DOI: 10.15276/aait.01.2021.5
UDC 004.93

IMPROVEMENT OF THE COLOR TEXT IMAGE BINARIZATION METHOD USING THE MINIMUM-DISTANCE CLASSIFIER

Marina V. Polyakova[1)]
ORCID: org/0000-0001-7229-7657; marina_polyakova@rambler.ru
Alexandr G. Nesteryuk[1)]
ORCID: org/0000-0002-0806-8259; nesteryuk@opu.ua
[1)] Odessa National Polytechnic University, 1, Shevchenka Ave., Odesa, 65044, Ukraine

ABSTRACT

Optical character recognition systems for the images are used to convert books and documents into electronic form, to automate accounting systems in business, when recognizing markers using augmented reality technologies and etc. The quality of optical character recognition, provided that binarization is applied, is largely determined by the quality of separation of the foreground pixels from the background. Methods of text image binarization are analyzed and insufficient quality of binarization is noted. As a way of research the minimum-distance classifier for the improvement of the existing method of binarization of color text images is used. To improve the quality of the binarization of color text images, it is advisable to divide image pixels into two classes, "Foreground" and "Background", to use classification methods instead of heuristic threshold selection, namely, a minimum-distance classifier. To reduce the amount of processed information before applying the classifier, it is advisable to select blocks of pixels for subsequent processing. This was done by analyzing the connected components on the original image. An improved method of the color text image binarization with the use of analysis of connected components and minimum-distance classifier has been elaborated. The research of the elaborated method showed that it is better than existing binarization methods in terms of robustness of binarization, but worse in terms of the error of the determining the boundaries of objects. Among the recognition errors, the pixels of images from the class labeled "Foreground" were more often mistaken for the class labeled "Background". The proposed method of binarization with the uniqueness of class prototypes is recommended to be used in problems of the processing of color images of the printed text, for which the error in determining the boundaries of characters as a result of binarization is compensated by the thickness of the letters. With a multiplicity of class prototypes, the proposed binarization method is recommended to be used in problems of processing color images of handwritten text, if high performance is not required. The improved binarization method has shown its efficiency in cases of slow changes in the color and illumination of the text and background, however, abrupt changes in color and illumination, as well as a textured background, do not allowing the binarization quality required for practical problems.

Keywords: Image binarization; minimum-distance classifier; optical character recognition; color text image; image background; image foreground

***For citation*: Polyakova M. V., Nesteryuk A. G. Improvement of the Color Text Image Binarization Method Using the Minimum-Distance Classifier. *Applied Aspects of Information Technology*. 2021; Vol.4 No.1: 57–70. DOI: 10.15276/aait.01.2021.5**

FORMULATION OF THE PROBLEM

The widespread use of scanning and digital cameras has led to the development of optical character recognition (OCR) systems in images [1-2]. These systems are used to convert books and documents into electronic form [3-4], to automate accounting systems in business, when recognizing markers using augmented reality technologies [5]. A separate application is the processing of raster images of technical drawings [6]. The result of using the OCR is text that can be viewed and printed in appropriate editors, compared with database fields, and used to synthesize voice messages. Currently, the most successful OCR systems with a sufficiently high quality recognize texts of standard fonts in images with a low level of noise and distortion.

However, even if the percentage of correct character recognition is 99.9 %, on one page of text (1500 characters), on average, there are 1-2 errors [7].

One of the tasks of OCR systems is to separate the foreground pixels (text, tables, graphic objects) from the background pixels in order to reduce the amount of information with which one has to work [7]. This task is reduced to the binarization of the image, that is, the transformation of a color or grayscale image into a two-level (black and white) image. Then, as a result of binarization of color images of documents, the text is displayed in black on a white background, regardless of the color of the text and background in the document [7-8].

The quality of optical character recognition, provided that binarization is applied, is largely determined by the quality of separation of the foreground pixels from the background. This is due to the fact that gaps in characters or mistakenly detected spots can negatively affect the recognition result. As a result, currently, when processing color images of text, high requirements are imposed on the quality of binarization [3]. At the same time, in a significant part of the tasks of processing such images, it becomes necessary to carry out binarization with high speed [1]; [5]. The known methods of binarization of color images of text [8-9] are not able to simultaneously satisfy the requirements for the quality of binarization and for its speed. These characteristics of text image binarization methods used in OCR systems.

As a way of splitting images into blocks in the methods of binarization of text images, it is possible to use block or pixel processing [9-10], as well as the analysis of connected components [11-12]. Methods that use block processing are characterized by high performance, but the result of binarization depends on the division of the processed image into blocks [13-14]. At the same time, pixel processing requires a significant investment of time, but allows obtaining high quality binarization [9]. For methods with the analysis of connected components [11]; [13], the processing time and quality of binarization is determined by the result of the selection of connected components.

Depending on the procedure for making a decision on assigning image pixels to the foreground or background, threshold and classification methods can be distinguished among the binarization methods for text images. For both groups of methods, the speed and quality of binarization are determined by the properties of the processed images, however, threshold methods are usually faster than classification methods, while the latter often provide better binarization. Thus, at present there is a contradiction between the modern requirements for the quality and speed of binarization of color images of text and the capabilities of modern binarization methods.

ANALYSIS OF RECENT RESEARCH AND PUBLICATIONS

Until recently, thresholding was used to binaries text images. The main parameter of this transformation is the threshold *T*, with the value of which the intensity of all pixels is then compared. After comparison, the pixel is assigned one of two possible values: “Foreground” or “Background” [15-16].

In the literature, threshold binarization methods are conventionally divided into two groups: global and local (adaptive) [17]. In global binarization methods, the threshold *T*, with which the image is divided into “Foreground” and “Background”, is determined for the entire image. Such binarizations can be classified as methods with block processing if the entire image is taken as a block. In the global Otsu method, the threshold T is calculated, which minimizes the average binarization error, that is, the average error from deciding whether the image pixels belong to the foreground or background by minimizing the intraclass variance, which is defined as the weighted sum of the variances of two classes – “Foreground” and “Background” [10]. The Otsu method is adapted to different kinds of images by choosing the optimal threshold and gives good results with a low level of noise,

In [8], the noise immunity of the Otsu method was increased for color images of text, which were first processed by a bilateral filter, and then global threshold processing was applied to each of the *R*, *G*, *B* components. The bilateral filter parameters and the threshold for each of the color components were selected by optimizing the percentage of correct classification of background pixels. At the same time, the percentage of correct classification of the foreground pixels was fixed. Taking into account the selected thresholds, binarization of each of the *R*, *G*, *B* image a component was performed, after which the pixels of the binary images of these components were classified using a naive Bayesian classifier. In the result classification, it was decided in which of the *R*, *G*, *B* components the information about the foreground for a given pixel of the image is better represented.

In [18], to reduce the level of noise before binarization, preliminary processing was carried out first with a median filter, and then with two linear filters – high and low frequencies. The preprocessing parameters and the binarization threshold were determined using Bayesian hyperparameter optimization.

Local (adaptive) binarization methods divide the image into blocks of a certain size, while the block size must be sufficient to preserve the features and details of the image, as well as to reduce the effect of noise on the binarization result. Next, for each block, a threshold is calculated based on the pixel intensity. Thus, local methods use either block processing [14-15] or connected component analysis [11-12], and also [16].

In [13] Sauwol et al. proposed the choice of the local threshold $T(x, y)$ in the processing window centered at a point with coordinates (x, y) according to the formula:

$$T(x, y) = m(x, y)(1+k(s(x, y)/R-1)), \qquad (1)$$

where: $m(x, y)$ – is the average intensity of pixels in the window;

$s(x, y)$ – is the standard deviation of the intensity of these pixels;

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

k – is a constant that takes values in the range [0.2; 0.5];

R – is the range of values of the standard deviation of the pixel intensity, which is calculated over the entire image and enhances the influence of $s(x, y)$ on the threshold value.

