Piotr LECH, Krzysztof OKARMA, Dominik WOJNAR

West Pomeranian University of Technology, Szczecin

doi:10.15199/48.2015.02.18

Binarization of document images using the modified local-global Otsu and Kapur algorithms

***Abstract**. In the paper an algorithm for binarization of grayscale images representing text documents together with its optimization is presented. In the proposed approach two classical global thresholding algorithms proposed by Otsu and Kapur are combined with their local versions applied for blocks. The experimental results have been obtained for the H-DIBCO test dataset containing handwritten text images together with their "ground-truth" binary equivalents.*

***Streszczenie.** W artykule zaprezentowano algorytm binaryzacji obrazów w skali szarości przedstawiających dokumenty tekstowe wraz z jego optymalizacją. Przedstawione podejście bazuje na połączeniu dwóch klasycznych metod progowania zaproponowanych przez Otsu i Kapura z ich lokalnymi wariantami zastosowanymi dla bloków. Wyniki eksperymentalne uzyskano dla bazy testowej H-DIBCO zawierającej obrazy rękopisów wraz z ich binarnymi odpowiednikami stanowiącymi wzorce.*
Binaryzacja obrazów dokumentów z użyciem zmodyfikowanych lokalno-globalnych algorytmów Otsu i Kapura

Keywords: image binarization, image analysis.
Słowa kluczowe: binaryzacja obrazu, analiza obrazu.

Introduction

Binarization of monochrome images based on global criteria using e.g. Kapur [1] or Otsu [2] methods often leads to some misrepresentations in the binary image caused by the non-uniformly illumination of the scene or massive large differences of luminance (stains, smudges, discolorations) as illustrated in Fig. 1. The most typical defects are overexposures or blackened areas causing the invisibility of details. Modern approaches are typically based on the algorithms related to the local thresholding characterized by higher computational complexity than the global ones [3]. The algorithm proposed in the paper should not be considered as the replacement for them but rather as a supplementary method allowing the extension of basic Kapur and Otsu methods which are treated as the reference. In the experimental verification and optimization of the discussed approach the high resolution handwritten text images from the H-DIBCO database [4] have been used and therefore the comparison of the obtained results with "ground truth" binary images has also been done.

Binarization is commonly used as the preprocessing step in many document image understanding and document image analysis applications as well as in some other Optical Character Recognition (OCR) algorithms, recognition of handwritten text or even for map digitization [5]. The result of binarization influences significantly the final result of the complex process of image analysis. The reduction of the amount of data conducted at the binarization stage decreases the overall complexity of image recognition algorithms and therefore allows shortening the time necessary for the whole process of the document image analysis.

The goal of the binarization is the reduction of unnecessary information leaving only the data useful for further processing. Many computer applications used in various fields of science and technology often utilize thresholding inside their procedures and the threshold value is set manually by the operator e.g. in preprocessing of handwritten or machine printed text images for the OCR algorithms, biomedical image analysis, geological image analysis etc. The approach described in the paper is devoted to the simplification of such task by giving the operator a choice of several pre-selected threshold values which should be better than obtained by the classical Kapur or Otsu methods.

Binarization algorithms can be divided into six main classes based on: clustering, entropy, spatial binarization, foreground attribute and histogram which are considered in this paper. The sixth group of methods consists some locally adaptive methods [6, 7]. Nevertheless, some other authors use another classification of binarization methods dividing them only into clustering-based, threshold-based and hybrid [8] where threshold-based ones can be further divided into local and global algorithms.

Local thresholding methods, more computationally complex in comparison to global algorithms discussed later, have been proposed e.g. by Bernsen [9], where the average values and contrast information from a local region are utilized, or Niblack [10] using the sliding rectangular window approach and Sauvola [11] who proposed its modified version based on different assumed distribution of pixels representing the foreground and background information.

Recently some adaptive methods have been proposed, especially useful for the binarization of document images, often with assumed non-uniform illumination or developed for degraded documents with differing background [12]. Various methods of such kind have been presented e.g. by Badekas [13], Valizadeh [14], Kuk [15] and Ramirez [16].

