

INTERNATIONAL CONFERENCE ON MACHINE VISION

A generalization of Otsu method for linear separation of two unbalanced classes in document image binarization*E.I. Ershov¹, S.A. Korchagin¹, V.V. Kokhan^{1,3}, P.V. Bezmaternykh^{2,3}*¹ *Institute for Information Transmission Problems, RAS, 127051, Moscow, Bolshoy Karetny per., 19, str. 1,*² *Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia, 117312, pr. 60-letiya Oktyabrya, 9,*³ *Smart Engines Service LLC, Moscow, Russia, 117312, pr. 60-letiya Oktyabrya, 9***Abstract**

The classical Otsu method is a common tool in document image binarization. Often, two classes, text and background, are imbalanced, which means that the assumption of the classical Otsu method is not met. In this work, we considered the imbalanced pixel classes of background and text: weights of two classes are different, but variances are the same. We experimentally demonstrated that the employment of a criterion that takes into account the imbalance of the classes' weights, allows attaining higher binarization accuracy. We described the generalization of the criteria for a two-parametric model, for which an algorithm for the optimal linear separation search via fast linear clustering was proposed. We also demonstrated that the two-parametric model with the proposed separation allows increasing the image binarization accuracy for the documents with a complex background or spots.

Keywords: threshold binarization, Otsu method, optimal linear classification, historical document image binarization.

Citation: Ershov EI, Korchagin SA, Kokhan VV, Bezmaternykh PV. A generalization of Otsu method for linear separation of two unbalanced classes in document image binarization. *Computer Optics* 2021; 45(1): 66-76. DOI: 10.18287/2412-6179-CO-752.

Acknowledgments: We are grateful for the insightful comments offered by D.P. Nikolaev. This research was partially supported by the Russian Foundation for Basic Research No. 19-29-09066 and 18-07-01387.

Introduction

There are many commonly used algorithms for image binarization [1] employed for various applications: digital documents compression [2], intelligent document recognition [3], medical image analysis [4], object detection [5, 6], digital image enhancement [7], porosity determination via computed tomography [8], forensic ballistics [9]. The diversification of methods is associated with the fact that the algorithm allowing for the equally high accuracy in all image processing tasks has not been developed yet. The efficiency of any chosen method depends on its statistical properties and constrictions.

Binarization methods can be classified according to the number of tuning parameters. The most convenient methods for users (i.e. for researchers engaged in the related fields or engineers) are the zero-parametric methods, which do not require the parameters tuning to match the expected properties of the input data. One of such methods is commonly used Otsu method [10], which automatically determines the optimal global threshold that separates pixels into two classes based on the intensity histogram. The simplest one-parametric method requires a user to set such threshold. Yet, in some cases, the global methods can demonstrate good results, and the corresponding conditions are of great interest.

The well-known Niblack's method [11] in its classical implementation has two tuning parameters, one of which is responsible for the window size and is regulated depending on the expected size of the segmented objects. The issue of the parameter adjustment for the certain application is not trivial and usually is resolved via the optimization on the subset of the available sample images, while some criterion of the binarization accuracy is selected [12].

Currently, the most accurate binarization methods are based on the artificial neural networks (ANNs) [13, 14]. Such networks employ a lot of tuning coefficients, the optimization of which requires many properly annotated pixel-precision examples of the classification. Despite this, the trained neural network is usually considered to be zero-parametric method suitable for employment by researchers or engineers with no machine learning experience. Nevertheless, this approach still requires a user to study the limits of the applicability of any chosen network, which could be nontrivial.

In this work, we consider document image binarization – the separation of two well-defined classes of pixels: contents of the document (pixels of printed or handwritten text lines, elements of the document structure, and other graphic primitives), and background regions which usually have uniform or weakly textured brightness pro-

file. This problem is characterized by the unbalanced weights of these two classes of pixels, i.e. document image consists of mainly background rather than the contents. And a priori information about the document can significantly simplify the binarization. Thus, in many systems, prior to the binarization, the document type for each input image is being determined in order to select the appropriate binarization algorithm and its parameters, which helps to resolve the interpretation uncertainties of pixel groups in complex cases (e.g. [15]).

Unlike in automated identification document processing (where the templates are known beforehand), the preliminary document type categorization for hand-written and historical document image binarization is not feasible.

Because of the importance of this application, to track the developments in historical document binarization, a special competition Document Image Binarization Contest (DIBCO) has been regularly organized since 2009 [16]. To compare the proposed methods, the formal criteria for accuracy evaluation were established, but there are no restrictions on running time and memory consumption. Such conditions led to the dominance of the ANNs methods. Nevertheless, the fast binarization algorithms are becoming increasingly important, e.g. a new competition Document Image Binarization (DIB) [17] has been recently established, which judges algorithms not only by the accuracy, but also by the running time.

The winner of DIBCO 2017 is a neural network binarization method described in [18]. This method requires the scaling of an input image, and the optimal scaling coefficient depends on the size of the text symbol. Thus, the preliminary scaling coefficient is a hidden parameter of this method. In other similar works, such an analysis was not performed. But we can assume that similar methods require normalization of the input data. Given that, the number of normalization parameters should be considered to be the number of method parameters. Document images are not self-similar and their normalization usually includes scaling [19], thus the effect demonstrated in [18] could be expected.

It must be noted that the histogram-based methods, which include Otsu method, are scaling-invariant and do not require any additional preprocessing.

Thus, it is interesting to analyze the accuracy of the document image binarization employing one of the fastest methods – the zero-parametric Otsu method and its less known modification for cases where the weights of the classes to be separated are not considered equal [20]. We also consider the feasibility of the accuracy improvement for this algorithm employing an additional parameter of the scale of the images.

Further in this work, the known Otsu method modifications and the scope of its applicability are discussed. Pseudo F-Measure, the metric employed in DIBCO, is analyzed. We demonstrate that the unbalanced Otsu method modification [20] is better than the classical Otsu method by approximately 3 points in pseudo F-Measure

metric. Also, we suggest a novel one-parametric (with window size as a parameter) generalization of Otsu method for two features using fast method of binary clustering [21]. The proposed method, which employs brightness average over window as the second feature, allows for the image binarization improvement of unevenly illuminated documents and documents with spots by 15 points. In Fig. 1, we illustrate the examples of the classical Otsu method, its unbalanced modification and the proposed one-parametric modification performance on the image from DIBCO dataset.

