ISSN 2464-4617 (print)
ISSN 2464-4625 (DVD)

Computer Science Research Notes
CSRN 3001

WSCG2020 Proceedings

# Optimization of Degraded Document Image Binarization Method Based on Background Estimation

Hubert Michalak
West Pomeranian University
of Technology in Szczecin
Aleja Piastów 17
70-310 Szczecin, Poland
michalak.hubert@zut.edu.pl

Krzysztof Okarma
West Pomeranian University
of Technology in Szczecin
Aleja Piastów 17
70-310 Szczecin, Poland
okarma@zut.edu.pl

## ABSTRACT

Binarization of document images is one of the most relevant pre-processing operations, leading to a significant decrease of the amount of information used during their further analysis. Since many document images, particularly historical, may be degraded over time, the application of some simple global thresholding methods usually lead to highly unsatisfactory results. A similar situation may occur for unevenly illuminated images, limiting the visibility of various shapes, representing not only the alphanumerical characters. A typical solution of this problem is the application of some adaptive thresholding methods, as well as more sophisticated solutions, proposed recently e.g. during Document Image Binarization Competitions (DIBCO) or TQ-DIB 2019 competition. Nevertheless, due to their relatively high computational demands, there is still a need of some faster methods, leading to high binarization accuracy for challenging benchmark datasets, such as DIBCO or Nabuco. Hence, the adaptation and optimization of the parameters of the fast thresholding method utilizing background estimation, proposed originally for the OCR purposes and verified for unevenly illuminated printed documents, is presented in this paper. The proposed solution has been optimized and verified using the state-of-the-art datasets containing 166 degraded document images together with their ground-truth binary equivalents, leading to better results, also in comparison to much slower adaptive thresholding methods. The performance of all methods used in comparisons has been determined using commonly accepted metrics, such as F-Measure, Accuracy, Distance Reciprocal Distortion (DRD) or Misclassification Penalty Metric (MPM), and relative execution time, calculated for all used image datasets.

## Keywords

Image thresholding, document images, background estimation, binarization

## 1 INTRODUCTION

Document image binarization is one of the most challenging problems related to various applications of a fast shape analysis and recognition. Although many relatively simple well-known image thresholding methods may be successfully applied in some typical applications, such as Optical Character Recognition (OCR) using well illuminated scanned documents or shape classification of uniformly lightened objects, their usefulness for lower quality images is often significantly limited. In some industrial applications, such as video inspection, dedicated machine vision illuminators may be applied to ensure the uniformity of lighting conditions. On the other hand, assuming the presence of image distortions, affecting the background and character readability in document images, more sophisticated adaptive methods are used, which usually require the analysis of the neighbourhood of each individual pixel to determine the local threshold. In some applications, where the analysis of the document's structure, e.g. text line segmentation, particularly in non-Latin languages [Has19], is important, the appropriate binarization might be a critical step as well. A similar influence may also be observed in some modern solutions based on the application of deep Convolutional Neural Networks (CNN) for various tasks related to image and video analysis [Lef19].

Comparing the influence of non-uniform illumination with distortions typical for degraded historical document images, there is no doubt that the latter ones are more challenging. Hence, newly developed image binarization methods, proposed during recent years, are typically verified using the datasets used in Document Image Binarization Competitions (DIBCO) [Pra17]. Considering the need of balancing the processing speed and obtained accuracy, recently some other datasets have also been proposed, namely Nabuco and LiveMemory, used in Time-Quality Document Binarization Competition organized during the IAPR International Conference on Document Analysis and Recognition – ICDAR 2019 [Lin19]. Since many image binarization methods may be useful in some other applications, e.g. related to automation and robotics, where some devices with low computing performance may be applied, the results of binarization are typically assessed in a general way, using commonly accepted classification metrics usually based on the calculation of the numbers of properly and improperly classified foreground and background pixels.

Considering this, filling the gap between the fast global (e.g. well-known Otsu [Ots79] or Kittler [Kit86] methods) and relatively slow adaptive thresholding algorithms (e.g. proposed by Niblack [Nib86] or Sauvola [Sau00]) is an interesting alternative, also for document image binarization. The approach presented in this paper is based on the estimation of the image background, which may be initially removed to make the final binarization easier. Since the simple background subtraction without additional operations does not lead to satisfactory results, the optimization of some parameters of the proposed method using the state-of-the-art benchmarking datasets is presented, as well as some modifications of the method [Mic19a], initially used for the pre-processing of unevenly illuminated photographs.

The rest of the paper is organized as follows: in Section 2 a short overview of the most popular image binarization methods is provided, whereas Section 3 contains the description of the proposed approach with experimental results provided in Section 4 and discussed in Section 5. Finally, Section 6 concludes the paper.

# 2 OVERVIEW OF POPULAR IMAGE BINARIZATION METHODS

## 2.1 Global and Adaptive Thresholding

The simplest approach to image binarization is the choice of a single threshold, which is applied for all image pixels. Such global thresholding differ by a strategy of choosing the most appropriate threshold. Apart from the use of a fixed threshold (e.g. 50% of the brightness range), the most popular approach, belonging to the family of histogram based methods, is the minimization of the sum of inter-class variances, equivalent to maximization of inter-class variance, proposed by Otsu [Ots79]. A similar idea applied for histogram's entropy replacing the variances has been presented by Kapur *et al.* [Kap85].

Both these approaches work well only for bi-modal histograms assuming uniform lighting conditions, hence a different method of unimodal thresholding has been proposed by Rosin [Ros01], who suggested the use of a simple minimization of the distance between the straight line connecting the peak and the high end of the histogram to determine the threshold. A modified version of such approach, known as T-point method [Cou10], utilizes piecewise linear regression of the descending slope of the histogram, although its usefulness is limited to specific images containing small objects, e.g. electron microscope images, so it is unsuitable for typical document images.

Kittler and Illingworth [Kit86] proposed a global binarization method assuming that the grey levels are normally distributed and may be described by a mixture of two Gaussian functions. The criterion derived in Kittler's method minimizes the classification error probability dividing the histogram into two parts.

