ELABORATION OF STRUCTURAL REPRESENTATION OF REGIONS OF SCANNED DOCUMENT IMAGES FOR MRC MODEL

A. Ishchenko
Senior Lecturer*
E-mail: alesya.ishchenko@gmail.com
M. Polyakova
Doctor of Technical Sciences,
Associate Professor*
E-mail: marina_polyakova@rambler.ru
V. Kuvaieva
Assistant
Department of Information Systems**
E-mail: vkuvayeva@opu.ua
A. Nesteryuk
PhD, Associate Professor
Department of Computer Systems**
E-mail: nesteryuk@opu.ua
*Department of Applied Mathematics and
Information Technologies**
**Odessa National Polytechnic University
Shevchenka ave., 1, Odessa, Ukraine, 65044

Модель змішаного растрового контенту (Mixed Raster Content) є поширеною формою представлення зображення відсканованого документа. Виконано подальший розвиток моделі змішаного растрового контенту, який полягає в структурно-статистичному представленні однорідних областей на кожному шарі зображення MRC моделі. Метою такого представлення зображення є подальше виділення областей зображення, які цікавлять, та ідентифікація для вирішення задачі сегментації зображень відсканованих документів.

Шар, що містить графічні та фотозображення, представлявся у вигляді об'єднання кількох областей за допомогою кусочно-постійної функції інтенсивності області зображення. Для цього графічне та фотозображення представлялися у вигляді розбиття на сегменти, що містять пікселі однорідної інтенсивності. Для визначення цих областей з метою відділення графіки від фотозображень розглядалися значення перепаду інтенсивності пікселів на границях множин, що містять дані області.

Шар, що містить текст, представлявся як зображення областей структурної текстури на однорідному фоні. Ці області містили фрагменти звичайного тексту та заголовка, що мають однакову інтенсивність пікселів і розрізняються формою та розміром символів, а також відстанню між ними. Таке уявлення шару дозволило врахувати просторові зв'язки між пікселями в зображенні символу і надалі виділити область тексту з фону.

Запропонована модель представлення зображення відсканованого документа дозволяє виділити шари зображення, що містять однорідні області, та звести процес сегментації всього зображення до сегментації окремих шарів зображення. Це дозволяє підвищити швидкодію обробки при збереженні високої якості сегментації зображення.

Ключові слова: зображення відсканованих документів, модель змішаного растрового контенту, виділення тексту

1. Introduction

With the increase in the capacity of personal computers and the need to store a large number of scanned documents, the question arises about the representation of images of scanned documents. The basis of any scanned document is the relative position of its elements, i.e. its structure. Images, text, lines and frames are the structure-forming elements of the document. Knowledge of the structure of the scanned document allows extracting information from structural blocks. Structural blocks are regions that contain uniform content, such as text only or image only.

The same document image can be represented using different models depending on the application.

To reduce the space for storing electronic documents and organize a quick search of information in them, the problem of image segmentation of a scanned document is solved. Regions of text, graphic and photographic images are extracted when segmenting the scanned document images. However, each of these structural blocks has different properties, hence the difficulty of choosing a system of features for the extraction of text regions, as well as regions of a graphic and/or photographic image from the background.

Therefore, there is a need to represent the image of the scanned document to solve the problem of segmentation of such an image in order to extract the image regions of interest such as text, photo and graphic from the background. The segmentation is presented with the requirement of the high performance speed of segmentation while ensuring the required quality of processing of scanned document images.

2. Literature review and problem statement

One of the most common representations of scanned document images is the Mixed Raster Content (MRC) model used in [1]. This model combines a hierarchical representa-

tion of a scanned document image with the determination of its structural blocks.

Using the MRC model allows reducing the size of the file with the scanned document image which contains text, photo and/or graphic, and provides the high quality of the image. This is due to the decision to divide the image into layers and to compress each layer with the most appropriate compression method. Usually the text layer is compressed using the JBIG2, and the rest of the content is compressed using JPEG/JPEG2000/ZIP.