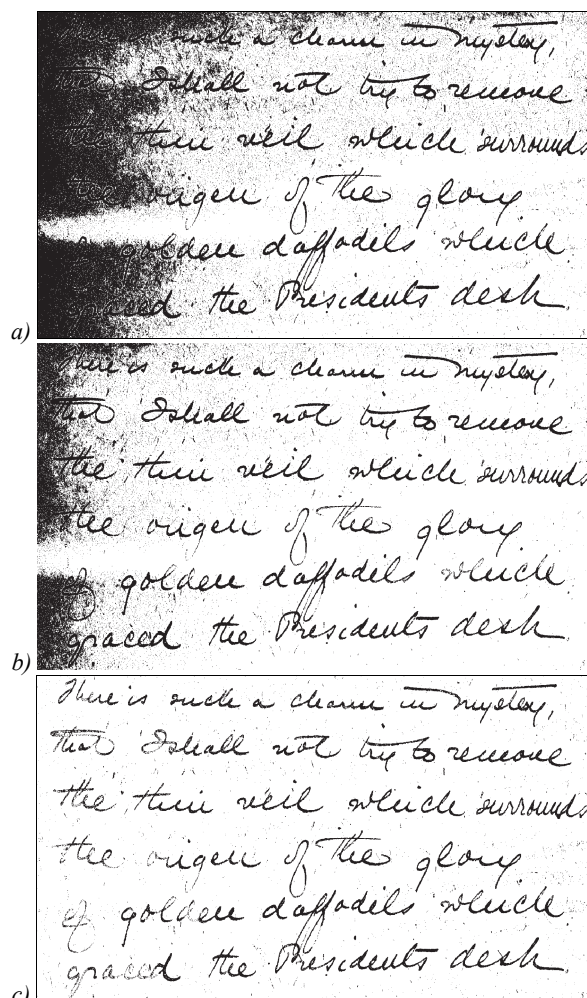


Fig. 1. The classical Otsu method (a), the unbalanced modification (b) and the proposed two-dimensional modification (c) performance on the hand-written document image

1. An overview of the known Otsu method modifications

Suggested in 1979, Otsu method was one of the most frequently mentioned thresholding methods in 2004 [22]. To date, when comparing the binarization algorithms, Otsu method is often used as the baseline. It is also true for DIBCO. In most cases, Otsu method demonstrates quite low accuracy, although sometimes it is not significantly worse than the competing algorithms. For example, in 2016, the best algorithm was better than the classical Otsu method by 5% [23], and in 2014, by 2% [24].

This can be associated with relative simplicity of the dataset (i.e. there were not many images with gradient illumination and overlapping distributions of background and text brightness).

To date, there are many known Otsu method modifications. The deviations involve the optimized criterion, search algorithms and even dimensionality of the analyzed feature space. Below we consider such modifications and their possible influence on the binarization accuracy in DIBCO.

Although there are many papers discussing Otsu method for various applications (the original paper on Otsu method was cited by approximately 32000 works in June 2019), there are not many works dedicated to its generalization. First, the paper by Kurita [20] should be acknowledged. It systematically describes the approach to the separability criterion search based on the maximum likelihood under assumption that the histogram is a mixture of two components, specifically, normal distributions. Three cases are considered: weights and components variances are equal (classical Otsu method), weights and components variances are arbitrary (the Kittler–Illingworth method [25]), and arbitrary weights while variances are equal (unbalanced Otsu method). It is worth mentioning that the classical Otsu method and the Kittler–Illingworth method are significantly more often featured in works than the unbalanced Otsu method, and thus, the first two are better studied than the latter. In particular, it is a widely known fact that the Kittler–Illingworth method is not stable (because it has too many degrees of freedom for a histogram-based method). To increase its robustness, in work [26], the authors suggested to introduce the restrictions for the target parameters.

The works by research group from King Saud University in Saudi Arabia should be also acknowledged. In [27], the authors suggested a separation criterion under the assumption that the mixture components follow lognormal distribution or Gamma distribution instead of the normal distribution. They demonstrated that under such assumptions the binarization accuracy could be increased, but the testing dataset was not large (only 13 images). In their other work [28], the authors proposed an approach to accelerate the classical Otsu method. They suggested to calculate the Otsu criterion only at local minima of the histogram, thus limiting the required computations. Prior to this work, the authors have published [29], which suggests the recursive binarization approach, where for the left part (in relation to the Otsu threshold) of the histogram the binarization threshold was calculated again in order to separate text from the complex background.

The development of Otsu method in certain cases allows for the binarization accuracy increase without involvement of the additional features. Nevertheless, there are certain effects, which are not compatible with any global threshold method. Let us consider the following list:

1. non-uniform lighting;
2. paper fold outlines;

3. ink spots and stains;
4. textured (“heraldic”) background;
5. high level of noise.

Under any condition from this list, the brightness of some regions of the background may be lower than that of the symbols in other regions, and then the proper binarization via the global threshold is not feasible. In such cases, the local binarization methods are used. One of the most popular local binarization methods was proposed by Niblack [11]. It classifies pixels globally via certain threshold function, and not in one-dimensional brightness space, but in three-dimensional feature space “brightness – its average over window – its window variation”. Parameters of the classifier are not determined by the distribution in the image, but should be set beforehand, which makes this method unstable, i.e. vulnerable to changes in the external conditions. Otsu method does not have any limitations regarding the input histogram dimensionality, which theoretically allows for the employment of the same features, and for the automatic search of the classifier parameters. Mainly due to this fact, there are generalizations for Otsu method for two-dimensional and even three-dimensional histograms.

For the first time, such a generalization for Otsu method was described in [30]. The authors suggested the image binarization via the analysis of the two-dimensional histogram, the first axis of which is the pixel brightness, and the second axis is the local average over window. Later in [31], Jian Gong et al. suggested the fast computation of the two-dimensional Otsu criterion described in [30] via cumulative summation: with cumulative histogram image computed once, it is feasible within the fixed timeframe to calculate weight (sum of all histogram pixels) of any orthotropic rectangle (which was used as the separating surface) with its vertex at the origin. This is also true for different statistics (mean, variance, and etc.). These statistics are further employed to calculate the criterion of the separation into two classes, rectangle and its complement, within the fixed timeframes. Later, in [32, 33] independently, there was suggested the same modification of the Jian Gong method: the criterion was calculated only via the most informative (according to the authors) part of the histogram. In 2016, in [34], it was demonstrated that the separation criterion calculation via two parameters (when the separating surface is a rectangle) can be performed independently via each parameter, which significantly reduces the computational time for the optimal separation without accuracy loss. Later, independently from the previous work, this approach was used in the paper dedicated to the development of the three-dimensional Otsu method [35].

Starting with Jian Gong’s research and later on, many authors work with rectangles, because the attempts to develop a fast brute-force search of the separating lines parameters failed, while lines are the optimal separation surfaces for equal covariance matrices in accordance with the original assumptions of Otsu method. An alternative

two-dimensional histogram-based binarization approach is described in [36]. The authors suggested that the pixel brightness and the sample mean do not correlate, thus the pixels distribution on the two-dimensional histogram is stretched along the diagonal containing zero. Based on this observation, the computations acceleration was suggested: the one-dimensional binarization performed via the projection of the two-dimensional histogram onto the diagonal.

Besides the modifications approach, there are some papers which suggest acceleration of the Otsu criterion calculation based on the genetic optimization algorithms [37] and GPU-accelerated computations [38].