A region-based method based on the separate calculation of Otsu's thresholds for image regions as well as some additional features has been proposed by Chou *et al.* [Cho10]. Their values are further used to construct the adaptive decision rules using Support Vector Machines (SVM) assuming the $3 \times 3$ pixels windows. Nevertheless, this method has been dedicated for typical document images, relatively well illuminated, and assessed considering the OCR performance rather than using classification metrics for individual pixels.

An adaptive version of Otsu's method, known as *AdOtsu*, has been proposed by Moghaddam and Cheriet [Mog12], where a multi-layer classifier has been applied in combination with grid-based modelling. Although in this method the background modelling is used, starting from rough double-scale Sauvola's binarization, the application of the multi-scale framework and further post-processing steps additionally increases the computational effort. Nevertheless, a simpler and more popular adaptive method has been proposed by Wayne Niblack [Nib86], who has proposed the use of the difference of the local average brightness, calculated for the $3 \times 3$ pixels sliding window, and its scaled standard deviation (with default multiplier $k = 0.2$) as the local threshold.

Since the results of this method are often noisy, in its modified version proposed by Sauvola and Pietikäinen [Sau00], the dynamic range of standard deviation has been included into Niblack's formula. This approach has been further improved utilizing integral images to speed up the computations [Sha08], leading to ca. 20-fold increase of computational performance. The description of the multi-scale version of Sauvola's method, useful for both small and large objects inside a single document, may be found in Lazzara's paper [Laz14], where the robustness to font size has been the priority, regardless of relatively weak performance for historical documents.

Some other adaptive methods, being the modifications of Niblack's approach, have been proposed by various researchers, leading to some improvements of obtained results. A fast implementation of this method has been presented by Samorodova and Samorodov [Sam16], which utilizes the integral image for the calculation of the local mean values and the second order integral image for the calculation of the local variance. One of the most interesting modifications of Niblack's method has been proposed by Wolf and Jolion [Wol04], where the decision is made in terms of contrast instead of greyscale values, combining the advantages of Sauvola's and Niblack's methods, although two passes are necessary increasing the computational complexity of the method.

Another extension presented by Feng *et al.* [Fen04] requires the pre-processing step based on median filtering and bilinear interpolation of threshold values, calculated only for the centres of the local windows, introducing the additional parameters defined for two local windows used during calculations. Relatively good results may be obtained using the method, known as NICK, proposed by Khurshid *et al.* [Khu09], which assumes the shifting down of the threshold to improve the binarization of brighter documents and the use of $19 \times 19$ pixels local windows. It is worth noticing that the computational complexity of this method is lower in comparison fo Wolf's and Feng's algorithms. A more detailed overview of various Niblack inspired methods may be found in some recent survey papers [Khu09, Sax19, Sul19].

One of the most popular adaptive binarization methods, utilizing the integral images to prevent long calculations, available e.g. in MATLAB environment as the *adaptthresh* function, has been proposed by Bradley and Roth [Bra07]. In this method the local threshold is set as the specified percentage of the average local brightness, although in MATLAB implementation the local median as well as Gaussian weighted mean may also be used. Another well-known method proposed by Bernsen [Ber86] is the calculation of the local threshold as the average of the maximum and minimum intensities inside the square local window, although pixels in low contrast regions are always classified as background. Even simpler approach is the direct use of the local mean as the threshold (known as the *meanthresh* method).

One of the recent approaches [Mic19b] is based on the application of the local entropy filter for image preprocessing, assuming the $19 \times 19$ pixels windows, followed by classical Otsu's thresholding applied for the entropy map. Then, a binary mask subjected to morphological dilation is used for background estimation and removal. The last step of the method is the binarization, which may be conducted using well-known thresholding methods. Another recent idea is the application of multi-layered stack of regions [Mic19c] for the OCR purposes assuming the non-uniform illumination of document images, where the local threshold is calculated as the average of thresholds obtained for the regions "covering" the pixel using a formula based on the modified average intensity in the image region.

Considering that many methods proposed in recent years are computationally demanding, the paper is focused on relatively fast methods, possible to implement efficiently e.g. in mobile devices or embedded systems. Hence, a detailed description of some specialized methods, e.g. based on deep learning [Ten17, Vo18] is skipped in this paper.

## 2.2 Performance Evaluation

Each of newly proposed image binarization method should be verified experimentally using widely accepted benchmarking datasets as well as appropriate evaluation methods [Nti13]. The most typical performance metrics used for evaluation of binarization methods come from machine learning and classification assessment and are based on the calculation of the number of pixels classified properly as foreground (true positives – TP) or background (true negatives – TN). Incorrectly classified pixels are denoted as false positives (FP) or false negatives (FN), respectively. For document images white pixels, usually represented by "ones" in binary images, are considered as background (negatives) and therefore, an additional image negation is often applied before the calculations. Since all such metrics are based on the comparison of resulting images with "ground truth" (GT), such binary images have to be provided in each dataset.

The most typically used evaluation metrics are:

- precision – defined as the ratio of true positives to all positives: $PR = \frac{TP}{TP+TN}$,
- sensitivity / recall – defined as the ratio of true positives to condition positives (sum of true positives and false negatives): $RC = \frac{TP}{TP+FN}$,
- specificity / selectivity – defined as the ratio of true negatives to condition negatives (sum of false positives and true negatives): $SPEC = \frac{TN}{FP+TN}$,
- accuracy – defined as the sum of true positives and true negatives divided by all samples (pixels): $ACC = \frac{TP+TN}{TP+TN+FP+FN}$,
- F-Measure / F1-score – defined as the harmonic mean of precision and recall: $FM = 2 \cdot \frac{PR \cdot RC}{PR+RC}$, usually multiplied by 100%.

Using the above values derived from the confusion matrix, some additional metrics, such as e.g. balanced accuracy (used for imbalanced data) or Matthews correlation coefficient (MCC), may be calculated, similarly as Receiver Operating Characteristics (ROC) graphs. Nevertheless, these metrics mostly lead to the same conclusions, and therefore only the results of the most relevant ones have been presented in the paper (namely F-Measure and accuracy).