An interesting approach based on background estimation and stroke edges has been presented by Lu [17] and the application of quality prediction of obtained binary images has been recently discussed by Rabeux [18]. Another adaptive algorithm for binarization of degraded documents with uneven background has been proposed by Singh [19] which consists of four main steps: local contrast analysis, contrast stretching, thresholding and noise removal.

Global binarization

Considering the binarization algorithm proposed by Kapur it can be noticed that two separate classes (representing the background and useful data) are described by independent probability distributions which do not overlap each other (similarly as in Otsu method). The main assumption of the algorithm is that the maximum entropy is interpreted as an indicator of high information content in the thresholded image. The image is described by n grayscale levels (from 0 to $n-1$). The object (ob) and the background (b) are characterized by the probability distributions briefly expressed as

(1) $$\frac{p_0}{p_{ob}}, \frac{p_1}{p_{ob}}, ..., \frac{p_t}{p_{ob}}$$

for the object and

(2) $$\frac{p_0}{p_{ob}}, \frac{p_1}{p_{ob}}, ..., \frac{p_t}{p_{ob}}$$

for the background, where t is the threshold, $p_{ob} = \sum_{i=0}^{t} p_i$ and $p_b = \sum_{i=t+1}^{n-1} p_i$.

The optimal threshold is the value maximizing the aggregated entropy:

(3) $$H_T = H_{ob} + H_b = -\sum_{i=0}^{t} \frac{p_i}{p_{ob}} \log_2\left(\frac{p_i}{p_{ob}}\right) - \sum_{i=t+1}^{n-1} \frac{p_i}{p_b} \log_2\left(\frac{p_i}{p_b}\right).$$

In the alternative method proposed by Otsu it is assumed that the threshold t divides the image into two classes (object and background) with the same definitions of p_{ob} and p_b as above. This method has an especially high performance for the images with bimodal histograms where two normal distributions partially overlapping each other may represent the histogram of the image. The image with separated two classes can be characterized by the mean values and variances for both classes as:

(4) $$\mu_{ob} = \sum_{i=0}^{t} \frac{i \cdot p_i}{p_{ob}} \quad \text{and} \quad \mu_b = \sum_{i=t+1}^{n-1} \frac{i \cdot p_i}{p_b}$$

representing the mean values and the variances:

(5) $$\sigma_{ob}^2 = \sum_{i=0}^{t} (i - \mu_{ob})^2 \frac{p_i}{p_{ob}} \quad \text{and} \quad \sigma_b^2 = \sum_{i=t+1}^{n-1} (i - \mu_b)^2 \frac{p_i}{p_b} .$$

At this stage the histogram of the input image is characterized by two pairs of numbers for two classes: mean intensity and variance for the object and mean intensity and variance for the background. It is equivalent to the estimation of the distribution of the luminance levels in the image by two normal distributions described by two parameters (mean value and variance).

(a) Original image (b) "Ground truth" binary image

(c) result of Otsu binarization (c) result of Kapur binarization

Fig.1. Illustration of binarization errors introduced by both considered algorithms

The global descriptors of the image are the global mean values and the global variance given as:

(6) $$\mu_T = \sum_{i=0}^{n-1} i \cdot p_i \quad \text{and} \quad \sigma_T^2 = \sum_{i=0}^{n-1} (i - \mu_T)^2 \cdot p_i$$

respectively.

The global variance can be expressed equivalently as the sum of the within-class variance σ_W and the between-class variance σ_B:

(7) $$\sigma_T^2 = \sigma_W^2 + \sigma_B^2 = \sum_{i=0}^{t} (i - \mu_T)^2 \cdot p_i + \sum_{i=t+1}^{n-1} (i - \mu_T)^2 \cdot p_i .$$

The global variance for a given image is constant independently on the threshold which influences only the values of the within-class and between-class variance which sum is always equal to the global variance. Minimization of the within-class variance is equivalent to the maximization of the between-class variance which is typically chosen as the criterion because of its lower computational cost. It is dependent only on the first order statistical moments:

(8) $$\sigma_B^2 = p_{ob} \cdot p_b \cdot (\mu_{ob} - \mu_b)^2 .$$

The maximum value of the between-class variance corresponds to the optimal separation of two classes in the image so the value t leading to such separation is considered as optimal threshold t_{opt}.