Currently, Otsu method and the Kittler–Illingworth method are more often employed as part of some combination of the methods. In particular, Otsu threshold is used as an input parameter during the pre-processing [39, 40], for the binarization of the “simple” images [15, 41–43], and for global threshold value computation [44].

2. Test datasets and accuracy measure

For this work, we selected 150 images from the following datasets:

1. DIBCO Dataset (Document image binarization competition) [13, 16, 23, 24].
2. PHIBC Dataset (Persian heritage image binarization competition) [45].
3. The Nabuco Dataset [46].
4. LiveMemory Dataset [47].

The first three sets include images of the historical hand-written and printed documents with various distortions (inconsistent paper color, ink fading, blots, spots, seals, text showing through on the other side of the sheet, etc.). LiveMemory Dataset includes scanned proceedings of technical literature. Here, the main problems are symbols bleeding through on the other side and darkened paper fold outlines. It turned out that not all of the images from these datasets have pixelwise ground truth, and such images were not used. Thus, the compiled set included 150 images. Then it was separated into three subsets according to the image contents: 1) “simple” group includes 32 images; according to their annotation, they can be sufficiently (with accuracy over 95%) binarized via global thresholding; 2) “complex” – the rest of the images; 3) “spots and shadows” – subgroup of “complex” group, which includes images with spots and shadows (21 images). The description of such a set and its separation is available at [48].

We evaluate the accuracy of the methods via one of DIBCO’s metrics. Let us describe it. In document image binarization, the recall is defined as the ratio of the number of pixels labeled correctly over all the possible correct pixels corresponding to the ground truth, including those that were not labeled right but that should be included in the result (false negatives). Precision is defined as the ratio of the number of all correct values retrieved over the total number of the values retrieved. F-Measure (harmonic

mean of recall and precision) is commonly used to rank classification algorithms.

According to [49], F-measure is not a proper way to evaluate binarization quality, since it does not take into account the localization of the mislabeled pixel in relation to the document contents. For example, the loss of the symbol’s pixel is worse than the background pixel labeled as text. Thus, the author suggested the following modifications to the metric: pseudo-recall, pseudo-precision, and pseudo F-Measure.

Pseudo-recall is calculated as follows:

$$R_{ps} = \frac{\sum_{x,y} B(x,y) \cdot G_w(x,y)}{\sum_{x,y} G_w(x,y)}, \quad (1)$$

where $B(x,y)$ is a binary image, $G_w(x,y)$ is a segmented binary image multiplied with a weight mask. The latter is formed in a way that grants false negative pixels in text regions for an ideal result the lesser penalty the thicker the line width of the symbol is and the larger the distance from the border is. The weighted accuracy is defined as follows:

$$P_{ps} = \frac{\sum_{x,y} G(x,y) \cdot B_w(x,y)}{\sum_{x,y} B_w(x,y)}, \quad (2)$$

where $G(x,y)$ is a segmented binary image, $B_w(x,y) = B(x,y)P_w(x,y)$, and $P_w(x,y)$ is a weight mask tuning penalties for background false negatives based on the distance to a symbol. The goal of weighting is to increase the penalties for noise around a symbol and for the merged symbols. Thus, the pseudo F-Measure is defined as follows:

$$F_{ps} = \frac{2R_{ps}P_{ps}}{R_{ps} + P_{ps}}. \quad (3)$$

This metric is employed in DIBCO. We will use it further in this work.

Even though F_{ps} became the standard approach to quality evaluation in document image binarization, this measure has certain drawbacks. In Fig. 2, we illustrate two images (possible binarization results) with different distortions, each featuring comparable numbers of black pixels. In the first image, there is a crossed-out piece of text, while its quality evaluation is $F_{ps} = 98.34$. And in the second image, there is a thick horizontal line added onto the white background, and the quality evaluation is 97.22. Thus, according to F_{ps} , it could be concluded that in some cases the information loss is not as important as the existence of the artifacts of the same scale. This is a very subtle effect, and it was not observed during the accuracy measurements of the considered methods on the selected set of images.

It is important to note that the structure of the Pseudo F-Measure is complex. The methods based on the ANNs can adapt to this measure and get the binarization accura-

cy increase which will not reflect any real improvements. The “engineered” methods requiring fewer parameters (including Otsu method), on the contrary, cannot adapt to this metric, and thus, are less sensitive to its drawbacks.

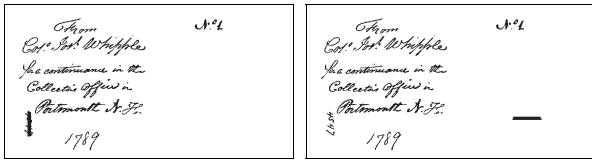


Fig. 2. The pair of the distorted images: the left image includes a crossed-out piece of text; the right image includes the black rectangle on the background

3. One-dimensional criteria of Otsu binarization

In the original work [10], Otsu suggested the minimization of the between-class variance. The latter is defined as follows:

$$Q(k) = \sigma_b^2(k) = \omega_1(k)\omega_2(k)(\mu_1(k) - \mu_2(k))^2, \quad (4)$$

where k is threshold level, and $\omega_1, \omega_2, \mu_1, \mu_2$ are the weights and the sample means of the first and second classes respectively. One of the key results of [10] is the formula for the separability criterion: the author demonstrated that the separability criterion search can be described in only the first-order cumulative moments.

The separability criterion for the case when the populations of object and background classes are unbalanced was suggested in [20]:

$$Q(k) = \sum_{i=1}^2 \omega_i(k) \ln(\omega_i(k)) - \ln(\sigma_w), \quad (5)$$

where i is the index of the class, σ_w is the within-class variance, which is defined as follows:

$$\sigma_w^2 = \sum_{i=1}^2 \omega_i(k) \cdot \sigma_i^2. \quad (6)$$

This criterion suits document images processing better, since the number of background pixels in document images almost always significantly surpasses the number of text pixels. This is confirmed by the experiments. For the entire selected dataset, the mean F_{ps} was 87.57 for the unbalanced classes criterion, versus 84.3 for the classical one.

Tab. 1 demonstrates that the unbalanced Otsu method for all sets, except for the “simple” one, gives an average increase of 4%, and for the “simple” set the methods perform similarly. These results suggest that cancelling equal weight restriction for text and background classes allows for a higher binarization quality even for the complex images. Tab. 2 demonstrates two clear examples of the different results for binarization via classical and unbalanced criteria.

These results suggest that the further increase of the freedom degrees would allow for even higher binarization accuracy. However, on the contrary, the further generalization

of the model (5) led to worse measure values: F_{ps} for Kittler–Illingworth algorithm for the entire images set was 73.69, i.e. 10% less than for the classical Otsu method, and 13% less, than for the unbalanced one. This is consistent with the previously reported results, in particular, with Haralick et al. findings [26].