In addition to the above metrics, two other approaches for performance evaluation of image binarization methods have been proposed, which do not utilize typical classification statistics based only on the number of misclassified pixels. The first one, known as Distance Reciprocal Distortion (DRD) [Lu04], has been proposed as an alternative for the Peak Signal to Noise Ratio (PSNR) and the other metrics, which do not match well with subjective quality assessment. This metric considers also the mutual relations between pixels as the perception of distortions in binary images is different than for natural images. Hence, the weighting matrix has been proposed to consider the distances between the flipped pixels from the central pixel, not only the number of the misclassified pixels. Another useful tool is the Misclassification Penalty Metric (MPM) [You05], where the flipped pixels are penalized by their distances from the objects. To calculate the MPM values the GT object's borders should be extracted followed by distance calculation for misclassified pixels (FN and FP). The final MPM score is calculated as:

$$MPM = \frac{1}{2 \cdot D} \times \sum_{i,j} \left( d_{FN}^{i} + d_{FP}^{j} \right) \qquad (1)$$

where $d_{FN}^{i}$ and $d_{FP}^{j}$ are the distances of the $i$-th false negative and $j$-th false positive, respectively. To make the results more independent on image size, the normalization factor $D$ has also been used, defined as the sum of the distances of all pixels to contours in the GT image. Nevertheless, although both these metrics are suitable for document images, they are not useful for halftone images (with dithering).

# 3 PROPOSED SOLUTION

Since the idea of the investigated approach is to have possibly low computational effort, preserving the performance comparable to adaptive algorithms, a relatively fast pre-processing step is used, which would not cause a significant increase of the computation time in contrast to e.g. median filtering. Hence, the global thresholding is assumed as the second stage, which should be conducted using an image with partially eliminated background information. Nevertheless, it has been assumed that some minor distortions may not be eliminated during the pre-processing stage, which should be suppressed by the final thresholding.

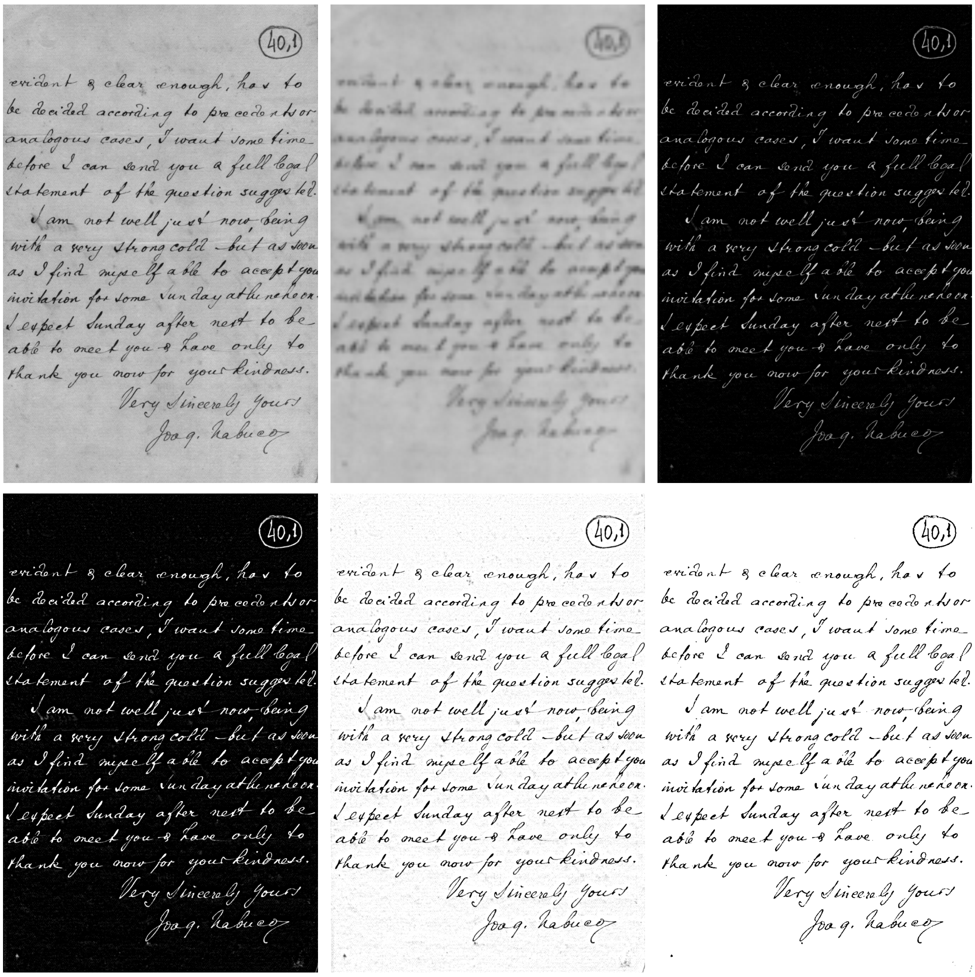
The initial background estimation is obtained using the image downsampling and further upsampling using one of typically used kernels: bilinear, bicubic and Lanczos (in two versions with *sinc* function of $\alpha$ equal to 2 and 3), as well as using simple nearest neighbour approach. For this purpose MATLAB's *imresize* function may be used with appropriate parameters, assuming the use of the same kernel for both operations. The effects obtained after the use of consecutive steps of the algorithm are shown on the top images of Fig. 1, where the influence of image resampling may be observed as well as the extracted details representing the foreground text, used in further steps. A crucial parameter, subjected to optimization, is the kernel size $sc$ (referred also as the scale factor), significantly influencing the obtained results as shown in four plots in Fig. 2.