The idea of binarization is based on the assumption that two histogram modes represent two separate classes: objects and background. The problems appear if such assumption is not fulfilled leading to some errors caused by the presence of only a single threshold value obtained by the global thresholding. In the case of local thresholding, the classification of each pixel as representing the object or the background may be independent upon the analysis of its neighbourhood (context). Depending on the method used for the threshold setting, they can be divided into global, local and hybrid algorithms. In general, each global method can be applied in a pseudo-local version by dividing the image into the sub-images (blocks of a fixed size) and the use of global thresholding to each block independently. The advantage of such solution is its simplicity and an easy implementation based on existing global functions.

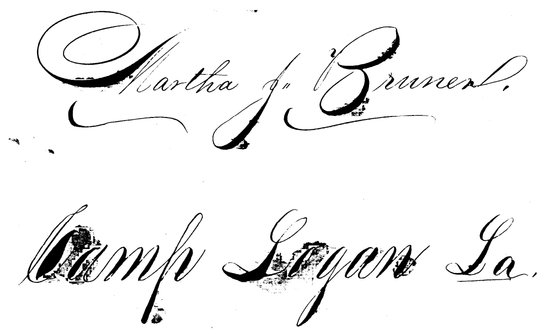
Nevertheless, this process may cause the presence of artefacts resulting from the independence of each block. On the other hand, analysing the images presented in Fig. 2 one can observe the text which is much better readable, even with artefacts, than the result of global binarization illustrated in Fig. 1.

Proposed hybrid thresholding and its experimental verification

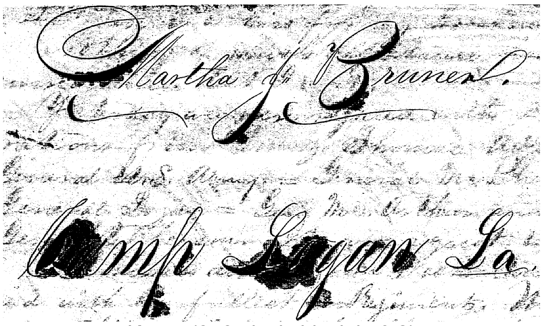
For elimination of the disadvantages of the local methods illustrated in Fig. 2 the weighted threshold value can be applied as the weighted average of the local and global threshold values (T_{local} and T_{global} respectively) obtained for Kapur and Otsu methods according to:

(9) $$T_{comb} = T_{global} \cdot k + T_{local} \cdot (1 - k)$$

where k can be any number from 0 to 1.



Otsu (16×16 pixels block k=0.5)



Kapur (8×8 pixels block k=0.2)

Fig.2. Improvement of the readability of the text (top) and arising artefacts (bottom) for the local thresholding methods

In order to verify the usefulness of the proposed approach some typical metrics [20] used for the verification of binarization and classification algorithms have been calculated for the images from H-DIBCO dataset [4]. Assuming the size of the block equal to 16×16, 32×32, 64×64 and 128×128 pixels and the values of the parameter k changing from 0 to 1 by the step 0.1, the Peak Signal-to-Noise Ratio (PSNR), F-Measure (based on Precision and Recall) and Distance Reciprocal Distortion (DRD) metric values [21] have been calculated for each of the obtained images. Setting the values of the parameter $k=0$ is equivalent to global thresholding and $k=1$ the local thresholding independently for each block. Exemplary representative values of the three mentioned measures for the tested image are shown in Tables 1 and 2 where the optimal values are indicated as bold.

Analysing the obtained results for all images the following cases can be distinguished:

1. three optimal measures for the same value of parameter k (except $k=0$ and $k=1$) are optimal,
2. at least two values of measures for the same value of parameter k (except $k=0$ and $k=1$) are optimal,
3. the optimal values of measures are obtained for different values of the parameter k,
4. the optimal values of measures are obtained for global thresholding.