Table 1. The comparison of classical and unbalanced Otsu criteria via F_{ps} measure

	Classical Otsu, %	Unbalanced Otsu, %
Entire set	84.30	87.57
“Spots and shadows”	69.82	73.91
“Complex”	80.56	84.10
“Simple”	98.38	98.59

The maximum accuracy for the global threshold binarization methods for the entire set is 93.76. This value is a result of the optimal thresholds (for each image from the dataset) set via manual segmentation. It surpasses the balanced Otsu by 9.46, and the unbalanced by 6.19 percent points. Thus, for the test data, replacement of the balanced criterion with the unbalanced one is significant in terms of the accuracy, since it reduces the number of pixel classification errors by a factor of 1.5.

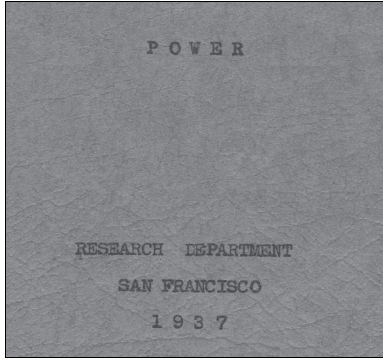
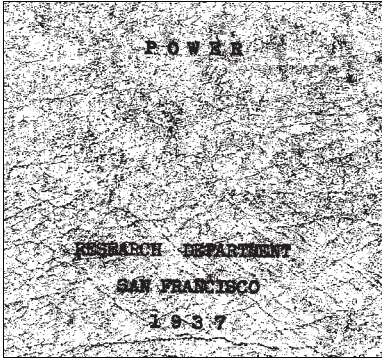
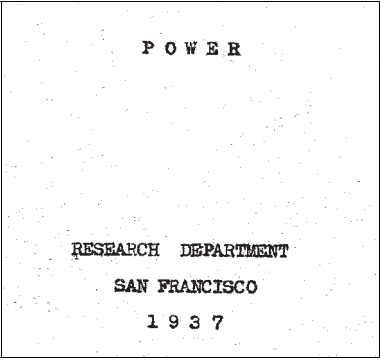
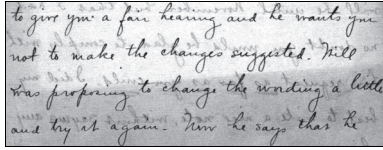
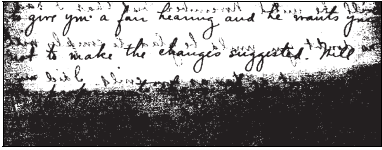
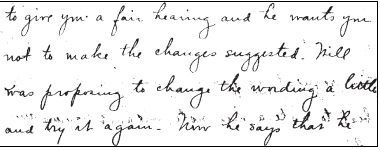
For further accuracy increase, there is a need for either new criteria of global threshold calculation, or for changes to the existing model. In the next section, we suggest the changes to the model, increasing the dimensionality of the problem, where the optimum threshold is computed via two features.

4. The Otsu method for image binarization via two features

Let us consider the class of images, which incorporate spots or inhomogeneous background (sample images are illustrated in Fig. 3). Such images make up approximately 15% of the considered set of images, which is significant. As shown in Section 3, utilization of both classical and unbalanced Otsu criteria for brightness histograms of such images results in unsatisfactory binarization accuracy (69.82% and 73.91% respectively). In this section, we suggest Otsu method for gray images binarization based on two features, allowing for successful binarization of the mentioned images.

Let us assign a two-dimensional vector to each pixel p of the image I to be binarized: $v(p) = \langle i, s \rangle$, where i is the intensity value of this pixel, and s is the local average intensity in the vicinity of δ , centered around p . Let us build the joint histogram H for the input image I based on the mentioned features. It can be considered as the digital image, where the vertical axis corresponds to i values, and horizontal axis corresponds to s values, and the intensity of the pixels corresponds to the number of vectors $v = \langle i, s \rangle$ in image I . Then the image binarization can be performed via the optimal separating surface in the joint histogram H image.

Table 2. The sample performance of the classical and the unbalanced Otsu methods

The input image	The classical Otsu criterion	The unbalanced Otsu criterion
 <p>a)</p>	 <p>b)</p>	 <p>c)</p>
 <p>d)</p>	 <p>e)</p>	 <p>f)</p>

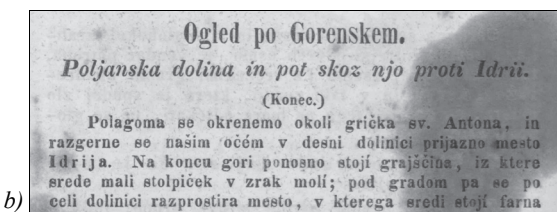
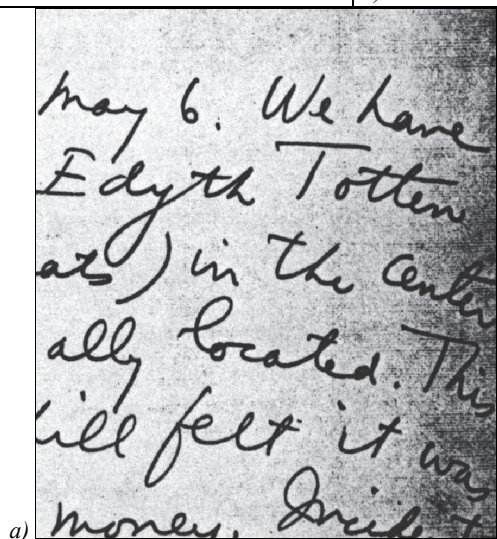


Fig. 3. Sample images featuring spots and inhomogeneous background

Let us consider two criteria for optimal separation of the histogram H into regions of two classes. These criteria, based on the maximum likelihood principle, are the generalization of one-dimensional unbalanced criterion for the case of two model features. The first of the latter is optimal for the unbalanced sets and covariation matrix with equal eigenvalues:

$$A_1(\theta) = \arg \max_{\theta \in \Omega} \sum_{j=1}^2 \omega_j(\theta) \ln \omega_j(\theta) - \ln(D_1(\theta) + D_2(\theta)), \quad (7)$$

where Ω is the set of feasible separations in this model, θ are the parameters of a single separation option, ω_j are the weights of the separated classes, and D_j are the variances of their values. For the optimum second criterion, the covariation matrices of the mixture components without eigenvalues ratio restriction should be equal:

$$A_2(\theta) = \arg \max_{\theta \in \Omega} \sum_{j=1}^2 \omega_j(\theta) \ln \omega_j(\theta) - \frac{1}{2} \ln \det \left(\sum_{j=1}^2 S_j(\theta) \right), \quad (8)$$

where S_j are the sample covariation matrices of each of the resulting classes.

