar comparison for the most demanding Courier font shape is illustrated in Figures 10 and 11. where the advantages of the proposed approach are clearly visible.

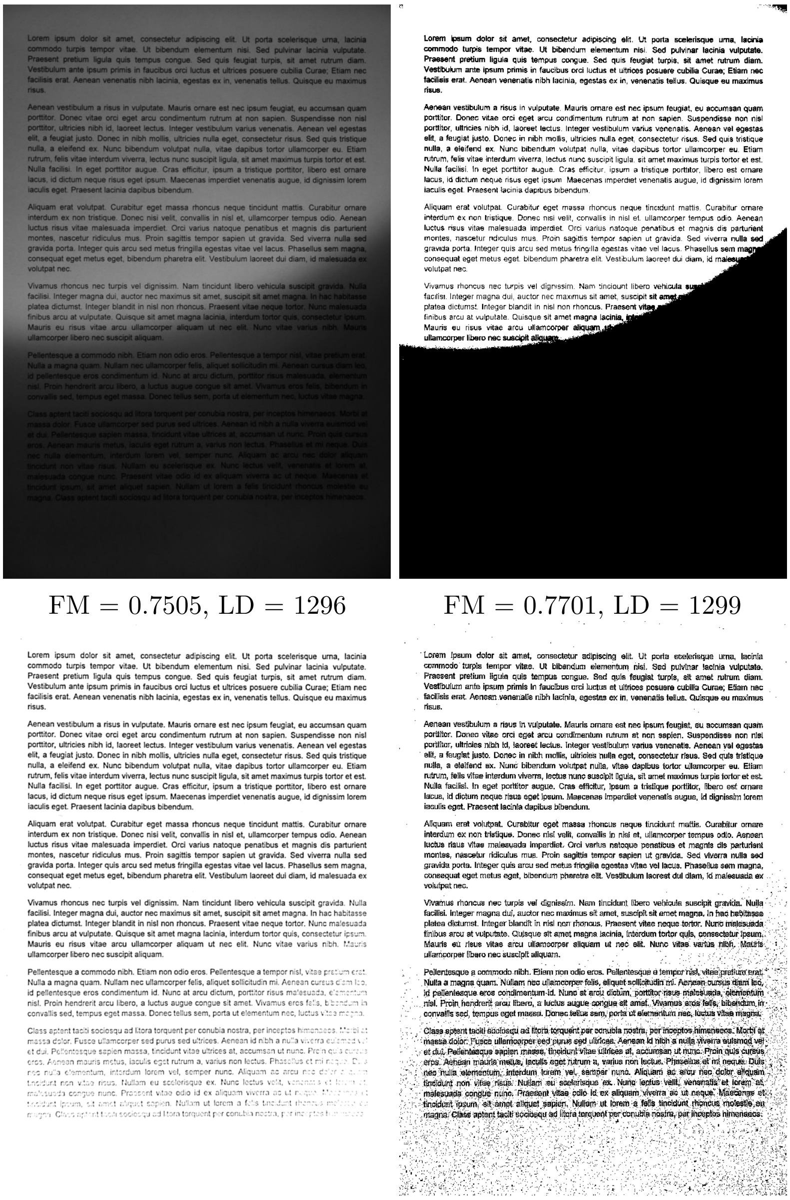


Figure 8: Binarization results obtained using various methods for exemplary unevenly illuminated image - from left: input image and global Otsu algorithm (top), Niblack and Feng methods (bottom), together with F-Measure values and Levenshein distances obtained after applying the OCR procedure