The next steps are the subtraction of two images to obtain the image with partially suppressed background, which is further enhanced by the increase of contrast. To maintain the speed of the method a simple multiplication by the contrast correction factor $k$ (with saturation of bright pixels) has been used, followed by image negation. Obviously, these two steps may be reversed, assuming the saturation of dark pixels for the increase of image contrast after negation. Such obtained image, using the optimized contrast correction factor is then subjected to final binarization, which may be conducted using a fixed threshold or any global thresholding method, e.g. Otsu's binarization, which has led to the best results during the experiments. The results of these three steps for a sample document image are illustrated in the bottom part of Fig. 1.

During the experiments, various interpolation kernels used during down- and upsampling, as well as different final thresholding methods, have been checked. Surprisingly, the best results have been achieved using the bilinear kernel combined with Otsu's thresholding in the final step. The illustration of the results of the metrics obtained during the optimization of the parameters is presented in Figures 2 and 3. Analysing these plots, the combination of the smallest reasonable contrast correction factor with the possibly universal scale factor, leading to good results for the datasets containing images of different sizes, should be chosen (in this case $k = 0.5$ for contrast and $sc = 32$, respectively).

# 4 EXPERIMENTAL RESULTS

To verify the usefulness of the proposed optimized method, several known algorithms, shortly presented in Section 2.1, have been used and the comparisons of the accuracy, F-Measure, DRD and MPM metrics have been made using four datasets. During the experiments the 116 images from DIBCO 2009 – 2018 datasets have been aggregated into a single dataset and the three remaining ones have been proposed in conjunction with the Time-Quality Document Binarization Competition (TQ-DIB). Two of them contain the images of letters and postcards written and typed by Brazilian writer and



40,1

evident & clear enough, has to
be decided according to precedents or
analogous cases, I want some time
before I can send you a full legal
statement of the question suggested.
I am not well just now, being
with a very strong cold – but as soon
as I find myself able to accept your
invitation for some Sunday at the noon.
I expect Sunday after next to be
able to meet you & have only to
thank you now for your kindness.
Very Sincerely Yours
Joaq. Nabuco

Figure 1: Results obtained after consecutive steps of the proposed method for a sample document image from Nabuco dataset – from left to right: original image, estimated background and their difference (top row), increased contrast, negation, final result after Otsu's thresholding (bottom row).

statesman Joaquim Nabuco – one of them, containing 15 images, is referred as "Nabuco dataset" whereas the second one consisting of 20 images, used in TQ-DIB 2019 competition [Lin19], is referred as "Nabuco competition". However, most of the images in those two datasets are different, hence they have been considered separately during the experiments. The fourth dataset, called LiveMemory, available at the website of the TQ-DIB organizers as well[1], contains 15 document images selected from the proceedings of the Brazilian Telecommunications Society. Obviously, all the datasets contain the respective binary ground-truth images.

Regardless of the performance metrics, the approximate execution times have also been determined using the same PC with Intel Core-i7 CPU and 16GB of RAM with MATLAB 2018b running on 64-bit Windows 10 operating system. Although the presented times should be considered as approximate, it is possible to compare the computational effort of individual algorithms, even though the implementation of the proposed method has not been optimized in terms of processing speed. As expected, the proposed method has turned out to be obviously slower than global thresholding but still significantly faster than all adaptive methods.

To make the comparison of the execution time more universal and hardware independent all execution times have been determined relatively to the global Otsu's method. The advantages of some adaptive methods utilising integral images may also be easily observed, e.g. comparing time rank positions of both Bradley methods. Nevertheless, it is worth to note that the proposed method is faster even in comparison with the global Kittler method.

Some more detailed results of the four considered performance metrics for individual datasets are presented in Tables 1 and 2, demonstrating the advantages of the proposed approach. For a reliable comparison of all methods both in terms of time and quality, they have been ranked according to each of four performance

[1] https://dib.cin.ufpe.br/

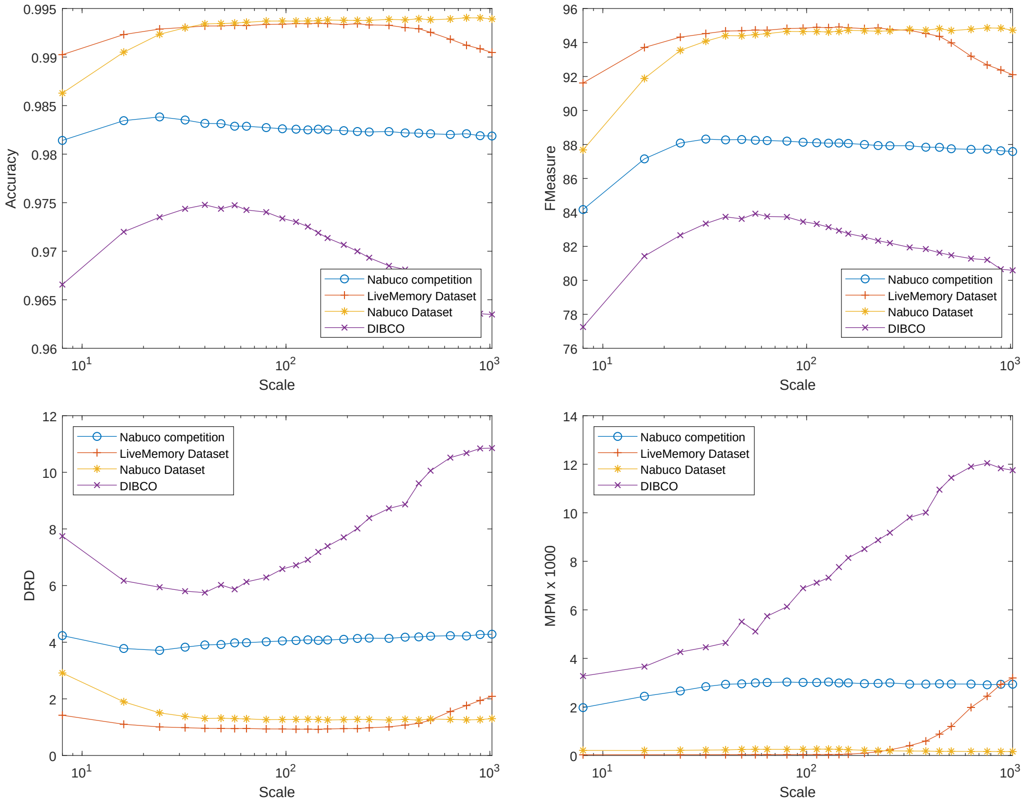



Figure 2: Illustration of results obtained during the optimization of parameter *sc*.