The basic three layers MRC model (Fig. 1) represents a color raster image as a composition of two color images, namely, a foreground layer, a background layer, and a binary image which is a mask layer.
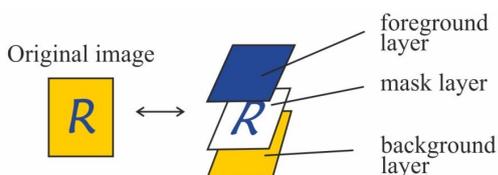

Fig. 1. Basic three layers MRC model of images [1]

The mask layer describes how to restore the original image using foreground and background images. If the mask pixel value is 1, then the corresponding foreground layer pixel is selected for the original image, and if the mask pixel value is 0, then the corresponding pixel is selected from the background layer.

There are two main types of MRC models of the image, specifically, region classification (RC) representation, and transition identification (TI) representation [1].

In the case of an RC representation, regions containing text and graphic images are identified and displayed on a separate foreground layer (Fig. 2, *a*). The mask is a binary image that contains large fragments of zeros and ones that define text and graphic regions, and the background layer contains the rest regions of the image, i. e. complex graphics and/or color images that do not include the foreground layer and mask. The pixels corresponding to the regions of text and graphic in the mask have an intensity of zero, and the pixels corresponding to the background regions are characterized by unit intensity.

In both the RC and TI representations, the background layer is defined in the same way. However, in the case of a TI representation, the mask and the foreground layer are formed differently (Fig. 2, *b*). The mask contains the contours of text symbols, drawings, and filled areas, and the foreground layer contains the colors of the elements whose contours are represented in the mask, that is, the colors of the text symbols and graphic objects. Then the foreground layer containing text or graphic colors is mapped through the mask onto the background layer.

One of the advantages of the MRC model is that such representation of the image does not destroy the structure of the original document and is convenient when encoding documents for further storage of images and information searching. If for some reason the MRC is not suitable for representing a particular image, you can select only one layer and compress it using the appropriate compression method. Another advantage of MRC is that it can work with classical compression methods such as JBIG-2 and JPEG-2000, ensuring compatibility with earlier methods such as JBIG, MMR and JPEG.
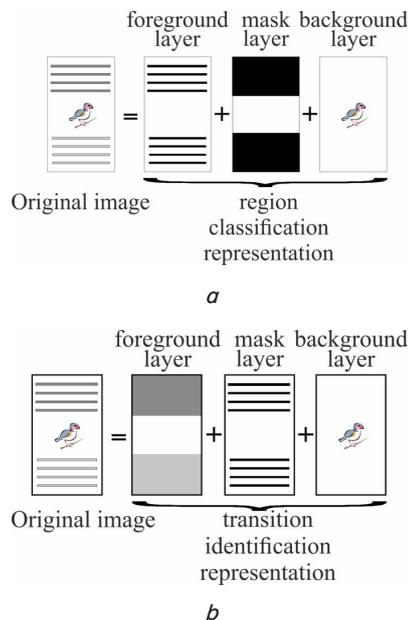

Fig. 2. The types of MRC models of the image:
a — RC representation; *b* — TI representation [1]

In [2, 3, 5–8], the RC representation was used mainly for compressing images containing text and illustrations. Segmentation was a preliminary stage of processing for the subsequent compression of images. Therefore, the construction of the RC representation was performed for the blocks of the original image, which greatly simplified the simulation.

Thus, in [2], the image of the scanned document was represented as a set of text and non-text regions, the latter including photos and graphics. In [2, 3], the RC representation of the MRC model was used to construct a two-level model of the image of a scanned document. At first such an image was divided into blocks, each of which belonged to one of five classes: background, text, photo, graphics, and also an overlapping block that could contain fragments of blocks of the first four classes. Then, for the overlapping blocks, the RC representation was formed, and a representation using tokens was used for text blocks [4]. In [5], the background layer of the RC representation, including illustrations, was divided into blocks, and the foreground layer, containing text and simple graphics, was represented using a map of basic color indices obtained as a result of quantizing the color feature of the images. In [6–8], to encode images on the screen containing text against the background of natural scenes, the image was divided into blocks when defining the RC representation, and for each block a binary mask was constructed by quantizing the intensity. This mask decomposes the image block into the foreground layer and the background layer.