Sauvol's method [13] with the choice of the threshold according to the formula (1) is widely applied to images with non-uniform intensity with thick lines and large objects, when processing thin intersecting lines, gaps may occur. This method is less robust to noise in the original image than, for example, the Otsu method. Also, Sauwola's method has difficulty in processing poorly lit images, especially in cases where the intensity values of the foreground and background pixels are close to each other.

In [14], binarization of text images was performed, on which such distortions and interferences as shadows, low contrast, changes in illumination, and blur were present. The values of the intensity of the original image in the neighboring background pixels were interpolated and the background image was constructed. Then the values of the original image were interpolated and the threshold was determined by comparison with the background image.

In [15], in the process of binarization, the text image was blurred with a Gaussian filter to reduce the level of noise. On the resulting image, the contours were highlighted by the Canny method. Further, the global threshold for binarization was determined from the local minimum of the histogram of the image intensity values with underlined contours. This local minimum is located between two peaks of the histogram corresponding to the intensity values to the left and to the right of the contour. To calculate the local threshold, a window containing foreground pixels was selected, and the average intensity of these points was determined. The resulting value was used as a threshold for classifying the foreground and background pixels of the image in the sliding window.

In [19], a method was proposed for combining the results of binarization by two local methods, which, however, requires tuning a significant number of parameters. According to the authors of [19], such a combination of methods makes it possible to achieve both high values of the noise immunity indices of binarization and to reduce the error in determining the boundaries of objects.

In [16], the analysis of connected components was used to determine the contours of symbols. Then a window was formed around each outline of the symbol – a rectangle of pixels. At the border of the window, the background pixels were selected, participating in the calculation of the background intensity. The intensity of the foreground was determined from the points of the contours of the symbols. The intensity of the current pixel in the image was compared to the local intensity of the foreground. The current pixel of the image was relative to the foreground or background depending on the result of this comparison. The local background intensity was used to determine if the background is lighter than the foreground or vice versa.

In [11-12], the method [16] was adapted for the binarization of color images, and in [11] the binarization threshold was chosen heuristically and applied to the value of the distance from the current pixel to the background estimate in the color space. The heuristic choice of the threshold significantly reduced the quality of binarization; therefore, in [12], the threshold calculated by the Otsu method was applied to the same distance, which made it possible to improve the quality of binarization.

Recently, along with threshold binarization methods, classification methods using pixel processing have also been used. So, in [9], binarization of color images of text was performed depending on the content of images in the color space *Lab*. If basic information was contained in an image intensity component, the global threshold was applied to that component using the Otsu method. When the main information about the image is contained in the color components a, b, the binarization of color images of the text was performed by the k-means method, and the values of these color components were used as features of the image pixels. The method does not work well with uneven illumination of the original image and does not take into account the spatial relationships between image pixels, which can reduce the quality of binarization.

AIM OF THE RESEARCH

To solve a number of problems such as optical character recognition, text extraction in color images, high quality binarization methods are required for color images of text. The analysis of the known methods of binarization of text images in [9]; [15] and [16] showed that the main factors affecting the quality of binarization are the incorrect choice of the binarization threshold, the appearance of various kinds of artifacts on the image during the acquisition and processing, as well as the incorrect choice of the size of the processing blocks. Images in addition, the quality of binarization of images depends significantly on the content of the images.

In order to improve the quality of binarization of color images of text, in this work, it is proposed to use classification methods instead of a heuristic choice of a threshold, namely, a classifier based on the minimum distance. This will eliminate the influence of the choice of the threshold on the quality of binarization and will ensure high performance of this procedure. In this case, the pixels of the image are divided into two classes, "Foreground" and "Background". To reduce the amount of processed information on the original image, it is advisable to select blocks of pixels for subsequent processing. This was done before applying the minimum distance classifier using connected component analysis

The aim of this research is the development of an improved method for binarization of color images of text using the analysis of connected components and a classifier at a minimum distance, which improves the quality of binarization.

IMPROVEMENT OF THE COLOR BINARIZATION METHOD IMAGE TEXT USING ANALYSIS OF RELATED COMPONENTS AND MINIMUM DISTANCE CLASSIFIER

As a result of the analysis of literature sources for binarization of color images of text, the method [11-12] was chosen as the basic method. This method first selects blocks of the image using the analysis of connected components, and then, taking into account the selected blocks, binarizes the image using a threshold selected heuristically [11] or according to Ots [12]. The analysis of the method [11-12] showed that the time of image processing by this method meets the requirements for the operability of the OCR systems, however, threshold processing reduces the quality of binarization, which is due to the complexity of the choice of the threshold. Improving the quality of binarization of color images of text makes it possible to achieve a sufficient quality of optical character recognition in these images.

The minimum distance classifier is based on the assumption that classes of objects, in this case, image pixels, can be represented by one or more representative objects – prototypes [20]. Then the minimum distance classifier is a special case of a linear classifier, the dividing surface of which is a hyperplane consisting of points equidistant from the centers of two classes.

Consider the case of uniqueness of the prototype. Depending on the content of the images, the feature values of pixels of each of the classes "Foreground" and "Background" may tend to group around some single set of feature values, the prototype of the class, which is representative of this class. A similar situation arises if the variability of pixel features is small for each of the "Foreground" and "Background" classes, and interference is easy to take into account. Then it is expedient to use the classifier by minimum distance to divide the pixels of the image into pixels of the foreground and background. Let the classes "Foreground" and "Background" admits representation using the prototypes of the classes' m_1 and m_2.

In the case of multiple prototypes, each class can be characterized not by a single set of feature values, but by several. Then any pixel belonging to class i (i = 1 for the "Foreground" class and i = 2 for the "Background" class) tends to cluster around one of the prototypes of this class: m_{i1}, m_{i2},..., m_{iMi}, where Mi is the number of class i prototypes.

As an alternative, a method has been developed for binarization of color images of text on the assumption of uniqueness or multiplicity of class prototypes. Stages 1-5 of this method are similar to the stages of the method [11] that perform the analysis of connected components, and to improve the quality of binarization, stages 6, 7 of the basic method are changed and classify according to the minimum distance.

Then the steps of the proposed method under the assumption of uniqueness of class prototypes are as follows.

1. Selection of connected components in the image based on information about the contours. To highlight the outlines on color images of text, the Canny method [21-22] was used, which was applied separately to each of the color components R, G, B. The outlines obtained for each of the color components were combined using the OR operation. After highlighting the contours on the color image of the text, 8-connected components were determined [11].

2. For each connected component, a rectangle described around it was formed; the set of points of the boundaries of this rectangle was denoted by E_B.

3. The rectangles described around connected components that do not correspond to text components were filtered out. For this, the aspect ratios and areas of each such rectangle with thresholds were compared. So, only connected components with the aspect ratio of the circumscribed rectangle in the interval [0.1; 10] pixels, with an area in the interval [15; 0.2 × max $\{N, M\}$] pixels, where $N \times M$ is the size of the original text image.

4. When selecting outlines, both the outer and inner outlines of characters in text images are determined. For example, the letter "D" is represented by two connected components and,

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

accordingly, by two rectangles circumscribed around: one corresponds to the outer contour of the symbol (indicated in Fig. 1, a in green), the other – to the inner contour (indicated in Fig. 1, a in blue). Rectangles, described around text characters, can contain one or two rectangles, described around the inner contours of the characters (Fig. 1a). So, one inner rectangle contains rectangles, described around the characters "A", "D", and two inner rectangles contain a rectangle, circumscribed around the character "B". On the other hand, if the rectangle circumscribed around the outer edge of the symbol contains more than two rectangles.