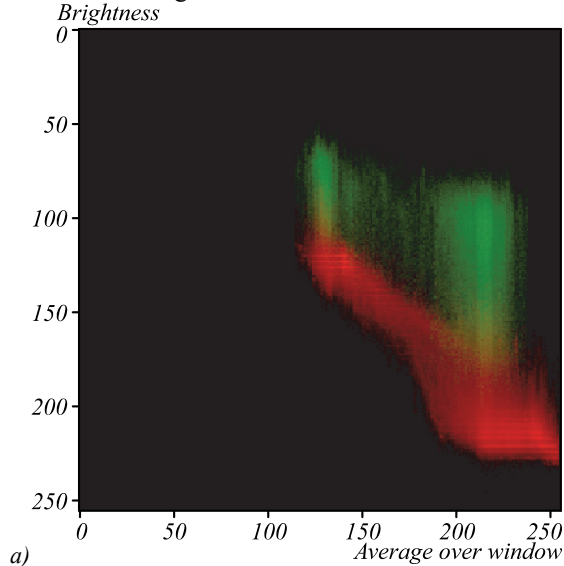
Fig. 4 illustrates the histograms for the images demonstrated in Fig. 3. They are visualized via three channel images, with red channel corresponding to the histogram of the background pixels, and green channel corresponding to the text pixels histogram. The cluster colors were obtained via segmentation. Fig. 4a, b demonstrate that the background and text clusters are elongated. These figures also demonstrate that the assumptions of criterion (7) are not met. But there is no indication that the criterion (8) could perform better, since the distributions are not uniform and are not coaxial (fig. 4b). Thus, the criterion yielding the higher accuracy results can be determined only via computational experiments.

Let us consider different sets of feasible separations Ω . For the global threshold segmentation based on the pixel intensity, set Ω_j includes the segmentations via lines parallel to the x-axis and constructed by the only parameter $\theta = \langle I_{\max} \rangle$. Then the set of text pixels is defined as $C_T = \{p(i,s) | i \geq I_{\max}\}$, and the set of background pixels is defined as $C_B = \{p(i,s) | i < I_{\max}\}$. The listed examples show that a horizontal line cannot sufficiently divide text and background into two classes.

The set of the Jian Gong method segmentations is the set of all possible orthotropic rectangles (i.e. rectangles with sides, parallel to the coordinate axes) with one of the

vertices at the origin. These rectangles can be constructed via two parameters: $q = \langle I_{\max}, S_{\max} \rangle$, where I_{\max} is the rectangle width, and S_{\max} is its height. After such separation, the following pixel set is assigned to the text class: $C_T = \{p(i, s) | i \leq I_{\max} \wedge s \leq S_{\max}\}$, and the following pixel set is assigned to the background class: $C_B = \{p(i, s) | i > I_{\max} \vee s > S_{\max}\}$. Fig. 4 demonstrates that such separation would result in classification errors.

Let us consider set Ω_L , containing lines, crossing the two-dimensional histogram. These lines can be set via two



parameters: $\theta = \langle k, i_0 \rangle$. Then the background class consists of the pixel set $C_B = \{p(i, s) | i \geq ks + i_0\}$, and the text class consists of $C_T = \{p(i, s) | i < ks + i_0\}$. The figures show that for two-dimensional histograms, there could be such slant lines selection, which would allow for good segmentation.

In [21], the algorithm for fast computation of (7) and (9) criteria on Ω_L segmentation set was suggested. Let us adopt it for the image binarization algorithm. The inputs of the algorithm are: the gray image I , and separation criterion A . The algorithm includes the following steps:

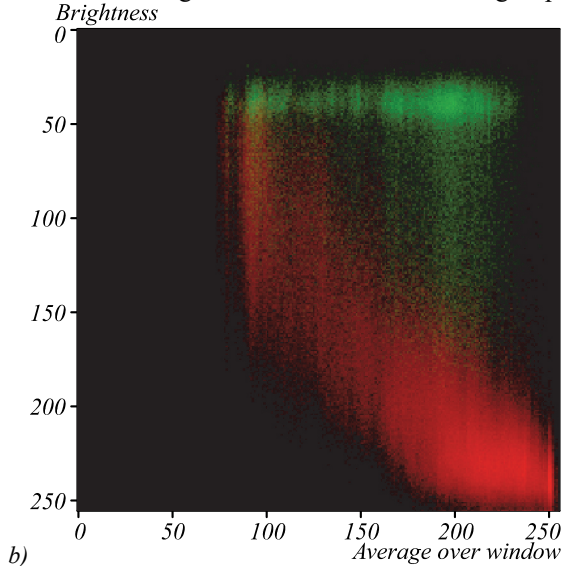


Fig. 4. The histograms of images illustrated in fig. 3

Step 1: For the input image I with sides w and h , the histogram H is built based on the features $v = \langle i, s \rangle$. The feature s is calculated via squared window with the side of $\delta = 0,1 \cdot \min(w, h)$. The number of the histogram bins is 256 for each axis.

Step 2: The criterion A is calculated via linear binarization clustering adopted from [21], and the parameters of the separating line $\langle k, i_0 \rangle$ are defined (the detailed description and pseudocode are given in [21]). It is worth to mention that linear binarization clustering method utilize fast Hough transform algorithm, which has wide range of application, such as [50]. The considered set of the separating lines is limited in the way that assures an angle between each line and the horizontal axis is less than 30 degrees. This allows for a higher algorithm performance.

Step 3: For each pixel $p \in I$, the vector of features $v(p)$ is calculated. Based on the computed parameters $\langle k, i_0 \rangle$ of the line, p is assigned to either background class or text class according to the rule above.

It must be noted that in Step 1 the choice of the histogram binning scale is not specified: there can be used either the scale of absolute values or the range between minimum and maximum calculated features. For criteria A_1 and A_2 , the choice of the scale affects binarization accuracy differently. The results reflect the accuracy for A_1 in terms of the scale with absolute values, and for A_2 in terms of the range, calculated via features values.

Let us evaluate the complexity of the suggested algorithm. The computations of Step 1 and Step 3 require $O(n, m)$, where n and m are the linear size parameters of the image. The complexity of Step 2 depends of the histogram size. Employing the method, suggested in [21], the complexity of the binarization criterion computation for all lines and its optimal value search is $O(h^2 \log h)$, where h is linear size of the histogram (as previously mentioned, $h = 256$). Thus, the total complexity is $O(h^2 \ln h + nm)$, which is comparable with Jian Gong algorithm complexity – $O(h^2 + nm)$ [31].

Fig. 5 demonstrates performance results on the images with complex background of the suggested algorithm for criteria A_1 and A_2 , and of the one-dimensional algorithm for the unbalanced criterion.

Tab. 3–5 demonstrate the results of the one-dimensional Otsu method for the unbalanced criterion, and of the proposed Otsu method for two-dimensional histogram for criteria A_1 and A_2 .