The disadvantage of constructing the RC representation for the blocks of the original image is that in the future this implies the use of block methods for the segmentation of scanned documents. These methods are distinguished by high performance speed, but the way the image is divided into blocks affects the quality of segmentation. Insufficient quality of segmentation can lead to inaccurate determination of the boundaries of the text region, which can cause erroneous character recognition, and reduce the quality of the illustration when recovering from compression.

From the obtained RC representation of each block of the image in [2, 3, 5–8], a background layer containing complex graphics and photos was extracted. This layer, in turn, was represented by the coefficients of the wavelet transform [2], the cosine decomposition [3, 5, 6–8] or the difference representation [5, 6–8]. In this case, the first two representations require more time to identify the image block than the last, and, consequently, increase the time of scanned document image segmentation.

It is advisable to use the RC representation when the text and illustrations do not overlap, the color of the text symbols does not change, and the text symbols are placed against a background of constant intensity. Otherwise, it is better to use the TI representation, whose mask defines the boundaries of text symbols and graphic elements, and the foreground layer defines their color. Thus, in [1, 9, 10] the TI representation was formed for compressing images containing text and illustrations. In [1], one of the models is based on the TI representation and assumes that regions containing simple graphics and text are represented on the foreground layer, the mask layer contains the colors of text symbols and graphics, and complex graphics and photos are placed on the background layer. In [10], the TI representation for the MRC model was used to segment images of scanned documents in 2 stages. At the first stage, preliminary segmentation of image blocks was performed by optimizing the cost function. In the second stage, the quality of segmentation was improved by identifying non-textual regions that were identified as textual. For this, the image labeling obtained at the first stage was modeled by a Markov random field. In [11], the TI representation of the MRC model is adapted for segmentation of the image on the screen by decomposing them into foreground and background layers. The foreground layer contained text and simple graphics, the background layer contained text colors and simple graphics, as well as illustrations. It was assumed that if the intensity or color of the image block changes slowly, then the image block is assigned to the background layer, otherwise to the foreground layer. To estimate the rate of change in the intensity or color of an image block, this block was decomposed into cosine functions.

The disadvantage of the TI representation is that it is more difficult to construct than the RC representation, since for the latter it is necessary to divide the image only into text and non-text regions. In addition, the TI representation implies a lower image compression ratio than the RC representation. This is due to the representation of the binary mask layer. In the case of the RC representation, the mask contains large arrays of zeros and ones, and in the case of the TI representation, the mask layer contains the text whose size and shape are the same as on the original image.

Another problem with the TI representation is the appearances of a halo effect, specifically, a halo around the text and image graphic that affect the quality of the segmentation. This is due to the fact that the boundaries of the text and graphics of the scanned document image are blurred, which is why using the binary mask layer it is impossible to display smooth transitions of the boundaries completely on the background layer or on the foreground layer.

As for the RC representation, in the case of the TI representation, the background layer structure with illustrations was modeled using the wavelet transform [1] and cosine decomposition [9, 11], as well as using a Markov random field [10]. Using such representations of the background layer

with illustrations reduces the performance speed of scanned document image segmentation.

An analysis of the image models of scanned documents based on the RC representation [1–3, 5–8] and the TI representation [1, 9–11] showed that such models were mainly developed for image compression. The use of the considered representations for the scanned document image segmentation can reduce the quality of processing and performance speed. To solve the problem of scanned document image segmentation, it is necessary that the regions containing only text, only photographic images and only graphics should be located on separate image layers for their identification and classification. To select a feature system for homogeneous regions of scanned document images, a structural representation of the background and foreground layers containing, respectively, illustrations and text, is required.

3. The aim and objectives of the study

The aim of this work is to elaborate and research the structural representation of homogeneous regions on each image layer for a model of mixed raster content based on the classification of regions. This will allow elaborating a method of identification of these regions for scanned document image segmentation.

To accomplish the aim, the following objectives have been set:

– to elaborate a structural representation of homogeneous regions on each image layer of a mixed raster content model for scanned document images;

– to verify the elaborated model.