Thus,

if a $N_{int}> 2$, then delete E_{Bout}, keep E_{Bint},
otherwise delete E_{Bint}, we leave E_{Bout}.

Here E_{Bout} – the set of points of the boundaries of the rectangle described around the outer contour of the symbol under consideration, E_{Bint} – the set of points of the boundaries of the rectangle described around the inner contour of the symbol contained in the rectangle E_{Bout}, N_{int} – the number of rectangles described around the inner contours of the symbol [11].

5. Determination of the background color for each rectangle described around the outer contour of the symbol, taking into account the color of the points in the corners of this rectangle [11]. Let's denote the coordinates of the pixel in the upper left corner of this rectangle (x_1, y_1), in the lower left corner – (x_2, y_2), in the upper right corner of this rectangle – (x_3, y_3), in the lower right corner – (x_4, y_4). Formed set of points (Fig. 1b):

$$B=\{(x_1-1, y_1-1), (x_1, y_1-1), (x_1-1, y_1), (x_2+1, y_2-1), (x_2, y_2-1), (x_2+1, y_2), (x_3-1, y_3+1), (x_3, y_3+1), (x_3-1, y_3), (x_4+1, y_4+1), (x_4, y_4+1), (x_4+1, y_4)\}. \quad (2)$$

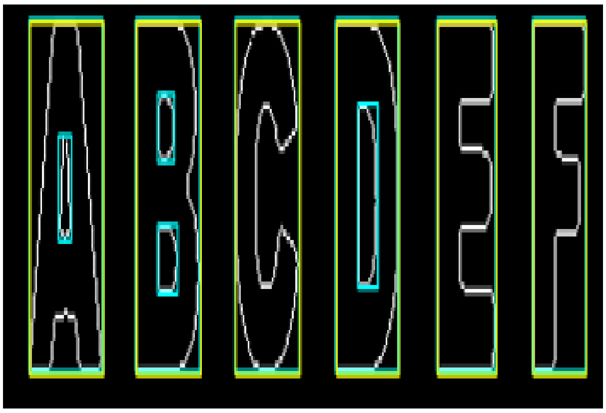
Further, in the La^*b^* color space, the medians of the pixel colors (2) were determined for each of the components L, a^*, b^*. We get the values of the background color components – the vector $\mathbf{B} = (B_L, B_a, B_b)$.

6. Determination of the foreground color for each rectangle described around the outer outline of the symbol. For each such rectangle, the foreground color was estimated as a vector of average values of the components L, a^*, b^*, calculated from the points of the symbol outline: $\mathbf{F} = (F_L, F_a, F_b)$, where

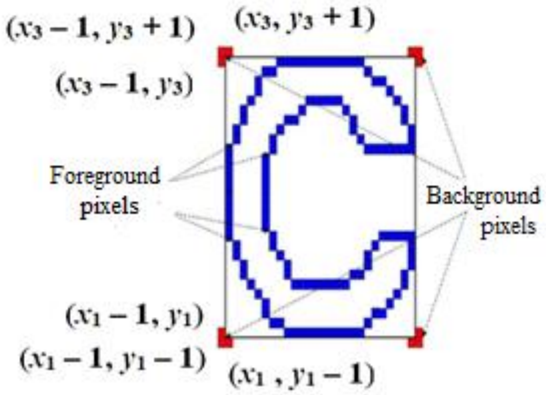
$$F_L = \frac{1}{N_E}\sum_{(x,y)\in E} L(x,y),$$

$$F_a = \frac{1}{N_E}\sum_{(x,y)\in E} a^*(x,y), \quad (3)$$

$$F_b = \frac{1}{N_E}\sum_{(x,y)\in E} b^*(x,y)$$



a



b

Fig. 1. **Rectangles, described around text characters:**
a – one or two rectangles, described around the inner contours of the characters [11];
b – calculation of the background color taking into account the colors of the points in the corners of the rectangle [11]
Source: **[11]**

In the formula (3) the following are indicated: E – the set of points of the outline of the symbol belonging to the rectangle described around it; $L(x, y)$, $a^*(x, y)$, $b^*(x, y)$ are the values of the color components L, a^*, b^*, respectively, for a pixel with coordinates (x, y), N_E is the number of elements of the set E.

7. Making a decision on assigning pixels of a color image of text to one of two classes: "Foreground" or "Background". For each pixel P, located inside the rectangle described around the outer contour of the symbol, the lesser of two distances was determined: to the background and to the foreground in color space. At the minimum of these distances, the current pixel was relative to the background or foreground.

If the pixel P corresponds to the values of the color components (P_L, P_a, P_b), then the difference between the colors of the background and this pixel was calculated by the formula

$$dE = \sqrt{(P_L - B_L)^2 + (P_a - B_a)^2 + (P_b - B_b)^2} \ . \quad (4)$$

The distance in color space from P to the vector F of foreground color values was calculated by the formula

$$dF = \sqrt{(P_L - F_L)^2 + (P_a - F_a)^2 + (P_b - F_b)^2} \ . \quad (5)$$

Taking into account (4), (5), the condition $dE \geq dF$. If it was executed, then the P pixel belonged to the foreground (text characters) and was painted black. If the condition $dE \geq dF$ was not executed, then the pixel P, located inside the rectangle described around the outer contour of the symbol, belonged to the background and was painted over with white color. Likewise, pixels that did not fall into any of the rectangles described around the outer outline of the symbol were also referred to the background and were painted white.

Assuming a multiplicity of prototypes of classes, stages 6, 7 of the proposed method for binarization of color images of text are transformed to the following form.

6. The foreground color was determined adaptively for each pixel P of the rectangle described around the outer outline of the symbol. For each such pixel P with Cartesian coordinates (x, y), the distances to the points of the contour around which the rectangle is described were calculated. Let's designate E – a contour pixel with Cartesian coordinates (x_E, y_E).

The distance between pixels P and E was estimated by the formula:

$$d(P, E) = \sqrt{(x - x_E)^2 + (y - y_E)^2} \ . \quad (6)$$

The closest in the sense of distance (6) pixel of the contour from the rectangle described around the outer contour of the symbol to each pixel P from the same rectangle was determined. Let's designate EP this pixel of the contour closest to the pixel P. Pixel color value $E_P = (E_{PL}, E_{Pa}, E_{Pb})$. This color was taken as the foreground color for pixel P.

7. The decision to assign the pixels of the color image of the text to one of two classes, “Foreground” or “Background”, was carried out similarly to the case of the uniqueness of the prototypes of these classes. The difference was in estimating the distance from the current pixel P to the foreground, since the assessment of the foreground color for each pixel P was carried out adaptively to the color of the pixel E_P determined in the previous step.

Then the distance from P to E_P was calculated in color coordinates by the formula:

$$dE_P = \sqrt{(P_L - E_{PL})^2 + (P_a - E_{Pa})^2 + (P_b - E_{Pb})^2} \ . \quad (7)$$

Taking into account (7), for each pixel P belonging to the rectangle described around the outer contour of the symbol, the condition $dE_P \geq dF$ instead of the condition $dE \geq dF$.