Tab. 3 shows that for the entire set, A_1 works the best. This fact suggests that the employment of the average over window as the additional feature allows for a higher binarization accuracy of the global threshold function (line) in two-dimensional feature space “brightness – its average over window”. Tab. 4 shows that all three binarization methods grant similarly accurate results for the “simple” images set.

Table 3. F_{ps} values for the entire set of images

	Mean	Median	Variance
Unbalanced Otsu	87.57	91.98	14.03
Criterion	91.22	93.90	10.35
Criterion	89.63	93.53	13.61

Table 4. F_{ps} values for "simple" set of images

	Mean	Median	Variance
Unbalanced Otsu	98.59	98.64	0.91
Criterion A_1	98.64	98.64	0.90
Criterion A_2	98.55	98.51	0.74

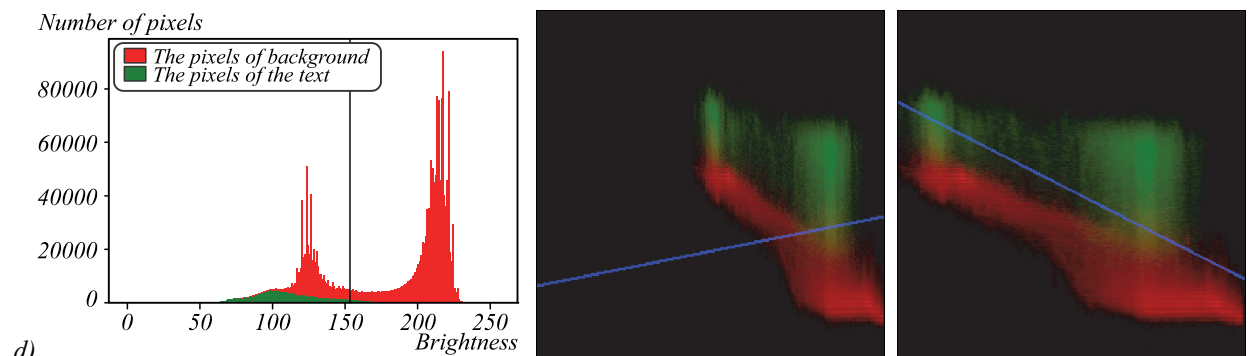
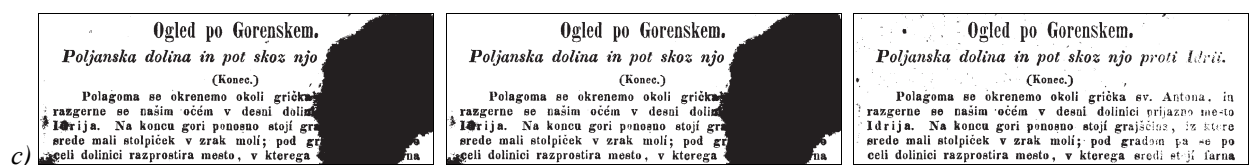
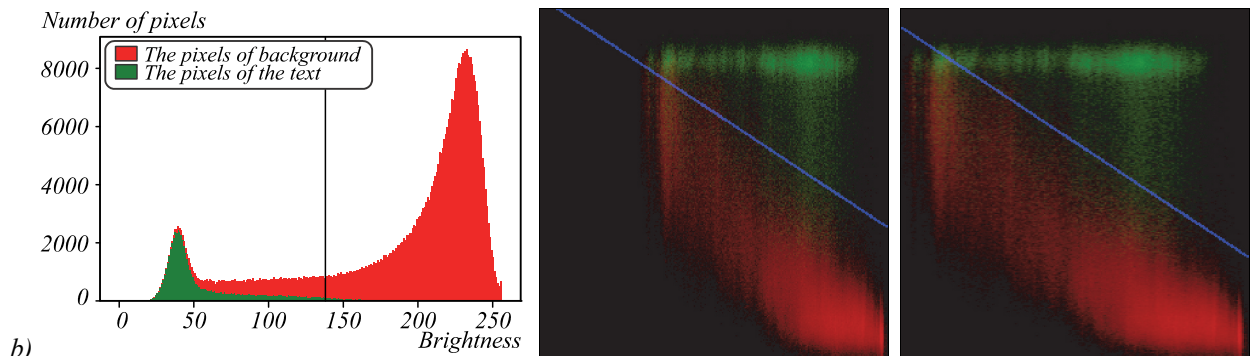
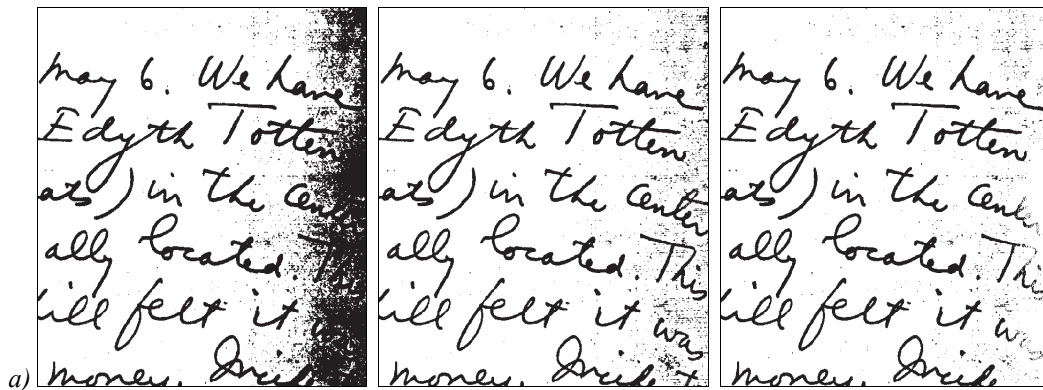


Fig. 5. The binarization algorithms performance results. The first column in a) and c) panels demonstrates the image binarization results by Otsu method with the unbalanced criterion; b) and d) panels demonstrate histograms and binarization threshold for a) and c) respectively. Similarly, results for two-dimensional modification of the method with criterion A_1 in the second column, and criterion A_2 in the third column, are presented

Tab. 5 details the binarization accuracy on “spots and shadows” dataset. It shows that no matter which criterion is used, two-dimensional method is more accurate. It is worth mentioning that the criterion A_2 is better for the image binarization of documents with uneven illumination or spots.

Table 5. F_{ps} values for “spots and shadows” set of images

	Mean	Median	Variance
Unbalanced Otsu	73.91	77.17	18.84
Criterion A_1	84.06	88.85	17.26
Criterion A_2	89.64	90.12	6.02

Conclusion

This work considers document image binarization via Otsu method with classical (balanced) and unbalanced criteria. Pseudo F-Measure on the set of 150 test images showed that the accuracy of these methods was 84.3 % and 87.57% respectively (the maximum Pseudo F-Measure value of this global binarization model was 93.76%). We demonstrated that for document image binarization the lesser-known Otsu method with unbalanced criterion allows for a higher binarization accuracy. The international document binarization competitions in certain years (when the dataset was suitable for global binarization) showed that one-dimensional Otsu method with unbalanced criterion was comparable with the best competing algorithms in terms of the accuracy, which confirms Otsu method practical importance.