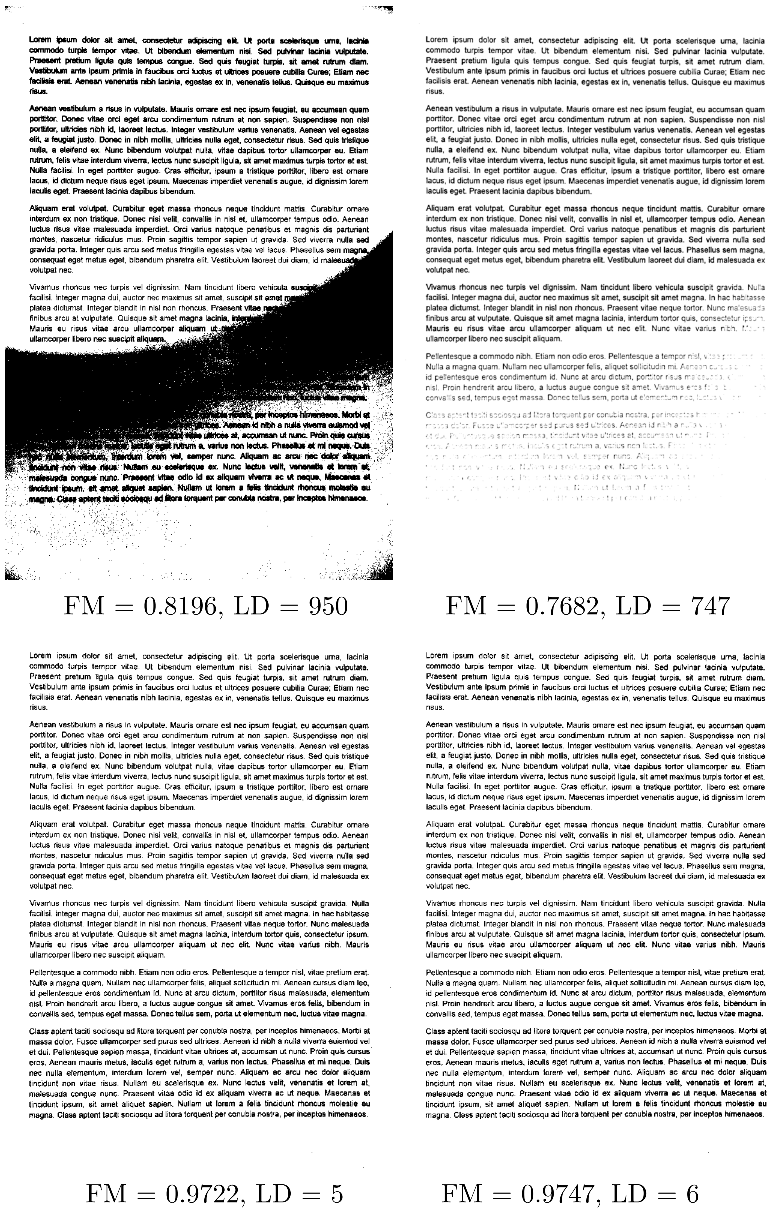
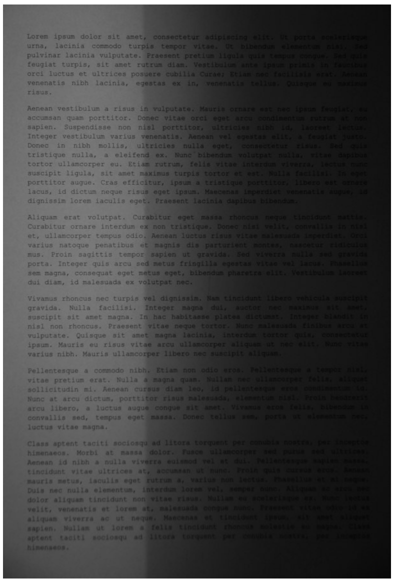
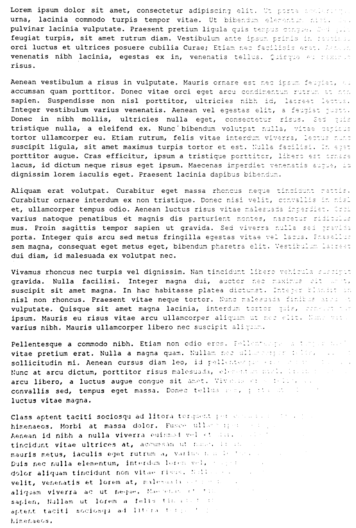
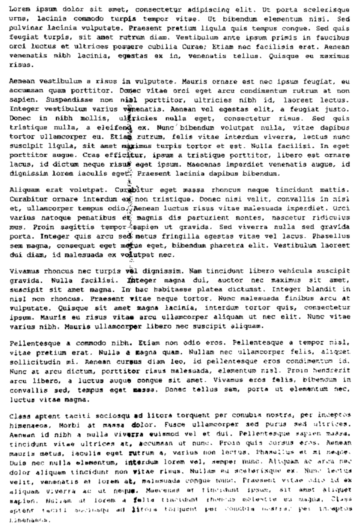
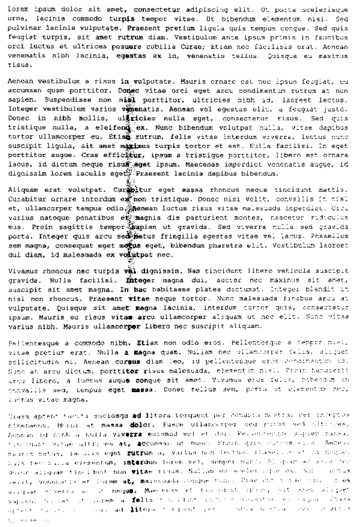