RESEARCH OF THE PROPOSED BINARIZATION METHOD COLOR IMAGE TEXT

As a result of experimental studies, the quality of binarization of color images of text was assessed using the proposed method and methods known from the literature, for example, [10-11] for the same test images of printed and handwritten text from the DIBCO 2009 database [23]. These color images of text with an average size of 400 × 1300 pixels were scanned at 300 dpi. For each test image, using the information available in the database, the result of marking the points by an expert was formed. In Fig. 2 shows the original color images of the text and the result of the expert marking for each such image. In Fig. 3 and Fig. 4 shows the result of binarization by the proposed method and the result of binarization by the method [11].

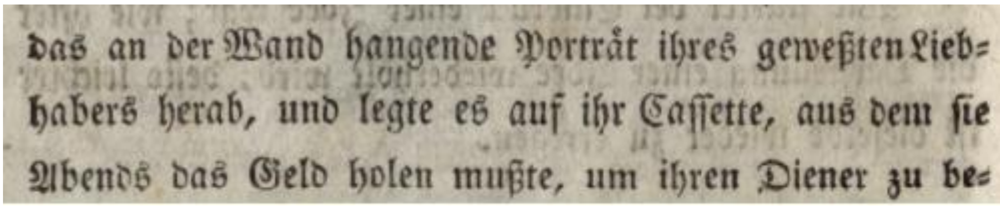
The quality of binarization of color images of text by the proposed method and methods of works [12-13] and others, compared with the markup of the image by an expert, was assessed by calculating the following indicators: F-measure, peak signal-to-noise ratio, metric of negative indicators, metric of penalty for incorrect classification [23].

The peak signal-to-noise ratio $PSNR$ was calculated according to the expression:

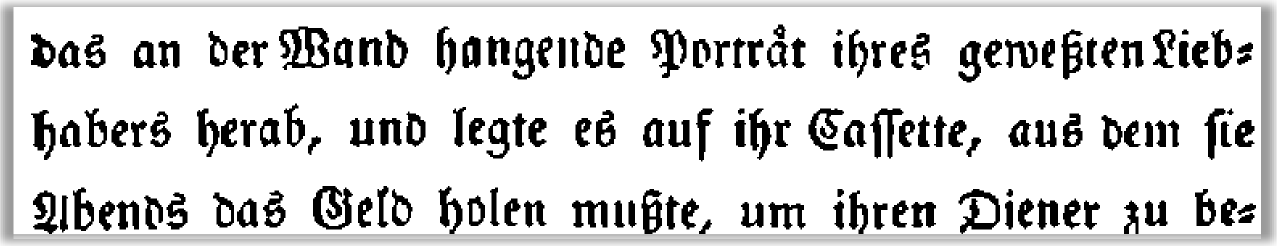
$$PSNR = 10 \lg\left(\frac{C^2}{MSE}\right), \quad (8)$$

where C is the difference between the estimates of the foreground and background intensities, MSE is the mean square error due to the difference in the image $I(x, y)$ binarized by the studied method, from image $I_0(x, y)$ marked by an expert. $PSNR$ characterizes the similarity of two images. The higher the $PSNR$, the more similar the two images are.

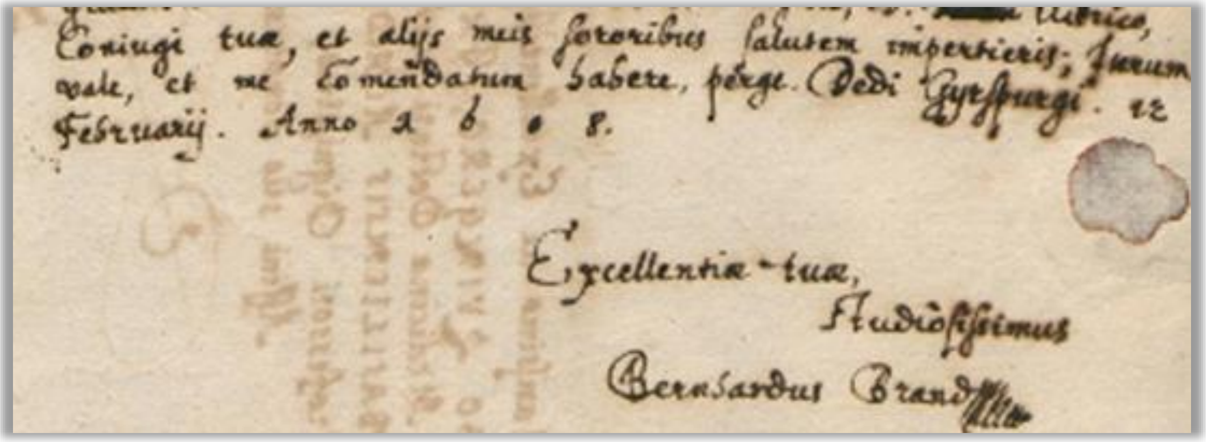
ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)


das an der Wand hangende Porträt ihres geweßten Lieb=
habers herab, und legte es auf ihr Cassette, aus dem sie
Abends das Geld holen mußte, um ihren Diener zu be=

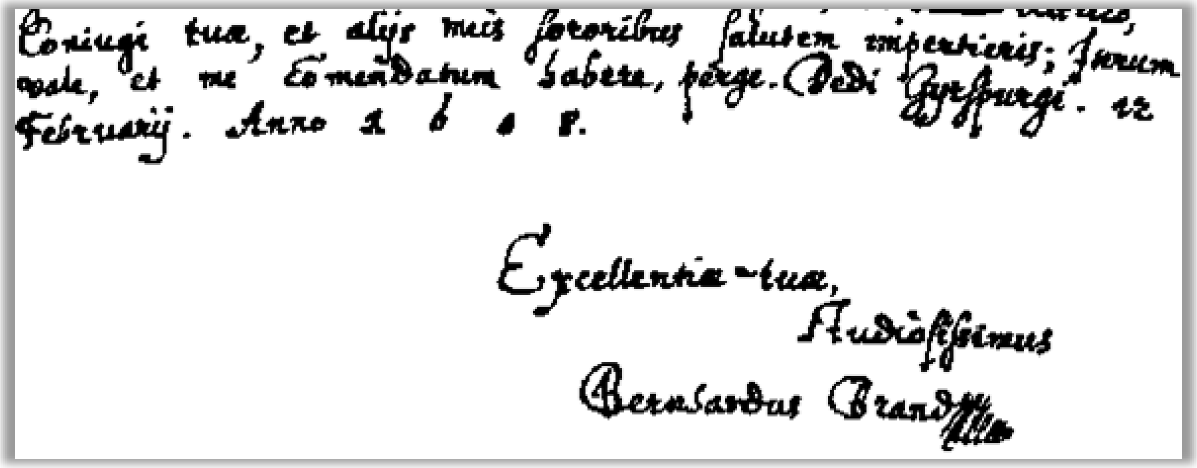
a


das an der Wand hangende Porträt ihres geweßten Lieb=
habers herab, und legte es auf ihr Cassette, aus dem sie
Abends das Geld holen mußte, um ihren Diener zu be=

b


Coniugi tuæ, et alijs meis sororibus salutem impertieris; Iterum
vale, et me Comendatum habere, perge. Dedi Gryphurgi. 12
Februarij. Anno 1 6 . 8.
Excellentiæ tuæ,
Studiosissimus
Bernardus Brand