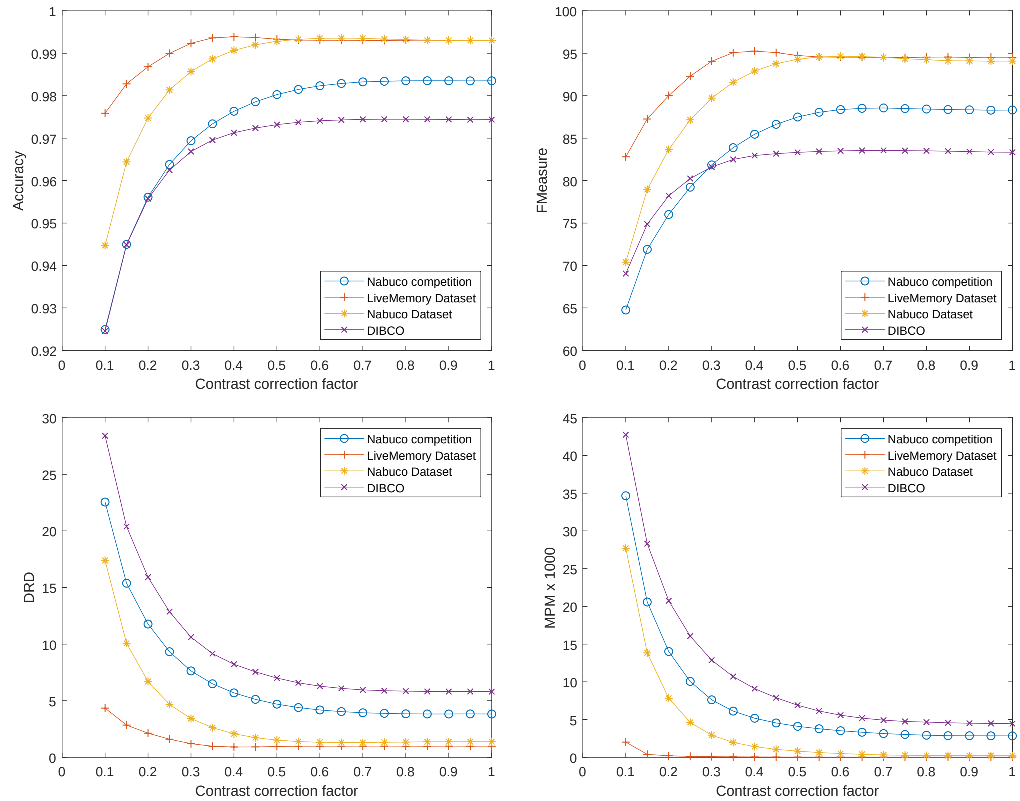



Figure 3: Illustration of results obtained during the optimization of parameter *k*.

| Binarization method | Reference | DIBCO | | | | LiveMemory | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | FM | ACC | DRD | MPM ×1000 | FM | ACC | DRD | MPM ×1000 |
| Otsu | [Ots79] | 78.77 | 0.948 | 16.72 | 22.22 | 89.64 | 0.986 | 3.78 | 6.65 |
| Kittler | [Kit86] | 70.96 | 0.920 | 28.24 | 32.94 | 70.27 | 0.949 | 11.19 | 16.16 |
| Meanthresh | – | 81.47 | 0.969 | 9.01 | 10.11 | 90.47 | 0.987 | 2.04 | 0.23 |
| Niblack | [Nib86] | 74.68 | 0.951 | 16.29 | 22.73 | 90.25 | 0.988 | 2.32 | 1.26 |
| Sauvola | [Sau00] | 78.08 | 0.957 | 12.11 | 12.13 | 88.97 | 0.985 | 2.43 | 0.17 |
| Wolf | [Wol04] | 76.24 | 0.953 | 13.43 | 13.27 | 86.46 | 0.981 | 3.08 | 0.22 |
| Feng | [Fen04] | 66.76 | 0.933 | 22.54 | 22.86 | 90.99 | 0.989 | 2.06 | 0.29 |
| Bradley (Mean) | [Bra07] | 77.43 | 0.954 | 16.07 | 22.20 | 88.66 | 0.985 | 2.48 | 0.20 |
| Bradley (Gaussian) | [Bra07] | 75.33 | 0.946 | 18.99 | 24.97 | 87.36 | 0.982 | 2.86 | 0.57 |
| NICK | [Khu09] | 69.90 | 0.960 | 9.21 | **3.50** | 89.22 | 0.989 | 1.68 | **0.03** |
| Chou | [Cho10] | 74.02 | 0.940 | 24.77 | 37.84 | 86.07 | 0.980 | 5.89 | 18.50 |
| Region-based (1 layer) | [Mic19c] | 76.29 | 0.956 | 13.30 | 14.89 | 91.36 | 0.989 | 1.82 | 0.17 |
| Region-based (2 layers) | [Mic19c] | 78.68 | 0.962 | 11.68 | 13.71 | 92.63 | 0.990 | 1.55 | 0.14 |
| Region-based (4 layers) | [Mic19c] | 79.34 | 0.963 | 11.25 | 13.07 | 92.90 | 0.991 | 1.49 | 0.13 |
| Region-based (6 layers) | [Mic19c] | 79.44 | 0.964 | 11.18 | 12.93 | 92.94 | 0.991 | 1.48 | 0.13 |
| Region-based (8 layers) | [Mic19c] | 79.51 | 0.964 | 11.13 | 12.82 | 92.96 | 0.991 | 1.48 | 0.13 |
| Region-based (12 layers) | [Mic19c] | 79.53 | 0.964 | 11.11 | 12.75 | 92.97 | 0.991 | 1.48 | 0.13 |
| Region-based (16 layers) | [Mic19c] | 79.55 | 0.964 | 11.09 | 12.73 | 92.98 | 0.991 | 1.48 | 0.13 |
| Entropy filter + Otsu | [Mic19b] | 80.73 | 0.966 | 11.22 | 10.79 | 94.43 | 0.993 | 1.18 | 0.09 |
| **Proposed (bilinear; k=0.5; sc=32)** | [Mic19a] | **83.33** | **0.973** | **7.00** | 6.91 | **94.74** | **0.993** | **0.96** | 0.04 |