4. Representation of a scanned document image as a set of its constituent layers

In this paper, for scanned document images, it is assumed that text symbols are placed against a background of constant intensity, the color changes of these symbols can be neglected, text fragments and illustrations do not overlap. Then, to solve the problem of scanned document image segmentation, it is advisable to use the RC representation. However, this image model represents text and simple graphics on the foreground layer. To solve the problem of scanned document image segmentation, it is important that these regions are located on separate layers of the image. Therefore, it is proposed to represent the image as layers, each of which contains only one class of regions: the first layer contains only text regions; the second layer contains photo image regions and graphics. Then, in each layer, separate segments that will contain regions of interest can be extracted: on the first layer these are regions of text, on the second layer these are regions of photos and graphics. With this representation, we obtain image layers containing text, or a graphic or photographic image on a uniform background, which will allow us to select a feature system that identifies homogeneous regions of the scanned document image, namely, text, graphics, photos, background.

As a basis for the foreground layer, the representation of the text by tokens [4, 5] was chosen. This is representation by patterns of symbols together with the coordinates of the location of patterns in the image. Since the information on each symbol is redundant for the extraction of text frag-

ments at the scanned document image segmentation, it is proposed to use texture primitives for the structural representation of text. The modeling of the background layer containing illustrations is proposed to elaborate on the basis of a difference representation [5–8], which provides high performance speed identification of image fragments.

Suppose that there was a scanned document image without noise, from which, thanks to non-ideal sources for scanning and/or lighting, an image of a scanned document containing noise was obtained. We represent this image $I(x, y)$ in grayscale with intensity values from 0 to 255 and decompose it on a foreground layer $I_1(x, y)$ containing text on a uniform background, and a background layer $I_2(x, y)$ containing photo and/or graphic on a uniform background. The binary mask $I_3(x, y)$ is a layer which allows selecting either text regions or regions of photo and/or graphic on the original scanned document image. Then the scanned document image is represented in the following form:

$$I(x,y) = I_1(x,y) \cdot I_3(x,y) + \\ + I_2(x,y) \cdot (1 - I_3(x,y)) + I_n(x,y), \tag{1}$$

where $I(x, y)$ is the intensity of the pixel located in the xth row and the yth column of the noisy image, $I_n(x, y)$ is a random additive Gaussian noise with zero mean.

Equation (1) represents the image of the scanned document as decomposition into a layer containing regions of text, and a layer containing regions of photo and/or graphic images. Next, a representation of the image layers of a scanned document containing regions of interest is elaborated, namely, a representation of an image layer containing text on a uniform background and a representation of an image layer containing a graphic and/or photographic image on a uniform background.

4. 1. Representation of an image layer containing text on a uniform background.

The image of the symbol will be considered as a two-dimensional array of pixels, in which, depending on the size, font, and other characteristics, there are informative pixels and background pixels. This is necessary in order to take into account the spatial relationships between pixels in the symbol image. An image of a text region is represented as a set of symbol images arranged in a certain way. The connected set of pixels of a symbol, characterized by a certain set of features, forms a texture primitive of the image [13]. Regularly or almost regularly distributed in space, texture primitives, in turn, constitute a structural texture.

It is assumed that an image includes several regions, the textural differences of which are due to the change in type or spatial organization of texture primitives. These regions are fragments of text that have different sizes of symbols (heading or normal text), which usually do not overlap. The parameters of the texture representing the text are the horizontal and vertical distance between the text symbols and the number of symbols in the column and row of the image that are assumed to be unknown. The image of a text region is an ordered set of text symbols that represent the texture primitives of the structural texture. Different regions of text, such as normal text and heading, have the same pixel intensity and differ in the shape and size of texture primitives, as well as in the distance between them. Therefore, we present the text as an image of the regions of the structural texture on a uniform background [14–16].