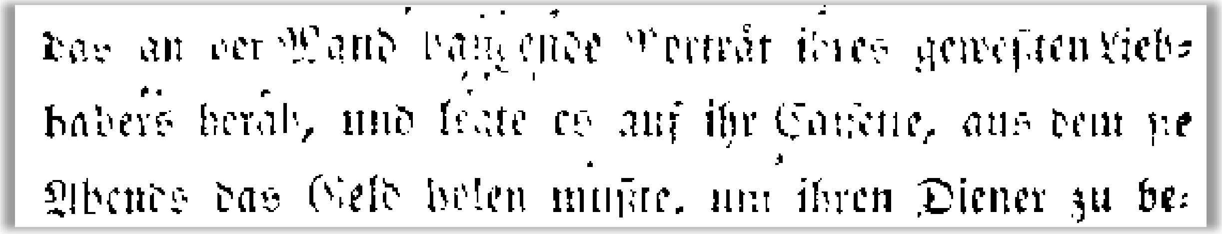
c


Coniugi tuæ, et alijs meis sororibus salutem impertieris; Iterum
vale, et me Comendatum habere, perge. Dedi Gryphurgi. 12
Februarij. Anno 1 6 . 8.
Excellentiæ tuæ,
Studiosissimus
Bernardus Brand

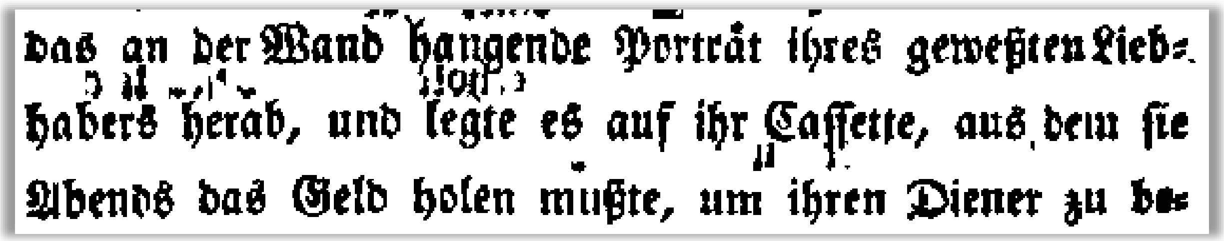
d

Fig. 2. **Images from the database of scanned printed documents DIBCO 2009 [23]: a – original image of printed text; b – result of expert marking of printed text; c – original image of handwritten text; d – result of expert marking of handwritten text**

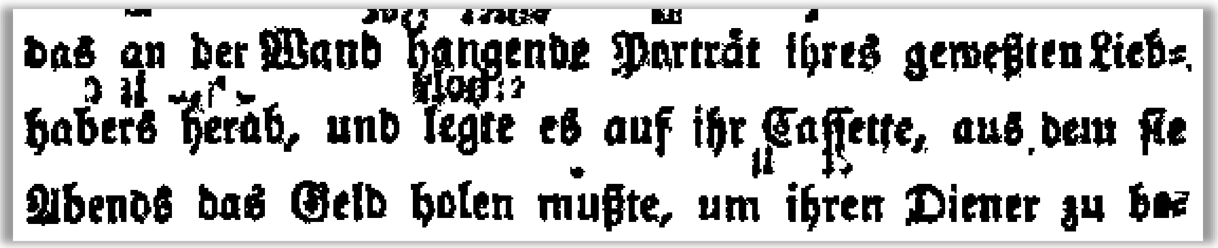
Source: **[23]**



a



b



c

Fig. 3. **Results of binarization of color images of printed text:**
a – the basic method; b –the developed method with the uniqueness of prototypes;
c – developed method with multiple prototypes
***Source:* compiled by the author**

The negative rate metric (*NRM*) is based on a pixel-by-pixel comparison of images $I(x, y)$ and $I_0(x, y)$:

$$NRM = \frac{1}{2}\left(\frac{FN}{FN+TP} + \frac{FP}{FP+TN}\right), \qquad (9)$$

where: *TP* is the percentage of pixels of images from the class labeled “Foreground” correctly assigned to the class labeled “Foreground”; *FP* is the percentage of pixels of images from the class labeled "Foreground" that are incorrectly assigned to the class labeled “Background”; *FN* is the percentage of pixels of images from the class labeled “Background” incorrectly assigned to the class labeled “Foreground”; *TN* is the percentage of pixels of images from the class labeled “Background” correctly assigned to the class labeled “Background”.

F-measure was determined by the formula:

$$F = \frac{2PrRc}{Pr+Rc}, \; Pr = \frac{TP}{TP+FP}, \; Rc = \frac{TP}{FN+TP}, \qquad (10)$$

where: *Pr* is the precision; *Rc* – recall.

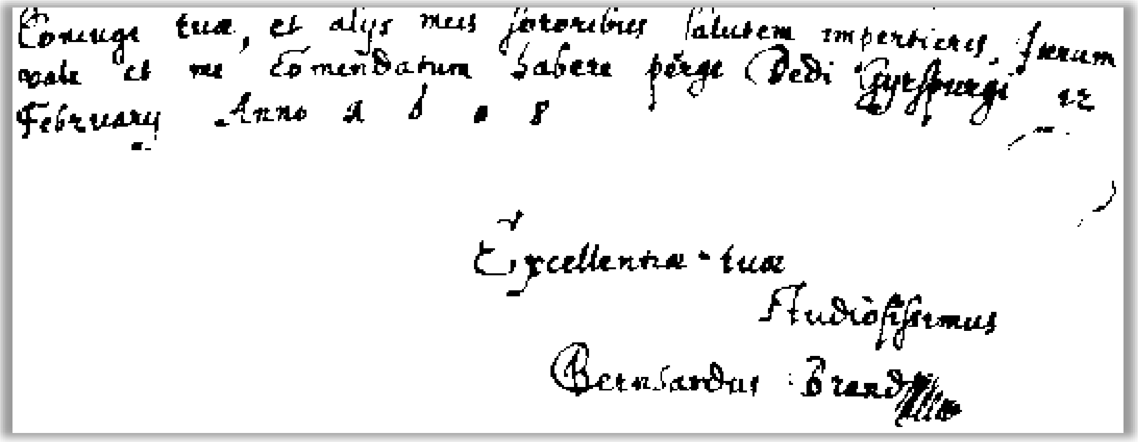
The misclassification penalty metric (*MPM*) characterizes the error in determining the boundaries of objects as a result of binarization and is calculated as

$$MPM = \frac{1}{2D}\left(\sum_{i=1}^{N_{FN}} d_{FN}^{i} + \sum_{i=1}^{N_{FP}} d_{FP}^{i}\right), \qquad (11)$$

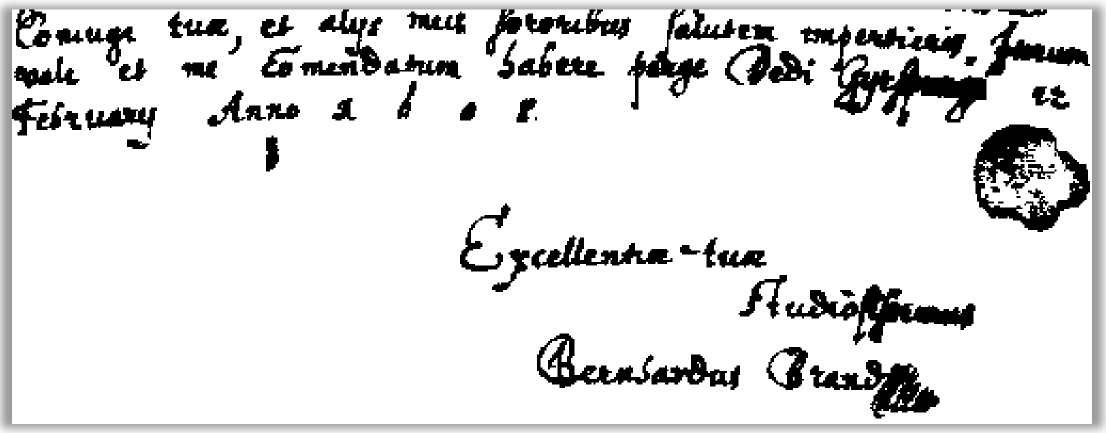
where: d_{FN}^{i} is the distance from the *i*-th pixel of the background, incorrectly classified as a pixel of the foreground, to the outline of the symbol in the image marked by the expert; d_{FP}^{i} is the distance from the *j*-th foreground pixel, incorrectly classified as a background pixel, to the character outline in the image marked by an expert. Also N_{FN} is the number of background pixels misclassified as foreground pixels; N_{FP} – the number of foreground pixels misclassified as background pixels; *D* is the sum of the distances from the foreground points to the outlines of the characters in the image marked by the expert.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