This work also demonstrates new generalization results for Otsu method in image binarization via two features. The proposed algorithm employs fast linear clustering of two-dimensional histogram for search of the parameters of the line separating background and text pixels on this histogram. We demonstrated that for many cases (if image includes inhomogeneous background or spots), the slant line is the optimal separating surface for two-dimensional histogram. Computational experiments for test datasets showed that Otsu method with two features employing the considered criteria allowed for average accuracy of 89.63 % and 91.22 % respectively.

References

- [1] Chaki N, Shaikh SH, Saeed K. A comprehensive survey on image binarization techniques. *Stud Comput Intell* 2014; 560: 5-15. DOI: 10.1007/978-81-322-1907-1_2.
- [2] Challa R, Rao KS. Efficient compression of binarized tainted documents. *Int J Adv Comput Sci Appl* 2018; 9(2): 663-667. DOI: 10.26483/ijarcs.v9i2.5520.
- [3] Arlazarov VL, Emel'janov NE, ed. *Razvitie bezbumazhnoj tehnologii v organizacionnyh sistemah* [In Russian]. Moscow: "URSS" Publisher; 1999. ISBN: 5-8360-0097-2.
- [4] Fan H, Xie F, Li Y, Jiang Z, Liu J. Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold. *Comput Biol Med* 2017; 85: 75-85. DOI: 10.1016/j.compbiomed.2017.03.025.
- [5] Vizilter YV, Gorbatevich VS, Vishnyakov BV, Sidyakin SV. Object detection in images using morphlet descriptions [In Russian]. *Computer Optics* 2017; 41(3): 406-411. DOI: 10.18287/2412-6179-2017-41-3-406-411.
- [6] Iskhakov AR, Malikov RF. Calculation of aircraft area on satellite images by genetic algorithm [In Russian]. *Vestnik JuUrGU. Ser. Matematicheskoe Modelirovanie i Programirovanie* 2016; 9(4): 148-154.
- [7] Mustafa WA, Haniza Y. Illumination and contrast correction strategy using bilateral filtering and binarization comparison. *J Telecommun Electron Comput Eng* 2016; 8(1): 67-73.
- [8] Kokhan V, Grigoriev M, Buzmakov A, Uvarov V, Ingacheva A, Shvets E, Chukalina M. Segmentation criteria in the problem of porosity determination based on CT scans. *Proc SPIE* 2020; 11433: 114331E. DOI: 10.1117/12.2558081.
- [9] Fedorenko VA, Sidak EV. Method of the binarization of images of traces on the shot bullets for the automatic assessment of their suitability to identification of the firearms [In Russian]. *Informacionnye Tehnologii i Vychislitel'nye Sistemy* 2016; 3: 82-88.
- [10] Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern Syst* 1979; 9(1): 62-66. DOI: 10.1109/tsmc.1979.4310076.
- [11] Niblack W. *An introduction to digital image processing*. Englewood Cliffs: Prentice Hall; 1986.
- [12] Aliev MA, Nikolaev DP, Saraev AA. Postroenie bystryh vychislitel'nyh shem nastrojki algoritma binarizacii Nibljeka [In Russian]. *Trudy ISA RAN* 2014; 64(3): 25-34.
- [13] Pratikakis I, Zagoris K, Barlas G, Gatos B. ICDAR2017 Competition on document image binarization (DIBCO 2017). 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) 2017: 1395-1403. DOI: 10.1109/icdar.2017.228.
- [14] Calvo-Zaragoza J, Gallego A-J. A selectional auto-encoder approach for document image binarization. *Patt Recogn* 2018; 86: 37-47. DOI: 10.1016/j.patcog.2018.08.011.
- [15] Chen Y, Leedham G. Decompose algorithm for thresholding degraded historical document images. *IEE Proceedings – Vision, Image and Signal Processing* 2005; 152(6): 702-714. DOI: 10.1049/ip-vis:20045054.
- [16] Gatos B, Ntirogiannis K, Pratikakis I. ICDAR 2009 Document image binarization contest (DIBCO 2009). 10th International Conference on Document Analysis and Recognition 2009: 1375-1382. DOI: 10.1109/icdar.2009.246.
- [17] Document image binarization. Source: <https://dib.cin.ufpe.br>.
- [18] Bezmaternykh PV, Ilin DA, Nikolaev DP. U-Net-bin: hacking the document image binarization contest. *Computer Optics* 2019; 43(5): 825-832. DOI: 10.18287/2412-6179-2019-43-5-825-832.
- [19] Tropin DV, Shemyakina YA, Konovalenko IA, Faradzhev IA. Localization of planar objects on the images with complex structure of projective distortion [In Russian]. *Informacionnye Processy* 2019; 19(2): 208-229.
- [20] Kurita T, Otsu N, Abdelmalek N. Maximum likelihood thresholding based on population mixture models. *Patt Recogn* 1992; 25(10): 1231-1240. DOI: 10.1016/0031-3203(92)90024-d.
- [21] Ershov EI. Fast binary linear clustering algorithm for small dimensional histograms [In Russian]. *Sensory systems* 2017; 31(3): 261-269.
- [22] Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. *J Electron Imaging* 2004; 13(1): 146-165. DOI: 10.1117/1.1631315.