Table 1. Values of all metrics obtained for the hybrid Kapur method for representative images from H-DIBCO dataset

block picture	**k**	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
16×16 H01	F-measure	35.79	41.16	46.60	54.03	62.83	71.22	77.88	82.40	84.76	**86.43**	86.10
	PSNR	4.40	5.38	6.31	7.59	9.15	10.78	12.30	13.53	14.27	**14.86**	14.72
	DRD	95.03	75.63	60.71	44.88	30.89	20.67	14.09	10.21	8.32	**7.03**	7.21
32×32 H03	F-measure	66.43	72.86	75.86	78.61	80.72	82.49	83.98	85.14	86.01	86.64	**87.02**
	PSNR	11.94	13.30	13.99	14.68	15.26	15.78	16.26	16.67	16.99	17.24	**17.38**
	DRD	17.56	12.46	10.39	8.66	7.40	6.41	5.57	4.94	4.47	4.12	**3.94**
64×64 H06	F-measure	71.09	78.15	81.01	82.83	83.42	83.85	**84.06**	83.98	83.94	83.65	83.44
	PSNR	13.43	15.11	15.90	16.45	16.66	16.82	16.92	16.93	**16.95**	16.90	16.85
	DRD	11.71	7.34	5.80	4.90	4.55	4.30	4.12	4.07	**4.01**	4.02	4.08
128×128 H02	F-measure	81.66	83.33	83.90	84.26	84.41	**84.47**	84.45	84.29	84.07	83.74	83.44
	PSNR	16.05	16.59	16.79	16.94	17.01	17.05	**17.07**	17.05	17.00	16.93	16.85
	DRD	5.50	4.64	4.35	4.12	4.01	3.93	**3.89**	**3.89**	3.91	3.97	4.08

Table 2. Values of all metrics obtained for the hybrid Otsu method for representative images from H-DIBCO dataset

block picture	**k**	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
16×16 H04	F-measure	35.85	87.08	**90.12**	89.99	89.52	88.93	88.25	87.55	86.87	86.15	85.63
	PSNR	5.75	16.62	**17.97**	17.96	17.79	17.57	17.34	17.11	16.90	16.69	16.53
	DRD	69.67	4.18	2.59	**2.56**	2.67	2.82	3.00	3.19	3.37	3.56	3.71
32×32 H08	F-measure	41.48	70.36	79.92	83.52	84.86	85.46	85.91	**86.16**	86.15	86.04	85.67
	PSNR	7.09	12.36	14.61	15.67	16.12	16.32	16.50	16.60	**16.61**	16.58	16.43
	DRD	48.52	13.08	7.03	5.09	4.39	4.07	3.78	3.59	3.54	**3.52**	3.66
64×64 H06	F-measure	47.57	66.47	75.10	78.38	79.68	80.38	80.77	80.96	81.00	80.99	**81.19**
	PSNR	9.81	13.23	15.06	15.87	16.22	16.42	16.55	16.61	16.64	16.65	**16.68**
	DRD	28.89	11.78	6.89	5.30	4.70	4.38	4.15	4.04	3.98	3.95	**3.91**
128×128 H09	F-measure	34.28	72.79	79.86	82.12	**82.26**	82.18	81.86	81.56	81.36	81.18	81.09
	PSNR	8.75	15.91	17.63	18.29	**18.35**	18.33	18.28	18.22	18.18	18.15	18.12
	DRD	51.44	8.01	4.61	3.61	**3.50**	3.51	3.54	3.59	3.62	3.65	3.67