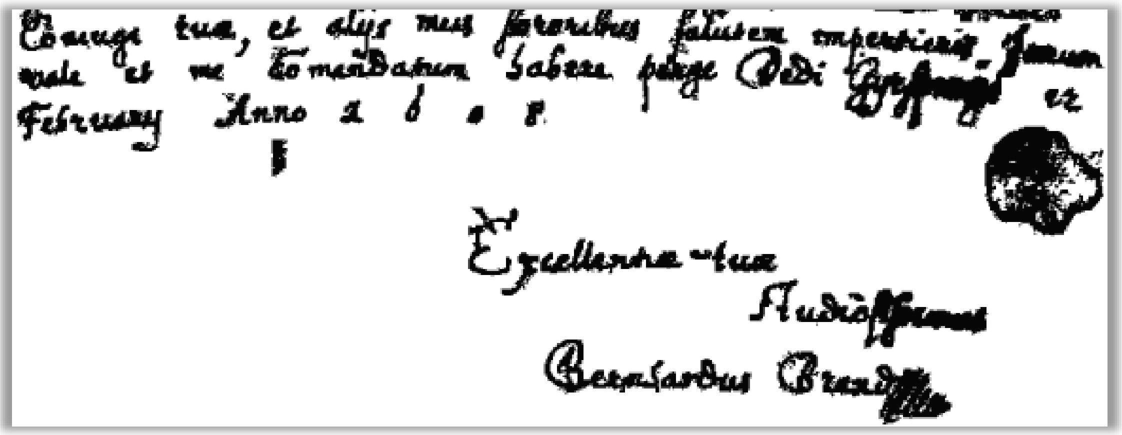
Note that test color images of printed and handwritten text differ significantly in their properties: the thickness of letters, the distance between them and the direction of the lines. Therefore, in this work, the characteristics of the quality of binarization were estimated for images of each of these two classes separately, in contrast to works [13-14] and others, in which the characteristics of the quality of binarization were averaged over images of both classes.



a



b



c

Fig. 4. **Results of binarization of color images of handwritten text:**
a – the basic method; b –the developed method with the uniqueness of prototypes;
c –developed method with multiple prototypes
***Source:* compiled by the author**

During the experiment, we first investigated the dependence of the characteristics of the proposed binarization method (*F*-measure according to formula (10), *PSNR* according to formula (8), *NRM* according to formula (9), *MPM* according to formula (11)) on the lower threshold of the Canny method, the value of which was chosen in the range 0.3…0.9. The value of the upper threshold of the canny method was fixed as 0.95.

For color images of printed text, the proposed method achieved a high quality of binarization in terms of the *F*-measure at threshold values of 0.3…0.7; according to the *PSNR* indicator – at threshold values of 0.5…0.6; according to the *NRM* indicator – at threshold values of 0.3…0.7; according to the *MPM* indicator – at threshold values of 0.6…0.9. Based on the study carried out for color images of printed text, the threshold of the Canny method was fixed as 0.6. In the case of images of handwritten text, the proposed method achieved a high quality of binarization in terms of the *F*-measure at threshold values of 0.3…0.4; according to the *PSNR* indicator – at threshold values of 0.3…0.5; according to the *NRM* indicator – at threshold values of 0.3…0.4; according to the *MPM* indicator – at a threshold value of 0.6…0.5. Therefore, for images of handwritten text, the threshold of the Canny method was fixed as 0.3.

The results are shown in Table 1, Table 2, Table 3 show that on the test images by the values of the *F*-measure proposed the method exceeds the known methods by 4-11 % for images of printed text and by 3-12 % for images of handwritten text. The *F*–measure represents the integral characteristic of the quality of the binarization method, calculated on the basis of *TP*, *FP*, *FN* therefore the values of these indicators are given in Table 2. Analyzing the values of indicators from Table 2, note that, among the classification errors, more often pixels of images from the class labeled “Foreground” were incorrectly assigned to the class labeled “Background”. The use in the proposed method for assessing the foreground color instead of the mean value of the median did not lead to a significant improvement in the characteristics of the method. The proposed method also used post-processing using merged group filtering, which improved the visual quality of the resulting binary images.

Table 1. **Results of evaluating the characteristics of binarization methods**

Link, year of publication	*F*-measure, %	*PSNR*	*NRM*	*MPM*
[13], 2000	85.41	16.39	6.94	3.20
[14], 2006	85.25	16.50	10,00	0.70
[15], 2014	83.66	15.60	4.81	3.29
[16], 2007	85.39	16.44	9.02	1.84
[18], 2017	90.58	18.13	5.50	2.26
[11], 2017 (printed text)	55.20	11.74	16.38	3.41
[12], 2019 (printed text)	88.98	15.17	8.44	1.72
The proposed method with uniqueness of prototypes (printed text)	94.38	13.38	5.23	3.91
The proposed method with multiple prototypes (printed text)	93.90	14.66	5.17	3.22
[11], 2017 (handwritten text)	56.57	13.82	18.15	0.12
[12], 2019 (handwritten text)	84.13	14.82	11.23	1.19
The proposed method with uniqueness of prototypes (handwritten text)	93.43	14.50	5.64	1.47
The proposed method with multiple prototypes (handwritten text)	92.90	12.31	6.32	1.93

Source: **compiled by the author**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Table 2. **Results of *TP, FP, FN, TN* evaluation for methods of binarization of images of printed text**

Link, year of publication	*TP*, %	*FP*, %	*FN*, %	*TN*, %
[11], 2017	45.38	54.62	0.10	99.90
[12], 2019	80.57	19.42	0.58	99.42
The proposed method with uniqueness of prototypes	93.55	6.45	4.53	95.46
The proposed method with multiple prototypes	91.34	8.66	2.72	97.28

Source: **compiled by the author**

Table 3. **Results of *TP, FP, FN, TN* estimation for methods of binarization of images of handwritten text**

Link, year of publication	*TP*, %	*FP*, %	*FN*, %	*TN*, %
[11], 2017	42.18	57.82	0.23	99.77
[12], 2019	75.14	24.86	2.19	97.81
The proposed method with uniqueness of prototypes	91.78	8.22	3.98	96.02
The proposed method with multiple prototypes	93.54	6.46	7.22	92.78

Source: **compiled by the author**

According to *PSNR* values – characteristics of the similarity between the obtained binary image and the binary image marked by an expert, the proposed method is worse known binarization methods [13-14], [15-16], [18] by 3-19 % for images of printed text and by 3-25 % for images handwritten text. According to the *NRM* values, the binarization error metric, the proposed method is the best known binarization methods up to two times. The *NRM* indicator is dimensionless, the lower its value, the fewer errors are obtained during binarization and the more the resulting image is similar to the expert markup. The *MRM* indicator, which characterizes the error in determining the boundaries of objects as a result of binarization, takes smaller values for images of better quality. According to the values of this dimensionless index, the proposed method for images of printed text is comparable to [13]; [15] and to [11], worse than [12]; [16], and [18] up to 2 times, and worse than method [14] up to 4 times, for images of handwritten text – comparable to [12]; [16] and exceeds [11]; [13], and also [15]; [18] up to 2 times.