Suppose that the set of image intensity values of the scanned document $I_1(x, y)$ includes the region of the structure texture $i(x, y)$. This texture $i(x, y)$ consists of equally spaced texture primitives, specifically, text symbols $t(x, y)$, and is determined by the formula:

$$i(x,y) = t(x,y) \sum_{k=1}^{L_y} \sum_{l=1}^{L_x} \delta(x - l\Delta x, y - k\Delta y), \tag{2}$$

where $\delta(\cdot, \cdot)$ is a delta function; Δx, Δy is the distance between the text symbols in the column and row of the image, respectively (Fig. 3); L_x, L_y is the number of symbols in the column and row of the image (Fig. 3); "*" is a convolution operator; $t(x, y)$ is representation of a text symbol. It is a random function of pixel intensity depending on spatial coordinates x, y, because the value of this function varies from one symbol to another.
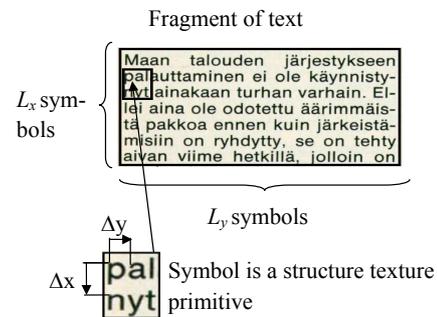

Fig. 3. Representation of the parameters of the scanned document image layer containing the text

The values of parameters of the heading region differ from the values of parameters of the normal text region. This means that regions containing the normal text differ in font size from regions containing text heading.

Let Δx_n, Δy_n be the distance between the symbols of the normal text in the column and line, respectively, and Δx_h, Δy_h be the distance between the heading symbols in the column and line, respectively. The difference between the corresponding parameters should not be less than x_{min}, y_{min}:

$$|\Delta x_h - \Delta x_n| > x_{min}, \quad |\Delta y_h - \Delta y_n| > y_{min}, \tag{3}$$

where x_{min}, y_{min} are model parameters.

Next, a representation of the layer containing a graphic and/or photographic image on a uniform background is elaborated.

4. 2. Representation of an image layer containing a graphic and/or photographic image on a uniform background

Let an image of a layer containing photo and/or graphic images on a uniform background without noise be represented as follows:

$$I_2(x,y) = \sum_{j=1}^{n} I_{2j}(x,y), \tag{4}$$

where $I_2(x, y)$ is an image layer containing graphic and/or photographic images on a uniform background; $I_{2j}(x, y)$ are specific graphic and/or photographic images, $j=1,…, n$; n is the number of graphic and/or photographic images of the layer. $I_{2j}(x, y)$ is defined by the expression:

$$I_{2j}(x,y) = I_2(x,y), \tag{5}$$

where $\chi_i(x, y)$ is the characteristic function of the set D_{2j}, on which the image regions containing the photo or graphic are defined, such that

$$\chi_i(x,y) = \begin{cases} 1, & (x,y) \in D_{2j}, \\ 0, & (x,y) \notin D_{2j}. \end{cases} \tag{6}$$

Each region containing a photo or graphic is represented as a piecewise constant function of the intensity $I_{2j}(x, y)$. This region is partitioned into m segments of uniform intensity R_{kj}. Here $k=1,...,m_j$; $m_j \in N$ is the number of segments of uniform intensity in the jth photo or graphic region, $j=1,...,n$ is the number of regions containing graphic or photographic images in the layer; all pixels of the R_{kj} region have the same intensity c_{kj} (Fig. 4). The same can be said about the background of the image, namely, the corresponding region in the image will also be characterized by the same intensity.

Fig. 4. Segments of uniform intensity of the scanned document image

Then the region of photographic and graphic image can be represented by the expression:

$$I_{2j}(x,y) = \sum_{k=1}^{m} \tilde{n}_{kj} \cdot \chi_{kj}(x,y), \quad (x,y) \in D_{2j}, \tag{7}$$

where $\chi_{kj}(x, y)$ is the characteristic function of the set R_{kj} defined by the following expression:

$$\chi_{kj}(x,y) = \begin{cases} 1, & (x,y) \in R_{kj}, \\ 0, & (x,y) \notin D_{2j} \setminus R_{kj}, \end{cases} \tag{8}$$

where

$$R_{kj} \bigcap R_{lj} = \varnothing,$$

with $k \neq l$, $k=1,...,m_j$; $l=1,...,m_j$,

$$D_{2j} = \bigcup_{k=1}^{m_j} R_{kj}.$$

Pixels with different intensity values fall into different sets R_{kj}. Then the intensity values of the pixels change at the boundaries of regions of uniform intensity, namely, the boundaries of regions of uniform intensity are characterized by edges. Images containing drawings or graphic are characterized by stronger edges, in contrast to images containing photographs.