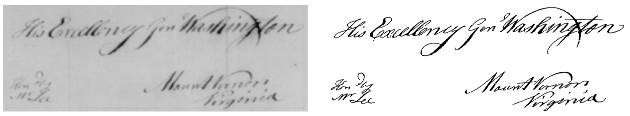
Table 3. The best results obtained for H10 image (for the block 16×16 pixels)

method picture	**k**	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
Otsu H10	F-measure	21.74	74.81	**85.96**	85.89	84.94	83.92	82.83	81.86	80.95	80.00	79.24
	PSNR	4.35	14.76	18.01	**18.10**	17.87	17.62	17.37	17.15	16.95	16.74	16.57
	DRD	151.82	11.89	4.38	**4.10**	4.29	4.54	4.83	5.10	5.37	5.66	5.94
Kapur H10	F-measure	19.02	38.93	77.77	**86.19**	86.39	86.06	85.47	84.74	83.79	82.79	81.78
	PSNR	3.46	7.87	15.30	17.88	**18.01**	17.94	17.79	17.61	17.35	17.10	16.84
	DRD	186.12	66.02	10.02	4.47	**4.20**	4.23	4.36	4.56	4.87	5.20	5.60

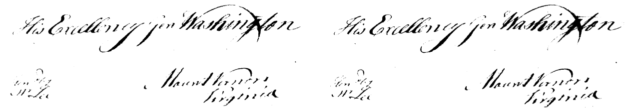
Unfortunately, for a single image any combination of the above optimal values of measures can take place as shown in Tables 1 and 2. The presence of the specific combination of the optimal values of measures is also dependent on the block size since a small block, e.g. 16×16 pixels, with relatively small weight may lead to significant better results than obtained for large blocks or the global thresholding.

The observed effects are quite similar for both algorithms analysed in the paper.

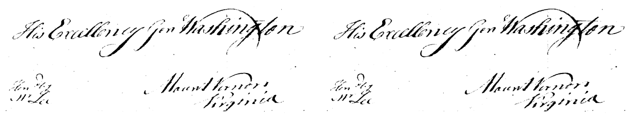
In Table 3 the best obtained results from the whole H-DIBCO dataset are presented which have been achieved for the H10 image, illustrating the advantages of the proposed hybrid methods. Images obtained after the binarization for the parameter k and the block size leading to the optimal values of the PSNR, DRD and F-Measure are presented in Fig. 3. For example, observing the handwritten letter "M", a significant improvement of its visibility can be easily noticed for the hybrid thresholding methods, especially for Kapur algorithm.



(a) Original H10 image (b) "Ground truth" for H10 image



(c) global Otsu binarization (d) global Kapur binarization



(e) Otsu – 16×16 pixels, k=0.3 (f) Kapur – 16×16 pixels, k=0.4

Fig.3. Presentation of resulting images corresponding to the results from Table 3

Summary and future work

Handwritten document images are not the only type of images which can be subjected to the proposed binarization algorithms since the useful information may be lost from many images after the global thresholding. Since the useful details present on the image can be different, there is no exact rule allowing the use of Otsu or Kapur algorithms instead of each other both in global and hybrid versions. It can be assumed that the main properties of each method in its global version are similar also in its hybrid variant. Nevertheless, for some specific images the results obtained using both global algorithms are very similar so it may be hard to predict which of hybrid algorithms would lead to better results. Both algorithms analysed in the paper can be accelerated using the Monte Carlo approach [22, 23].

The solution proposed in the paper allows to obtain better readability of handwritten text extracted from the grayscale document images in comparison to classical Kapur and Otsu binarization algorithms. It has been shown that using the appropriate weighting coefficients for the hybrid method, combining the local and global thresholding, the visibility of details on the resulting binary image can be significantly improved. Due to the application of the proposed methods, some overexposure errors as well as blackened areas can be partially reduced.

The next stages of our research will concentrate on the development of a method for the automatic choice of weighting coefficient and size of the block as well as the application of the Monte Carlo method.