The speed of binarization by the proposed method in comparison with the methods [11-12] was estimated by the processing time. For images of the printed text of the test database, the processing time by the methods [11-12] and the proposed method with the uniqueness of class prototypes was 1-2 s. For images of handwritten text, the processing time by the same methods was 0.5-4.5 s. The research was carried out using an Intel Core i5-7400 processor, 3 GHz CPU, 16GB memory, operating system Windows 10, 64 bit. The processing time by the proposed method with a multiplicity of class prototypes was 1.5-3.5 s for images of printed text and 0.8-18 s for images of handwritten text. The large scatter of processing time values is due to the properties of handwritten text images, namely, the number of connected components on them.

CONCLUSIONS

The method of binarization of color images of text has been improved by making a decision on assigning image pixels to the foreground or background using classification. This method differs from those known from literary sources in that after the selection of blocks of text using the edge detector and the analysis of connected components, the pixels of the image were related to the foreground or background using the classifier at the minimum distance. The proposed method made it possible to improve the quality of binarization of color images of printed and handwritten text with a comparable processing time (Table 2, Table 3). Among binarization errors, more often pixels of images from the class labeled “Foreground” were incorrectly assigned to the class labeled “Background”.

The proposed method of binarization with the uniqueness of class prototypes is recommended to be used in problems of processing color images of printed text, for which the error in determining the boundaries of characters as a result of binarization is compensated by the thickness of letters. With a plurality of class prototypes, the proposed binarization method is recommended to be used in problems of processing color images of handwritten text, if high performance is not required.

Improved this method of binarization has shown its efficiency in cases of slow changes in the light and illumination of the text and background, however, abrupt changes in color and illumination, as well as a textured background, do not allow providing the binarization quality required for solving practical problems. The latter makes it possible to use the developed method of binarization in systems without a special lighting-focusing installation.

Directions for further research can be taking into account the influence of various artifacts of text images, for example, uneven illumination, adding a preprocessing stage [24], processing various damaged documents [25- 26], in particular, archival documents [27-28]. The developed method of binarization can be used to separate text from illustrations in the task of mass processing of color scanned documents in order to create archives of enterprises and fill libraries [29-30].

REFERENCES

1. Ishchenko, O. V. “Rozrobka modulia intelektualnoi systemy obrobky vidskanovanykh dokumentiv na bazi kombinovanoho metodu sehmentatsii zobrazhen. Suchasnyi stan naukovykh doslidzhen ta tekhnolohii v promyslovosti”. *The current state of research and technology in industry* (in Ukrainian). 2019; No. 2(8):44–53. DOI: 10.30837/2522-9818.2019.8.044.

2. Ishchenko, A., Polyakova, M. & Nesteryuk, A. “The technique of extraction text areas on scanned document image using linear filtration”. *Applied Aspects of Information Technology*. *Publ. Science i Technical*. Odesa: Ukraine. 2019; No.2(3): 206-215. DOI: 10.15276/aait.03.2019.3.

3. Polyakova, M., Ishchenko, A., Volkova, N. & Pavlov, O. “The combining segmentation method of the scanned documents images with sequential division of the photo, graphics, and the text areas”. *Eastern-European Journal of Enterprise Technologies*. 2018; No. 5/2 (95): 6–16. DOI: 10.15587/1729-4061.2018.142735.

4. Ishchenko, A., Polyakova, M., Kuvaieva, V. & Nesteryuk, A. “Elaboration of structural representation of regions of scanned document images for MRC model”. *Eastern-European Journal of Enterprise Technologies*. 2018; No. 6/2 (96): 32–38. DOI: 10.15587/1729-4061.2018.147671.

5. Polyakova, M. V., Dishlyuk, V. O. & Makovetskiy, O. S. “Mobilnyi dodatok na osnovi tekhnolohii dopovnenoi realnosti dlia navchannia anhliiskii movi ditei doshkilnoho viku”. *Practical and theoretical nutritional development of science and education (part III): materials of the 25th International Conference on Practical Science* (in Ukrainian). 2020. Lviv: Lviv Science Forum. 2020. p.54–56.

6. Molchanova, V.S. “Adaptivnyj porogovyj metod binarizacii rastrovykh izobrazhenij tekhnicheskikh chertezhej”. *Radioelectronika. Informatics. Management* (in Russian). 2015; No. 2: 62–70.

7. Bolotova, Yu. A., Spitsyn, V. G. & Osina, P. M. “Obzor algoritmov detektirovaniya tekstovykh oblastej na izobrazheniyakh i videozapisyakh”. *Computer optics* (in Russian). 2017; Vol. 41, Issue 3: 441–452.

8. Almeida, M., Lins, R. D., Bernardino, R., Jesus, D. & Lima, B. “A new binarization algorithm for historical documents”. *Journal of Imaging*. 2018; Vol. 4, No. 27: 110–122. DOI: 10.3390 / jimaging4020027.

9. Ostapov, D. S. “Binarizaciya “fon-ob’ekt” predobrabotkoj izobrazheniya i nechyotkim algoritmom k-means”. *Bulletin of the Astrakhan State Technical University. Series: Management, computer technology and information* (in Russian). 2016; No. 3: 32–39.

10. Gonzalez, R. & Woods, R. “Digital Image processing (4th edition)”. London: Pearson Education. 2018. 1020 p. DOI / PMID / ISBN: 9780133356724.

11. Das, A. & Chowdhury, S. “Adaptive method for multi colored text binarization”. *Proc. of theInt. Conf. on Systems, Signals and Image Processing (IWSSIP)* (Poznan, Poland, 22-24 May, 2017). Poznan: Poland. 2017: p.359–366.

12. Chaban, S. V. & Polyakova, M. V. “Adaptyvna binaryzatsiia kolorovykh zobrazhen tekstu za dopomohoiu metodu Otsu”. *Program, Project, Portfolio p3management: materials of the IV International Scientific and Practical Conference* (Odessa, 6 – 7 December, 2019) (in Ukrainian). Odessa, 2019. p. 22–28.

13. Savuola, J. & Pietikainen, M. “Adaptive document image binarization”. *Pattern Recognition*. 2000; Vol. 33, No. 2: 225–236.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

14. Gatos, B., Pratikakis, I. & Perantonis, S. "Adaptive degraded document image binarization". *Pattern Recognition*. 2006; Vol. 39, No. 3: 317–327.

15. Biswas, B., Bhattacharya, U. & Chaudhuri, B. B. "A global-to-local approach to binarization of degraded document images". *Proc. of the 22nd Int. Conf. on Pattern Recognition* (Stockholm, Sweden, 24–28 August, 2014). Washington: 2014. p. 3008–3013.

16. Kasar, T., Kumar, J. & Ramakrishnan, A. G. "Font and background color independent text binarization". *Proc. of the 2nd Int. Workshop on Camera-Based Document Analysis and Recognition* (Curitiba, Brazil, September 22, 2007). Washington: 2007. p. 3–9.

17. Israfilov, Kh. S. "Investigation of methods for binarization of images". *Bulletin of Science and Education* (in Russian). 2017; Vol. 2, No. 6 (30): 14–19.

18. Vats, E., Hast, A. & Singh, P. "Automatic Document Image Binarization using Bayesian Optimization". *4th International Workshop on Historical Document Imaging and Processing (HIP2017)*. ACM. New York: NY, USA. 2017: p.89–94. DOI: 10.1145 / 3151509.3151520.