Table 1: Results of performance metrics obtained for DIBCO and LiveMemory datasets

| Binarization method | Reference | Nabuco competition (TQ-DIB) | | | | Nabuco dataset | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | FM | ACC | DRD | MPM ×1000 | FM | ACC | DRD | MPM ×1000 |
| Otsu | [Ots79] | 86.29 | 0.978 | 5.36 | 4.15 | 94.01 | 0.993 | 1.63 | 0.37 |
| Kittler | [Kit86] | 73.22 | 0.950 | 13.24 | 9.34 | 78.71 | 0.967 | 9.01 | 3.40 |
| Meanthresh | – | 85.74 | 0.978 | 5.59 | 6.47 | 90.98 | 0.988 | 3.14 | 3.60 |
| Niblack | [Nib86] | 77.10 | 0.961 | 11.76 | 18.97 | 82.81 | 0.974 | 8.28 | 13.64 |
| Sauvola | [Sau00] | 75.25 | 0.957 | 11.09 | 8.00 | 83.25 | 0.975 | 6.27 | 2.39 |
| Wolf | [Wol04] | 76.24 | 0.959 | 10.54 | 7.41 | 83.86 | 0.976 | 6.04 | 1.83 |
| Feng | [Fen04] | 76.14 | 0.962 | 10.94 | 10.72 | 82.11 | 0.973 | 7.92 | 7.12 |
| Bradley (Mean) | [Bra07] | 74.80 | 0.953 | 12.27 | 11.96 | 80.97 | 0.971 | 7.92 | 7.25 |
| Bradley (Gaussian) | [Bra07] | 73.05 | 0.949 | 13.76 | 12.53 | 78.20 | 0.966 | 9.91 | 8.14 |
| NICK | [Khu09] | 84.58 | **0.982** | **4.43** | **2.89** | 89.90 | 0.988 | 2.58 | 0.71 |
| Chou | [Cho10] | 77.96 | 0.963 | 15.05 | 21.95 | 80.36 | 0.965 | 18.56 | 34.04 |
| Region-based (1 layer) | [Mic19c] | 78.64 | 0.965 | 9.78 | 12.25 | 85.13 | 0.978 | 6.09 | 6.96 |
| Region-based (2 layers) | [Mic19c] | 81.23 | 0.970 | 8.46 | 11.20 | 87.63 | 0.982 | 5.05 | 6.55 |
| Region-based (4 layers) | [Mic19c] | 81.96 | 0.971 | 8.10 | 10.50 | 88.39 | 0.983 | 4.72 | 6.15 |
| Region-based (6 layers) | [Mic19c] | 82.08 | 0.971 | 8.04 | 10.38 | 88.50 | 0.983 | 4.68 | 6.08 |
| Region-based (8 layers) | [Mic19c] | 82.14 | 0.971 | 8.00 | 10.30 | 88.55 | 0.984 | 4.65 | 6.02 |
| Region-based (12 layers) | [Mic19c] | 82.17 | 0.971 | 7.97 | 10.24 | 88.58 | 0.984 | 4.64 | 5.99 |
| Region-based (16 layers) | [Mic19c] | 82.19 | 0.971 | 7.96 | 10.22 | 88.59 | 0.984 | 4.63 | 5.98 |
| Entropy filter + Otsu | [Mic19b] | 87.24 | 0.980 | 4.78 | 3.70 | **94.48** | **0.993** | **1.44** | **0.32** |
| **Proposed (bilinear; k=0.5; sc=32)** | [Mic19a] | **87.49** | 0.980 | 4.70 | 4.11 | 94.29 | **0.993** | 1.52 | 0.82 |

Table 2: Results of performance metrics obtained for both Nabuco datasets

| Binarization method | Relative execution time | Time rank | Performance (Quality) rank | Time-Quality rank |
|---|---|---|---|---|
| Otsu | 1.00 | **1** | 6 | 2 |
| Kittler | 22.87 | 5 | 18 | 10 |
| Meanthresh | 37.37 | 6 | 2 | 3 |
| Niblack | 71.09 | 9 | 15 | 11 |
| Sauvola | 71.68 | 10 | 11 | 9 |
| Wolf | 76.14 | 11 | 14 | 13 |
| Feng | 197.89 | 14 | 17 | 19 |
| Bradley (Mean) | 18.09 | 4 | 15 | 5 |
| Bradley (Gaussian) | 172.63 | 13 | 20 | 20 |
| NICK | 65.86 | 8 | 4 | 4 |
| Chou (without SVM) | 6.72 | 2 | 18 | 6 |
| Region-based (1 layer) | 53.70 | 7 | 13 | 6 |
| Region-based (2 layers) | 112.78 | 12 | 12 | 11 |
| Region-based (4 layers) | 221.94 | 15 | 10 | 13 |
| Region-based (6 layers) | 342.16 | 16 | 9 | 13 |
| Region-based (8 layers) | 474.42 | 17 | 8 | 13 |
| Region-based (12 layers) | 713.10 | 19 | 7 | 18 |
| Region-based (16 layers) | 959.53 | 20 | 5 | 13 |
| Entropy filter + Otsu | 646.34 | 18 | 2 | 6 |
| **Proposed (bilinear; k=0.5; sc=32)** | 12.06 | **3** | **1** | **1** |

Table 3: Approximate aggregated execution times relative to Otsu's method and obtained ranking scores