- [23] Pratikakis I, Zagoris K, Barlas G, Gatos B. ICFHR2016 handwritten document image binarization contest (H-DIBCO 2016). 15th International Conference on Frontiers in Handwriting Recognition (ICFHR) 2016: 619-623. DOI: 10.1109/ICFHR.2016.0118.
- [24] Ntirogiannis K, Gatos B, Pratikakis I. ICFHR2014 competition on handwritten document image binarization (H-DIBCO 2014). 14th International conference on frontiers in handwriting recognition 2014: 809-813. DOI: 10.1109/ICFHR.2014.141.
- [25] Kittler J, Illingworth J. Minimum error thresholding. *Patt Recogn* 1986; 19(1): 41-47. DOI: 10.1016/0031-3203(86)90030-0.
- [26] Sungzoon C, Haralick R, Yi S. Improvement of Kittler and Illingworth's minimum error thresholding. *Patt Recogn* 1989; 22(5): 609-617. DOI: 10.1016/0031-3203(89)90029-0.
- [27] AlSaeed DH, Bouridane A, ElZaart A, Sammouda R. Two modified Otsu image segmentation methods based on Lognormal and Gamma distribution models. *International Conference on Information Technology and e-Services* 2012: 1-5. DOI: 10.1109/ICITeS.2012.6216680.
- [28] AlSaeed DH, Bouridane A, ElZaart A. A novel fast Otsu digital image segmentation method. *Int Arab J Inf Technol (IAJIT)* 2016; 13(4): 427-433.
- [29] Cheriet M, Said JN, Suen CY. A recursive thresholding technique for image segmentation. *IEEE Trans Image Process* 1998; 7(6): 918-921. DOI: 10.1109/83.679444.
- [30] Liu J, Li W, Tian Y. Automatic thresholding of gray-level pictures using two-dimensional Otsu method. *International Conference on Circuits and Systems* 1991: 325-327. DOI: 10.1109/ciccas.1991.184351.
- [31] Gong J, Li L, Chen W. Fast recursive algorithms for two-dimensional thresholding. *Patt Recogn* 1998; 31(3): 295-300. DOI: 10.1016/S0031-3203(97)00043-5.
- [32] Lu C, Zhu P, Cao Y. The segmentation algorithm improvement of a two-dimensional Otsu and application research. *Software Technology and Engineering (ICSTE)* 2010; 1: V1-76-V1-79. 3-5. DOI: 10.1109/ICSTE.2010.5608908.
- [33] Chen Q, Zhao L, Lu J, Kuang G, Wang N, Jiang Y. Modified two-dimensional Otsu image segmentation algorithm and fast realization. *Image Process* 2012; 6(4): 426-433. DOI: 10.1049/iet-ipr.2010.0078.
- [34] Sha C, Hou J, Cui H. A robust 2D Otsu's thresholding method in image segmentation. *J Vis Commun Image Represent* 2016; 41: 339-351. DOI: 10.1016/j.jvcir.2016.10.013.
- [35] Zhang X, Zhao H, Li X, Feng Y, Li H. A multi-scale 3D Otsu thresholding algorithm for medical image segmentation. *Digit Signal Process* 2017; 60: 186-199. DOI: 10.1016/j.dsp.2016.08.003.
- [36] Zhang J, Hu J. Image segmentation based on 2D Otsu method with histogram analysis. *2008 International Conference on Computer Science and Software Engineering* 2008: 6: 105-108. DOI: 10.1109/CSSE.2008.206.
- [37] Zhang Y, Zeng L, Zhang Y, Meng J. 2D Otsu segmentation algorithm improvement based on FOCPSO. *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCLOUD/SocialCom/SustainCom)* 2018; 809-815. DOI: 10.1109/BDCLOUD.2018.00121.
- [38] Zhu X, Xiao Y, Tan G, Zhou Sh, Leung C-S, Zheng Y. GPU-accelerated 2D OTSU and 2D entropy-based thresholding. *J Real Time Image Process* 2020; 17: 993-1005. DOI: 10.1007/s11554-018-00848-5.
- [39] Nafchi HZ, Moghaddam RF, Cheriet M. Phase-based binarization of ancient document images: Model and applications. *IEEE Trans Image Process* 2014; 23(7): 2916-2930. DOI: 10.1109/TIP.2014.2322451.
- [40] Bolotova YA, Spitsyn VG, Osina PM. A review of algorithms for text detection in images and videos. *Computer Optics* 2017; 41(3): 441-452. DOI: 10.18287/2412-6179-2017-41-3-441-452.
- [41] Boudraa O, Hidouci WK, Michelucci D. A robust multi stage technique for image binarization of degraded historical documents. *5th International Conference on Electrical Engineering – Boumerdes (ICEE-B)* 2017. DOI: 10.1109/ICEE-B.2017.8192044.
- [42] Farrahi MR, Cheriet M. AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization. *Patt Recogn* 2012; 45(6): 2419-2431. DOI: 10.1016/j.patcog.2011.12.013.
- [43] Bolotova YU, Spitsyn VG, Rudometkina MN. License plate recognition algorithm on the basis of a connected components method and a hierarchical temporal memory model. *Computer Optics* 2015; 39(2): 275-280. DOI: 10.18287/0134-2452-2015-39-2-275-280.
- [44] Lech P, Okarma K. Binarization of document images using the modified local-global Otsu and Kapur algorithms. *Przegląd Elektrotechniczny* 2015; 1(2): 73-76. DOI: 10.15199/48.2015.02.18.
- [45] Ayatollahi SM, Nafchi HZ. Persian heritage image binarization competition (PHIBC 2012). *First Iranian Conference on Pattern Recognition and Image Analysis (PRIA)* 2013. DOI: 10.1109/PRIA.2013.6528442.
- [46] Lins RD. Nabuco – Two decades of document processing in Latin America. *J Univers Comput Sci* 2011; 17(1): 151-161. DOI: 10.3217/jucs-017-01-0151.
- [47] Lins RD, Silva G, Torreão G. Content recognition and indexing in the LiveMemory platform. In Book: Ogier J-M, Liu W, Lladós J, eds. *Graphics recognition. Achievements, challenges, and evolution*. Berlin, Heidelberg, New York: Springer; 2009: 220-230. DOI: 10.1007/978-3-642-13728-0_20.
- [48] Ershov E, Korchagin S. Description of the collected dataset. 2019. Source: ftp://vis.iitp.ru/dataset_description.txt.
- [49] Ntirogiannis K, Gatos B, Pratikakis I. Performance evaluation methodology for historical document image binarization. *IEEE Trans Image Process* 2012; 22(2): 595-609. DOI: 10.1109/TIP.2012.2219550.
- [50] Bocharov DA. A linear regression method robust to extreme stationary clutter [In Russian]. *Sensory Systems* 2020; 34(1): 44-56. DOI: 10.31857/S0235009220010059.

Authors' information

Egor Ivanovich Ershov (b. 1990) received a Ph.D. degree from Radio Engineering and Computer Technology Faculty in 2019. He is a senior researcher in Vision System Laboratory in Institute of Information Transmission Problems (Kharkevic Institute) of the Russian Academy of Science (IITP RAS). His main areas of research are image processing and analysis, particularly color computer vision, color reproduction technologies, colorimetry, human vision system, Hough transform, binarization. He is a recipient of several awards for his scientific and professional work. E-mail: ershov@iitp.ru.

Sergey Andeevich Korchagin, (b. 1997), studying at Lomonosov Moscow State University. Works as junior researcher in Vision System Laboratory in Institute of Information Transmission Problems (Kharkevic Institute) of the Russian Academy of Science (IITP RAS). Research interests: image processing, digital and analog photography. E-mail: korchagin@visillect.com.

Vladislav Vladimirovich Kokhan, (b. 1995) graduated from Moscow Aviation University in 2017. Works as researcher at the IITP RAS, a researcher-programmer at Smart Engines Service LLC. His major research interests include Computed Tomography, Image Processing. E-mail: v.kokhan@smartengines.com.

Pavel Vladimirovich Bezmaternykh (b. 1987) graduated from Moscow Institute of Steel and Alloys in 2009. Works as a software engineer at Institute for Systems Analysis of RAS. Research interests are image processing and analysis, document recognition, text layout analysis. E-mail: bezmpavel@smartengines.com.

Received May 14, 2020. The final version – November 26, 2020.