Figure 9: Binarization results obtained using various methods for exemplary unevenly illuminated image - from left: Bradley (Gaussian) algorithm and mean thresholding method (top), proposed approach for 24-fold bilinear interpolation with 0.75 threshold and 32-fold Lanczos3 kernel with Otsu threshold (bottom), together with F-Measure values and Levenshein distances obtained after applying the OCR procedure

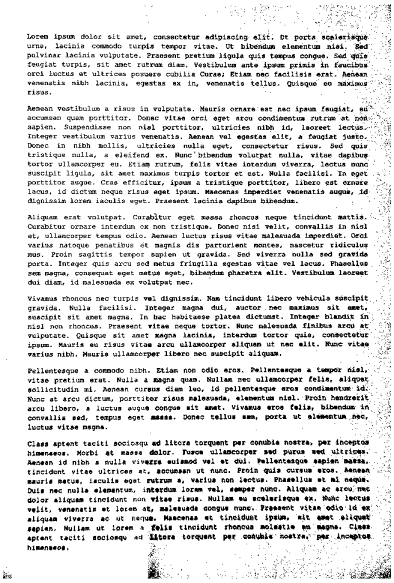



FM = 0.2031, LD = 2776 FM = 0.7457, LD = 1209

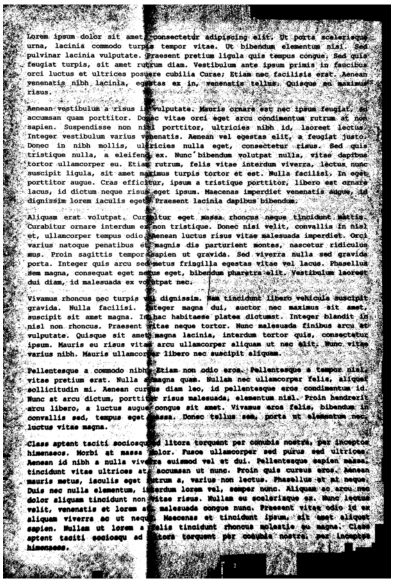



FM = 0.8525, LD = 90 FM = 0.7857, LD = 437

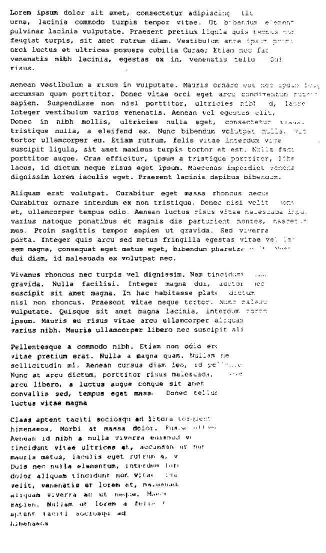
Figure 10: Binarization results obtained using various methods for Courier font shape - from left: input image and Niblack algorithm (top), Sauvola and Wolf methods (bottom), together with F-Measure values and Levenshein distances obtained after applying the OCR procedure