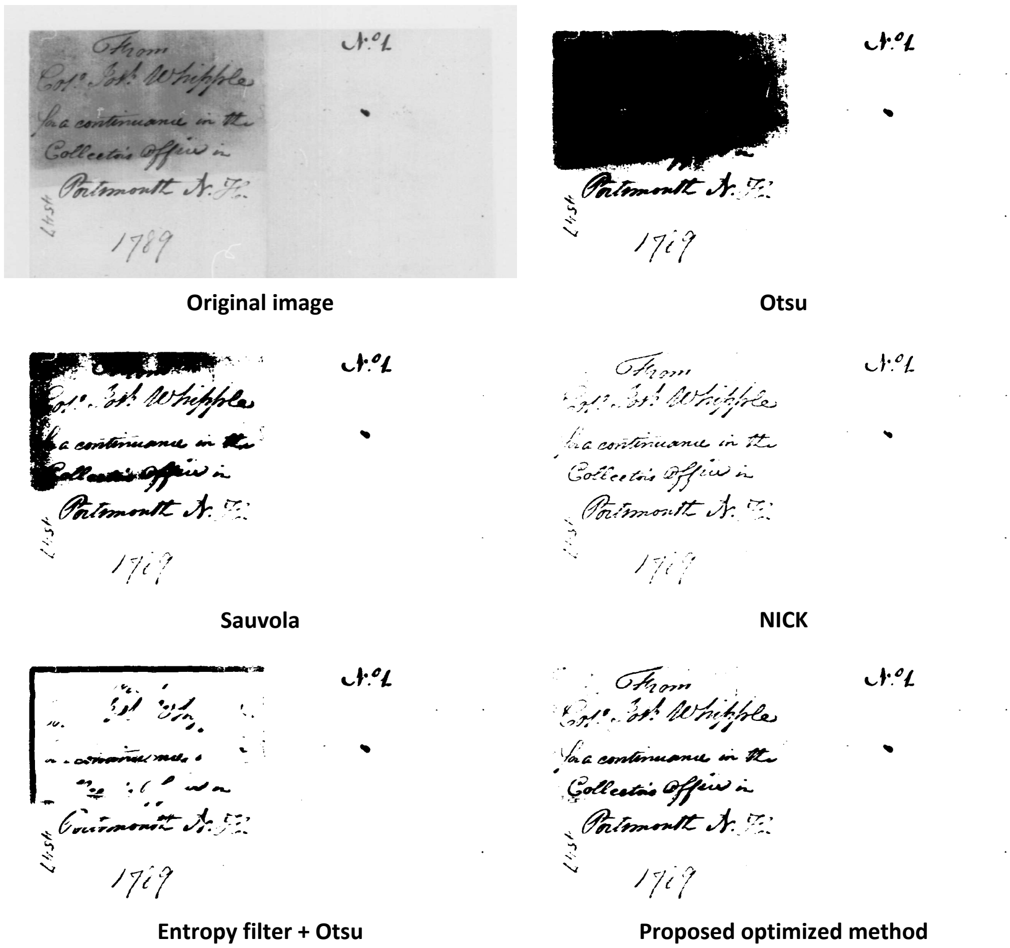



Figure 4: Results of the use of selected binarization methods for a sample challenging image used in experiments.

metrics for 4 datasets, leading to 16 rankings. The final quality ranking has been obtained after calculating the sum of the obtained positions, whereas time-quality ranking has been obtained in the same manner using positions in the final quality and execution time ranking.

## 5 DISCUSSION OF RESULTS

Analysing the results presented in Tables 1 and 2, much worse results may be noticed for DIBCO dataset, caused by much higher number of images, as well as the presence of some more challenging images in the latest DIBCO subsets. The Nabuco dataset seems to be much less challenging, as relatively good performance for these images may be noticed even using classical Otsu's method. Regardless of the dataset and the metric, the proposed method is always in top three methods. Since for both Nabuco datasets slightly better results may be obtained using the entropy based method [Mic19b] or NICK thresholding [Khu09], an important advantage of the method based on background estimation is its lower computational complexity, as illustrated in Table 3, where achieved execution times in relation to Otsu's method have been presented together with ranking scores.

A comparison of the results obtained for a sample challenging image is presented in Fig. 4, where the advantages of the proposed method are clearly visible. Since the presented sample image contains different shadows combined with rapid changes of intensity at the edges, the use of global thresholding methods is inappropriate. Nevertheless, serious problems may be observed also using adaptive methods, such as Sauvola and NICK, which have performed much better than the other methods in this case, in contrast to the proposed approach. Hence, the usefulness of the binarization based on background estimation for degraded document images has been confirmed, also considering the processing time constraints.

Nevertheless, analysing the potential limitations of the proposed approach, the main challenge might be a combination of degradations typical for document images with highly non-uniform illumination, where some more sophisticated time-consuming methods may be necessary, especially for images captured by mobile devices in unknown lighting conditions.

## 6 CONCLUDING REMARKS

The proposed optimized binarization method based on the fast background estimation using the down- and upsampling with bilinear kernel combined with increase of contrast and Otsu's thresholding, makes it possible to achieve very good performance for state-of-the-art image binarization datasets. Since it combines a good performance with relatively short processing time, it may be an interesting alternative for classical algorithms, particularly for low computational power devices. As the practical usefulness of this approach has also been verified in terms of the OCR accuracy, it may be considered as a universal binarization method for low quality document images.

A similar performance may be obtained using another two-stage method proposed recently [Mic19b], utilizing entropy filtering as the pre-processing step, followed by Otsu's thresholding as well. Nevertheless, this method is much slower even in comparison to adaptive methods. A combination of both these preprocessing approaches may be however considered in future research, although such an approach would be time-consuming and potentially useful only for off-line document image analysis. Another direction of our further work may be the application of the proposed approach for challenging pre-processing of images of unevenly illuminated industrial nameplates, captured by cameras mounted in mobile devices, subjected to further text recognition.