REFERENCES

[1] Kapur J., Sahoo P., Wong A.: A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing* 29 (1985) No. 3, 273–285

[2] Otsu N.: A threshold selection method from gray-level histograms. *IEEE Trans. Systems, Man and Cybernetics* 9 (1979), No. 1, 62–66

[3] Stathis P., Kavallieratou E., Papamarkos N.: An evaluation technique for binarization algorithms. *Journal of Universal Computer Science* 14 (2008), No. 18, 3011–3030

[4] Pratikakis, I., Gatos, B., Ntirogiannis, K.: ICDAR 2009 Document Image Binarization Contest (DIBCO 2009). *Proc. 10th Int. Conf. Document Analysis and Recognition ICDAR*. pp. 1375–1382

[5] Mandal S., Biswas S., Das A.K., Chanda B.: Binarisation of colour map images through extraction of regions. *Lecture Notes in Computer Science* 8671 (2014), 418–427

[6] Sezgin M., Sankur B.: Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging* 13 (2004) No. 1, 146–168

[7] Chen S., Li D.: Image binarization focusing on objects. *Neurocomputing* 69 (2006), No. 16–18, 2411–2415

[8] Wen J., Li S., Sun J.: A new binarization method for non-uniform illuminated document images. *Pattern Recognition* 46 (2013), No. 6, 1670–1690

[9] Bernsen J.: Dynamic thresholding of grey-level images. *Proc. Int. Conf. Pattern Recognition ICPR'86*, 1251–1255

[10] Niblack W.: An Introduction to Digital Image Processing (1986), 115–116, Prentice-Hall, Englewood Cliffs

[11] Sauvola J., Pietikäinen M.: Adaptive document image binarization. *Pattern Recognition* 33 (2000), No. 2, 225–236

[12] Gatos B., Pratikakis I., Perantonis S.J.: Adaptive degraded document image binarization. *Pattern Recognition* 39 (2006), No. 3, 317–327

[13] Badekas E., Papamarkos N.: Estimation of appropriate parameter values for document binarization techniques. *Int. J. Robotics Autom.* 24 (2009), No. 1, 66–78

[14] Valizadeh M., Komeili M., Armanfard N., Kabir E.: Degraded document image binarization based on combination of two complementary algorithms. *Proc. ICACTEA'09*, IEEE 595–599

[15] Kuk J.G., Cho N.I.: Feature based binarization of document images degraded by uneven light condition. *Proc. 10th Int. Conf. Document Analysis and Recognition (ICDAR'2009)*, 748–752

[16] Ramírez, M.A., Tapia, E., Rojas, R., Cuevas, E.: Transition thresholds and transition operators for binarization and edge detection. *Pattern Recognition* 43 (2010), No.4, 3243–3254

[17] Lu S., Su B., Tan C.L.: Document image binarization using background estimation and stroke edges. *Int. J. Doc. Anal. Recognit.* (IJDAR) 13 (2010), No. 4, 303–314

[18] Rabeux V., Journet N., Vialard A., Domenger J.P.: Quality evaluation of degraded document images for binarization result prediction. *Int. J. Doc. Anal. Recognit.* (IJDAR) 17 (2014), No. 2, 125–137

[19] Singh B.M., Sharma R., Ghosh D., Mittal A.: Adaptive binarization of severely degraded and non-uniformly illuminated documents. *Int. J. Doc. Anal. Recognit.* (IJDAR) in press (2014)

[20] Ntirogiannis K., Gatos B., Pratikakis I.: Performance evaluation methodology for historical document image binarization. *IEEE Trans. Image Processing* 22 (2013), No. 2, 595–609

[21] Lu H., Kot A., Shi Y.: Distance-reciprocal distortion measure for binary document images. *IEEE Signal Processing Letters* 11 (2004), No. 2, 228–231

[22] Lech P., Okarma K.: Optimization of the fast image binarization method based on the Monte Carlo approach. *Elektronika Ir Elektrotechnika* 20 (2014), No. 4, 63–66

[23] Lech P., Okarma K., Tecław, M.: A fast histogram estimation based on the Monte Carlo method for image binarization. *Advances in Intelligent Systems and Computing* 233 (2014), 73–80

***Authors**: dr inż. Piotr Lech, dr hab. inż. Krzysztof Okarma, Dominik Wojnar, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, Wydział Elektryczny, ul. Sikorskiego 37, 70-313 Szczecin, E-mail: piotr.lech@zut.edu.pl, okarma@zut.edu.pl*