19. Su, B., Lu, S. & Tan, C. L. "Combination of Document Image Binarization Techniques". *2011 International Conference on Document Analysis and Recognition*. Beijing: China. 2011: p.22–26. DOI: 10.1109 / ICDAR.2011.14.

20. Tu, J. & Gonzalez, R. "Principy raspoznavaniya obrazov". *Publ. Mir* (in Russian). Moscow: Russian Federation. 1978. 412 p.

21. Canny, J. "A computational approach to edge detection". *IEEE Trans. on PAMI*. 1986; Vol.8: 679–698.

22. Akinlar, C. & Topal, C. "Color ED: Color edge and segment detection by Edge Drawing (ED)". *Journal of Visual Communication and Image Representation*. 2017; Vol. 44: 82–94.

23. Gatos, B., Ntirogiannis, K. & Pratikakis, I. "DIBCO 2009: document image binarization contest". *Int. Journal on Document Analysis and Recognition*. 2011; Vol. 14, No. 1: 35–44.

24. Michalak, H. & Okarma, K. "Improvement of Image Binarization Methods Using Image Preprocessing with Local Entropy Filtering for Alphanumerical Character Recognition Purposes". *Entropy*. 2019; 21 (6): 562 p. DOI: https://doi.org/10.3390/e21060562.

25. Michalak, H. & Okarma, K. "Optimization of Degraded Document Image Binarization Method Based on Background Estimation". *28th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision WSCG 2020*. Pilsen: Czech Republic. *Journal of WSCG28* (1-2). p.89-98. DOI: 10.24132 / CSRN.2020.3001.11.

26. Sehad, A. "Ancient degraded document image binarization based on texture features". *8th International Symposium on Image and Signal Processing and Analysis, ISPA 2013*, Trieste: Italy. September 4-6, 2013. p.189-193. DOI:10.1109 / ISPA.2013.6703737.

27. Sehad, A., Chibani, Y., Hedjam, R. & Cheriet, M. "Gabor filter-based texture for ancient degraded document image binarization". *Pattern Analysis and Applications*. 2019; Vol. 22: 1–22. DOI: https://doi.org/10.1007/s10044-018-0747-7.

28. Sulaiman, A., Omar, K. & Nasrudin, M. F. "Degraded Historical Document Binarization: A Review on Issues, Challenges, Techniques, and Future Directions". *Journal of Imaging*. 2019; No. 5 (4): 48–73 DOI:10.3390 / jimaging5040048.

29. Bukhari, S. S., Shafait, F. & Breuel, T. "Improved document image segmentation algorithm using multiresolution morphology". *In Proc. of the 18th Document Recognition and Retrieval Conf., Document Recognition and Retrieval XVIII - DRR 2011*. San Jose: CA, USA. 2011. DOI: 10.1117 / 12.873461.

30. Polyakova, M., Ishchenko, A. & Huliaieva, N. "Document image segmentation using averaging filtering and mathematical morphology". *14th International Conference on Advanced Trends in Radioelecrtronics. Telecommunications and Computer Engineering (TCSET)*. Lviv-Slavske: Ukraine. 2018. p.966-969. DOI: https://doi.org/10.1109/TCSET.2018.8336354.

Conflicts of Interest: the authors declare no conflict of interest.

Received 08.02.2021
Received after revision 11.03.2021
Accepted 16.03.2021

DOI: 10.15276/aait.01.2021.5
УДК 004.93

ВДОКОНАЛЕННЯ МЕТОДУ БІНАРИЗАЦІЇ КОЛЬОРОВИХ ЗОБРАЖЕНЬ ТЕКСТУ З ВИКОРИСТАННЯМ КЛАСИФІКАТОРА ЗА МІНІМУМОМ ВІДСТАНІ

Марина В'ячеславівна Полякова[1)]
ORCID: org/0000-0001-7229-7657; marina_polyakova@rambler.ru
Олександр Геннадійович Нестерюк[1)]
ORCID: org/0000-0002-0806-8259; nesteryuk@opu.ua
[1)] Одеський національний політехнічний університет, пр. Шевченка, 1, Одеса, 65044, Україна

АНОТАЦІЯ

Системи оптичного розпізнавання символів на зображеннях використовуються для конвертації книг і документів в електронний вигляд, для автоматизації систем обліку в бізнесі, при розпізнаванні маркерів технологіями доповненої реальності та ін. Якість оптичного розпізнавання символів за умови застосування бінаризації в значній мірі визначається якістю відділення пікселів переднього плану від фону. Проаналізовано методи бінаризації зображень тексту та відзначено недостатню якість бінаризації. Щоб підвищити якість бинаризации кольорових зображень тексту, доцільно для поділу пікселів зображення на два класи, «Передній план» і «Фон», використовувати замість евристичного вибору порога класифікаційні методи. Як напрямок досліджень вибрано вдосконалення існуючого методу бінаризації кольорових зображень тексту шляхом використання класифікатора за мінімумом відстані. Для скорочення обсягу оброблюваної інформації до застосування класифікатора доцільно виділити блоки пікселів для подальшої обробки. Це виконувалося за допомогою аналізу зв'язкових компонент на оригінальному документі. Розроблено вдосконалений метод бінаризації кольорових зображень тексту із застосуванням аналізу зв'язкових компонент та класифікатора за мінімумом відстані. Дослідження розробленого методу показало, що він краще існуючих методів бінаризації за показниками завадостійкості бінаризації, але гірший за показником, що характеризує похибку визначення границь об'єктів. Серед помилок розпізнавання частіше пікселі зображень з класу з міткою «Передній план» неправильно ставилися до класу з міткою «Фон». Запропонований метод бінаризації при єдиності прототипів класів рекомендується використовувати в задачах обробки кольорових зображень друкованого тексту, для якого похибка визначення меж символів в результаті бинаризации компенсується товщиною букв. При множинності прототипів класів запропонований метод бінаризації рекомендується використовувати в задачах обробки кольорових зображень рукописного тексту, якщо не потрібна висока швидкодія. Вдосконалений метод бінаризації показав свою працездатність у випадках повільного зміни кольору і освітленості тексту і фону, проте стрибкоподібні зміни кольору і освітленості, а також текстурований фон не дозволяють забезпечити необхідну для вирішення практичних завдань якість бінаризації.

Ключові слова: бінаризація зображень; класифікатор за мінімумом відстані; оптичне розпізнавання символів; кольорове зображення тексту; фон зображення; передній план зображення

ABOUT THE AUTHORS

Marina V. Polyakova – Dr. Sci. (Eng) (2014), Professor of Department of Applied Mathematics and Information Technologies Department, Odessa National Polytechnic University, 1, Shevchenka ave., Odesa, 65044, Ukraine
ORCID: org/0000-0001-7229-7657; marina_polyakova@rambler.ru
***Research field*:** Digital Image Processing; Pattern Recognition
Марина В'ячеславівна Полякова – доктор технічних наук, професор кафедри Прикладної математики і інформаційних технологій, Одеський національний політехнічний університет, пр. Шевченка, 1, Одеса, 65044, Україна

Alexandr G. Nesteryuk – PhD (Eng) (2017), Associate Professor of the Department of Computer Systems, Odessa National Polytechnic University, 1, Shevchenka ave., Odesa, 65044, Ukraine
ORCID: org/0000-0002-0806-8259; nesteryuk@opu.ua
***Research field*:** Hybrid systems, Digital Image Processing; Pattern Recognition

Олександр Геннадійович Нестерюк – кандидат технічних наук, доцент кафедри Комп'ютерних систем, Одеський національний політехнічний університет, пр. Шевченка, 1, Одеса, 65044, Україна

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)