## 7 REFERENCES

[Ber86] Bernsen, J. Dynamic thresholding of grey-level images, in *Proc. 8th International Conference on Pattern Recognition (ICPR)*, pp. 1251-1255, 1986.

[Bra07] Bradley, D., and Roth, G. Adaptive thresholding using the integral image. *Journal of Graphics Tools* 12(2):13-21, 2007.

[Cho10] Chou, C.-H., Lin, W.-H., and Chang F. A binarization method with learning-built rules for document images produced by cameras. *Pattern Recognition* 43(4):1518-1530, 2010.

[Cou10] Coudray, N., Buessler, J.-L., and Urban, J.-P. Robust threshold estimation for images with unimodal histograms. *Pattern Recognition Letters* 31(9):1010-1019, 2010.

[Fen04] Feng,, M.-L., and Tan, Y.-P. Adaptive binarization method for document image analysis, in *Proc. 2004 IEEE Internatonal Conference on Multimedia and Expo (ICME)*, vol. 1, pp. 339-342, 2004.

[Has19] Hassina, B. Title segmentation in Arabic document pages, in *Proc. 27. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)* – part II (*Computer Science Research Notes*, vol. 2902), pp. 45-50, 2019.

[Kap85] Kapur, J.N., Sahoo, P.K., and Wong, A.K.C. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing* 29(3):273-285, 1985.

[Khu09] Khurshid, K., Siddiqi, I., Faure, C., and Vincent, N. Comparison of Niblack inspired binarization methods for ancient documents, in *Proc. SPIE – Document Recognition and Retrieval XVI*, article no. 724709, 2009.

[Kit86] Kittler, J., and Illingworth, J. Minimum error thresholding. *Pattern Recognition* 19(1): 42-47, 1986.

[Laz14] Lazzara, G., and Geráud, T. Efficient multiscale Sauvola's binarization. *International Journal on Document Analysis and Recognition* 17:105-123, 2014.

[Lef19] Lefkovits, S., Lefkovits, L., Szilágyi, L. CNN approaches for dorsal hand vein based identification, in *Proc. 27. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)* – part II (*Computer Science Research Notes*, vol 2902), pp. 51-60, 2019.

[Lin19] Lins, R, Bernardino, R., and Jesus, D.M. A quality and time assessment of binarization algorithms, in *Proc. 15th IAPR International Conference on Document Analysis and Recognition (ICDAR)* pp. 1444-1450, 2019.

[Lu04] Lu, H., Kot, A.C, and Shi, Y.Q. Distance-reciprocal distortion measure for binary document images. *IEEE Signal Processing Letters* 11(2):228-231, 2004.

[Mic19a] Michalak, H., Okarma, K. Fast binarization of unevenly illuminated document images based on background estimation for optical character recognition purposes. *Journal of Universal Computer Science* 25(6):627-646, 2019.

[Mic19b] Michalak, H., Okarma, K. Improvement of image binarization methods using image preprocessing with local entropy filtering for alphanumerical character recognition purposes. *Entropy* 21(6), article no. 562, 2019.

[Mic19c] Michalak, H., Okarma, K. Adaptive Image binarization based on multi-layered stack of regions, in *Proc. of International Conference on Computer Analysis of Images and Patterns (CAIP)*, LNCS, vol. 11679, pp. 281-293, 2019.

[Mog12] Moghaddam, R.F, and Cheriet, M. AdOtsu: an adaptive and parameterless generalization of Otsu's method for document image binarization. *Pattern Recognition* 45(6):2419-2431, 2012.

[Nib86] Niblack W. An introduction to digital image processing. Prentice Hall, Englewood Cliffs, 1986.

[Nti13] Ntirogiannis, K., Gatos, B., and Pratikakis, I. Performance evaluation methodology for historical document image binarization. *IEEE Transactions on Image Processing* 22(2):595-609, 2013.

[Ots79] Otsu, N. A threshold selection method from gray-level histograms/ *IEEE Trans. on Systems, Man, and Cybernetics* 9(1):62-66, 1979.

[Pra17] Pratikakis, I., Zagoris, K., Barlas, G., and Gatos, B. ICDAR2017 competition on document image binarization (DIBCO 2017), in *Proc. 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1395-1403, 2017.

[Ros01] Rosin, P. Unimodal thresholding. *Pattern Recognition* 34(11):2083-2096, 2001.

[Sam16] Samorodova, O.A., and Samorodov, A.V. Fast implementation of the Niblack binarization algorithm for microscope image segmentation. *Pattern Recognition and Image Analysis* 26(3):548-551, 2016.

[Sau00] Sauvola, J., and Pietikäinen, M. Adaptive document image binarization. *Pattern Recognition* 33(2):225-236, 2000.

[Sax19] Saxena, L.P. Niblack's binarization method and its modifications to real-time applications: a review. *Artificial Intelligence Review*, 51(4):673-705, 2019.

[Sha08] Shafait, F., Keysers, D., and Breuel, T.M. Efficient implementation of local adaptive thresholding techniques using integral images, in *Proc. SPIE – Document Recognition and Retrieval XV*, article no. 681510, 2008.

[Sul19] Sulaiman, A., Omar, Kh., and Nasrudin, M.F. Degraded historical document binarization: a review on issues, challenges, techniques, and future directions. *Journal of Imaging* 5(4), article no. 48, 2019.

[Ten17] Tensmeyer, C., and Martinez, T. Document image binarization with fully convolutional neural networks, in *Proc. of 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pp. 99-104, 2017.

[Vo18] Vo, Q.N., Kim, S.H., Yang, K.J., and Lee, G. Binarization of degraded document images based on hierarchical deep supervised network. *Pattern Recognition* 74:568-586, 2018.

[Wol04] Wolf, C., and Jolion, J.-M. Extraction and recognition of artificial text in multimedia documents. *Formal Pattern Analysis & Applications* 6(4):309-326, 2004.

[You05] Young, D.P., and Ferryman, J.M. PETS metrics: on-line performance evaluation service, in *Proc. IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 317-324, 2005.