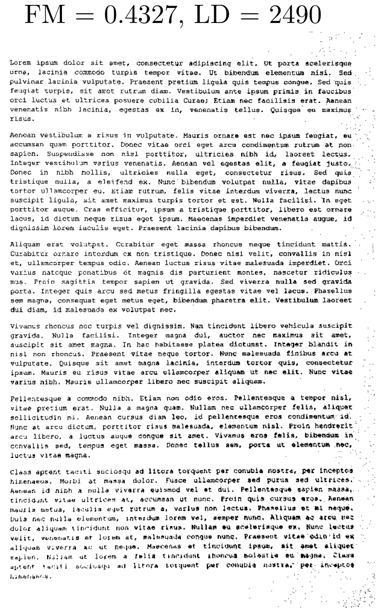
FM = 0.8525, LD = 124



FM = 0.4327, LD = 2490



FM = 0.7375, LD = 974



FM = 0.9473, LD = 62

Figure 11: Binarization results obtained using various methods for Courier font shape - from left: Bradley (mean) algorithm and Bernsen method (top), region based method and proposed approach with 24-fold bilinear interpolation and Otsu threshold (bottom), together with F-Measure values and Levenshein distances obtained after applying the OCR procedure

Table 3: Average OCR results obtained using Sauvola and proposed binarization methods for different fonts

Font shape	F-Measure		Levenshtein distance	
	Sauvola	proposed	Sauvola	proposed
Arial	0.9396	0.9644	93.68	16.04
Times New Roman	0.9476	0.9731	115.04	14.79
Calibri	0.9413	0.9618	89.64	15.97
Courier	0.9245	0.9358	120.00	59.13
Verdana	0.9402	0.9661	126.54	13.28
Font modification	**F-Measure**		**Levenshtein distance**	
	Sauvola	proposed	Sauvola	proposed
none	0.9366	0.9546	112.94	31.17
bold	0.9377	0.9671	96.14	17.49
italic	0.9351	0.9575	127.86	26.06
bold + italic	0.9456	0.9620	98.97	20.65

6 Conclusions

The proposed fast method of document image binarization can be efficiently applied for text recognition purposes from unevenly lightened document images. Presented approach combines short execution time with good OCR accuracy obtained after binarization and outperforms both recently proposed region based method and popular adaptive thresholding algorithms. Due to its high speed and good recognition results, one of potential areas of applications of the proposed method is preprocessing of document images captured by smartphone cameras. Hence, the proposed approach is much faster than many state-of-the-art methods employing more sophisticated methods, e.g. based on deep learning [Tensmeyer and Martinez, 2017]. The use of local features and Gaussian Mixtures, according to details presented in the paper [Mitianoudis and Papamarkos, 2015], requires an average of 7.7 sec for a 640×480 pixels DIBCO image, therefore being much slower in comparison to our approach, leading to comparable results.

Comparing two possibilities of using the fixed threshold and Otsu's global binarization after unevenly illuminated background elimination, the OCR accuracy and processing speed may be balanced depending on specific implementation. Application of Otsu's method allows to obtain the best OCR results outperforming the other methods with higher processing speed than achieved by Niblack or Sauvola methods. On the other hand the fastest (apart from purely global Otsu's thresholding) versions of the proposed method utilizing the bilinear interpolation and the fixed threshold still allow to achieve better OCR accuracy than the other methods - much faster than the simplest version of Bradley's method.

One of the directions of our future research may be the investigation of the relation between the optimal scale factors and the resolution of document images, in particular considering the font size and shape together with the number of text lines as well as their height and interline spacing. Another investigated issue will be the relation of the letter size and the size of the filter employed for subsampling and therefore an extension and additional verification of the proposed approach is planned, also using some other databases [Lins et al., 2017]. Nonetheless, they will require some modifications as they contain mainly evenly illuminated document images with some other types of visible distortions.