Therefore, to determine regions of uniform intensity for the image segmentation and separating the graphic regions from the photo images, it is advisable to consider the strength of edges at the boundaries of regions of uniform intensity. The edge is a connected set of pixels lying on the boundary between two regions [17].

Fig. 5 shows an edge with a change of intensity from a low level to a high level. The edge is characterized by the strength h and slope of ramp θ. We can assume that the edge exists when its slope of ramp and strength are greater than a certain predetermined threshold.
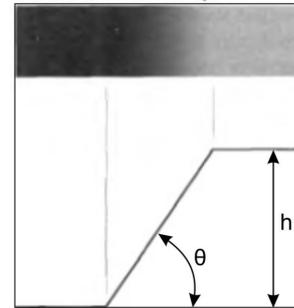

Fig. 5. The model of the ramp edge

Let $c(x, y)$ be the intensity value of a pixel (x, y), $c(x_1, y_1)$ be the intensity value of a pixel (x_1, y_1) from a neighborhood of a pixel with coordinates (x, y). We define the threshold value τ for determining the class of the analyzed image layer region, specifically is it a graphic image or a photographic one. Inequality

$$|c(x_1,y_1) - c(x,y)| < \tau, \tag{9}$$

shows that if low values of the strength of the image edges prevail, then it can be classified as a photographic image. Otherwise $I_{2j}(x, y)$ contains a graphic image.

5. Verification of the proposed mathematical model of the image of the scanned document

To elaborate a mathematical model of the image of the scanned document, the input and output data of the model were determined, and relations between them were established.

The verification stage implies a statistical analysis of the adequacy of the model, for which various procedures are used to compare model relationships with the properties of images of scanned documents.

The proposed mathematical model includes the representation of an image of a scanned document in the form of three layers (a foreground layer, a mask layer and a background layer) and representation of image layers. Representation of image layers includes a representation of the layer containing graphic and/or photographic images on a uniform background, and a representation of the layer containing text on a uniform background.

To verify the representation of a layer containing text on a uniform background, we take into consideration that textural differences of text regions are caused by a change in type or spatial organization of texture primitives. In order to show this, model images were constructed containing black rectangles in the places of text symbols of a text fragment of a scanned document image.

Hence the fragments of the images containing the heading and the normal text are cut. The model images were

generated with the same size as the cut images. The generated images were filled with black rectangles on a white background, the sizes of which were equal to the rectangles describing the most frequently encountered symbol. The distance between the rectangles in rows was the same, and the distance between the rectangles in columns was the same also. The most common letter was determined visually. In Fig. 6, *a*, the heading fragment is shown and the model image for it is shown in Fig. 6, *b*. In Fig. 6, *c*, the normal text fragment is shown and the model image for it is shown in Fig. 6, *d*.

Next, the correlation coefficient was calculated between the symbols of the cut fragment of the image and the constructed model image.

For the researched set of scanned document images, the correlation coefficient was obtained in the range of 0.5–0.65. For example, for the heading image from Fig. 6, *a*, and for the normal text image from Fig. 6, *c* the correlation coefficients 0.6035 and 0.5089 were obtained respectively. Based on the obtained values of the correlation coefficients, it can be argued that the fragments of scanned document images and the constructed model images are dependent. Therefore, the representation of the image layer containing the text on a uniform background (2.5) is adequate to the text fragments of scanned document images.
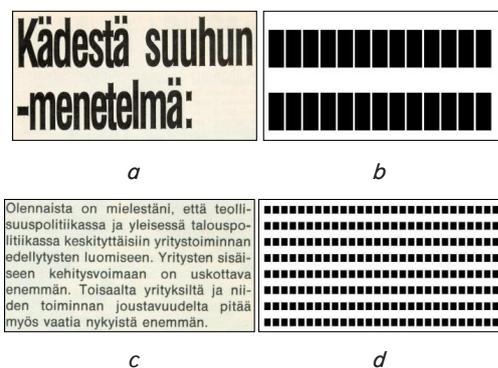