References

[Bernsen, 1986] Bernsen, J.: Dynamic thresholding of grey-level images. Proc. 8th Int. Conf. on Pattern Recognition (ICPR), pp. 1251–1255, (1986).

[Bradley and Roth, 2007] Bradley, D., Roth, G.: Adaptive thresholding using the integral image. Journal of Graphics Tools 12(2):13–21, (2007).

[Chen and Wang, 2017] Chen, Y., Wang, L.: Broken and degraded document images binarization. Neurocomputing 237:272–280, (2017).

[Chou et al., 2010] Chou, C.-H., Lin, W.-H., Chang, F.: A binarization method with learning-built rules for document images produced by cameras. Pattern Recognition 43(4):1518–1530, (2010).

[Erol et al., 2008] Erol, B., Antúnez, E., Hull. J.: HOTPAPER: multimedia interaction with paper using mobile phones. Proc. 16th Int. Conf. on Multimedia, pp. 399–408, (2008).

[Feng and Tan, 2004] Feng, M.-L., Tan, Y.-P.: Adaptive binarization method for document image analysis. Proc. 2004 IEEE Int. Conf. on Multimedia and Expo (ICME), vol. 1, pp. 339–342, (2004).

[Feng, 2019] Feng, S.: A novel variational model for noise robust document image binarization. Neurocomputing, 325:288–302, (2019).

[Gatos et al., 2006] Gatos, B., Pratikakis, I., Perantonis, S.: Adaptive degraded document image binarization. Pattern Recognition 39(3):317–327, (2006).

[Kapur et al., 1985] Kapur, N., Sahoo, P.K., Wong, A.K.: A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision Graphics and Image Processing, 29:273–285, (1985).

[Khitas et al., 2018] Khitas, M., Ziet, L., Bouguezel, S.: Improved degraded document image binarization using median filter for background estimation. Elektronika ir Elektrotechnika, 24(3):82–87, (2018).

[Khurshid et al., 2009] Khurshid, K., Siddiqi, I., Faure, C., Vincent, N.: Comparison of Niblack inspired binarization methods for ancient documents. Document Recognition and Retrieval XVI, Proc. SPIE, vol. 7247, pp. 7247–7247–9, (2009).

[Lech and Okarma, 2014] Lech, P., Okarma, K.: Optimization of the fast image binarization method based on the Monte Carlo approach. Elektronika Ir Elektrotechnika 20(4):63–66, (2014).

[Leedham et al., 2003] Leedham, G., Yan, C., Takru, K., Tan, J. H. N., Mian, L.: Comparison of some thresholding algorithms for text/background segmentation in difficult document images. Proc. 7th Int. Conf. on Document Analysis and Recognition (ICDAR), pp. 859–864, (2003).

[Lins et al., 2017] Lins, R.D., de Almeida, M.M., Bernardino, R.B., Jesus, D., Oliveira, J.M.: Assessing binarization techniques for document images. Proc. ACM Symposium on Document Engineering (DocEng), pp. 183–192, (2017).

[Lu et al., 2004] Lu, H., Kot, A.C., Shi, Y.Q.: Distance-reciprocal distortion measure for binary document images. IEEE Signal Processing Letters 11(2):228–231, (2004).

[Lu et al., 2010] Lu, S., Su, B., Tan, C.L.: Document image binarization using background estimation and stroke edges. International Journal on Document Analysis and Recognition (IJDAR) 13(2):303–314, (2010).

[Michalak and Okarma, 2018] Michalak, H., Okarma, K.: Region based adaptive binarization for Optical Character Recognition purposes. Proc. 2018 International Interdisciplinary PhD Workshop (IIPhDW), pp. 361–366, (2018).

[Michalak and Okarma, 2019] Michalak, H., Okarma, K.: Fast adaptive image binarization using the region based approach. Artificial Intelligence and Algorithms in Intelligent Systems, AISC, vol. 764, pp. 79–90. Springer International Publishing, (2019).

[Mitianoudis and Papamarkos, 2015] Mitianoudis, N., Papamarkos, N.: Document image binarization using local features and Gaussian mixture modeling. Image and Vision Computing 38:33–51, (2015).

[Moghaddam and Cheriet, 2012] Moghaddam, R.F., Cheriet, M.: AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization. Pattern Recognition 45(6):2419–2431, (2012).

[Mustafa and Kader, 2018] Mustafa, W.A., Kader, M.M.M.A.: Binarization of document images: a comprehensive review. Journal of Physics: Conference Series 1019, 012023, (2018).

[Mysore et al., 2016] Mysore, S., Gupta, M.K., Belhe, S.: Complex and degraded color document image binarization. Proc. 3rd Int. Conf. on Signal Processing and Integrated Networks (SPIN), pp. 157–162, (2016).

[Niblack, 1986] Niblack, W.: An introduction to digital image processing. Prentice Hall, Englewood Cliffs, (1986).

[Ntirogiannis et al., 2013] Ntirogiannis, K., Gatos, B., Pratikakis, I.: Performance evaluation methodology for historical document image binarization. IEEE Transactions on Image Processing 22(2):595–609, (2013).

[Okamoto et al., 2013] Okamoto, A., Yoshida, H., Tanaka, N.: A binarization method for degraded document images with morphological operations. Proc. 13th IAPR Int. Conf. on Machine Vision Applications (MVA), pp. 294–297, (2013).

[Otsu, 1979] Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9(1):62–66, (1979).

[Pratikakis et al., 2018] Pratikakis, I., Zagoris, K., Kaddas, P., Gatos, B.: ICFHR 2018 Competition on Handwritten Document Image Binarization (H-DIBCO 2018). Proc. 16th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), pp. 489–493, (2018).

[Samorodova and Samorodov, 2016] Samorodova, O.A., Samorodov, A.V.: Fast implementation of the Niblack binarization algorithm for microscope image segmentation. Pattern Recognition and Image Analysis 26(3):548–551, (2016).

[Sauvola and Pietikäinen, 2000] Sauvola, J., Pietikäinen, M.: Adaptive document image binarization. Pattern Recognition 33(2):225–236, (2000).

[Saxena, 2017] Saxena, L.P.: Niblack's binarization method and its modifications to real-time applications: a review. Artificial Intelligence Review, pp. 1–33, (2017).

[Shrivastava and Srivastava, 2016] Shrivastava, A., Srivastava, D.K.: A review on pixel-based binarization of gray images. Proc. Int. Congress on Information and Communication Technology (ICICT), AISC, vol. 439, pp. 357–364. Springer Singapore, (2016).

[Stathis et al., 2008] Stathis, P., Kavallieratou, E., Papamarkos, N.: An evaluation technique for binarization algorithms. Journal of Universal Computer Science 14(18):3011–3030, (2008).

[Su et al., 2013] Su, B., Lu, S., Tan, C.L.: Robust document image binarization technique for degraded document images. IEEE Transactions on Image Processing 22(4):1408–1417, (2013).

[Tensmeyer and Martinez, 2017] Tensmeyer, C., Martinez, T.: Document Image Binarization with Fully Convolutional Neural Networks. Proc. 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), pp. 99–104, (2017).

[Wen et al., 2013] Wen, J., Li, S., Sun, J.: A new binarization method for non-uniform illuminated document images. Pattern Recognition 46(6):1670–1690, (2013).

[Wolf and Jolion, 2004] Wolf, C., Jolion, J.-M.: Extraction and recognition of artificial text in multimedia documents. Formal Pattern Analysis & Applications 6(4):309–326, (2004).

[Young and Ferryman, 2005] Young, D.P., Ferryman, J.M.: PETS metrics: on-line performance evaluation service. Proc. 14th Int. Conf. on Computer Communications and Networks (ICCCN), pp. 317–324, (2005).