Fig. 6. Model images for fragments of scanned document:
a — fragment of heading; *b* — model image for the heading;
c — fragment of normal text; *d* — model image for
the normal text

To verify the representation of a layer containing text on a uniform background, note also that at the boundaries of different text regions, text parameters may vary. For example, the heading differs from the normal text by its parameters. The distance Δx_h between the heading symbols and the distance Δx_n between the symbols of the normal text in the column should not be less than x_{min}. Similarly, the distance Δy_h between the heading symbols and the distance Δy_n of the normal text in the image line should not be less than y_{min}. If you select, for example, $x_{min}=5$ pixels, and $y_{min}=12$ pixels, then inequalities (3) will be satisfied for the text shown in Fig. 6, *a*, *c*.

To verify the representation of a layer of an image containing graphic and/or photographic images on a uniform background, it was analyzed how the photographic image differs from the graphic one.

Since the grayscale images with pixel intensities from 0 to 255 were analyzed in the paper, the edges were considered. It was noticed that often the strength of the edge at the boundary between the regions of the graphic is greater than between the regions of the photographic images. This property allows distinguishing regions containing photo images from regions containing graphic. To verify this, values of the probability for the edge strength for images containing these regions were determined. For this purpose, a spatial filtering of image regions containing graphic and photographic images was performed using masks of the Sobel filter, which calculates an approximate value of the image intensity gradient and detects the boundaries of objects in the image. In Fig. 7, the plot of the probability of the edge strength on the edge strength for researched images containing photo, and for images containing graphic is shown.
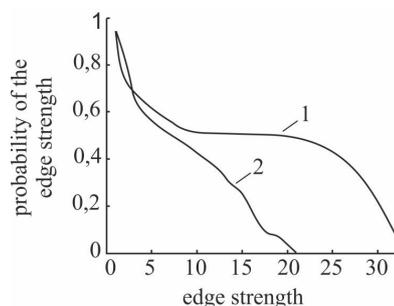

Fig. 7. Dependence of the probability of the edge strength on the edge strength: 1 — for images containing a photo; 2 — for images containing a graphic

Analyzing Fig. 7, we note that small values of the edge strength are characteristic of both images containing photos and images containing graphics, while large values of the edge strength are present mainly in the last images. Therefore, images containing graphics are indeed different from images containing photos by the presence of large values of the strength of the edges between regions of uniform intensity.

Also the standard deviation of values of the edge strength was calculated for some images containing regions of graphics and photo images. For images containing graphic regions, it was 98.7–102.8, for images containing photo regions, it was 63.4–68.2, which also indicates that variability of the edge strength for a graphic image is higher than for a photo image.

6. Discussion of the research results of the proposed mathematical model for the representation of scanned document images

The elaborated mathematical model of representation of the scanned document images is the expansion of the MRC model. A research of the elaborated mathematical model of the scanned document image showed that the structural representation of each of the layers of the proposed model is adequate to regions of text and illustrations of the scanned document images. The advantage of this representation is in its use to select features of text regions, graphics and photos for the scanned document image segmentation in order to increase the performance speed of such images [18]. The proposed image representation can be used to solve the problem of scanned document image segmentation by extracting and labeling the image regions of interest.

The elaborated model of representation of the image contains restrictions (3) on the distance between the text symbols of the normal text and the symbols of the heading. If the symbols of the text are unevenly spaced at different distances, then inequalities (3) are not satisfied. Another limitation (9) of the model allows representing graphic or photographic images, depending on the predominance of high or low values of the edge strength in the image. If the analyzed images contain darkened photographic images, the intensity values of which are close to zero, or the grayscale intensity values of which are close to each other, then this inequality will also not be satisfied. Scanned document images of this type for which at least one of the considered inequalities is not satisfied cannot be represented using the proposed model. These restrictions must be considered when choosing the analyzed documents.

The disadvantage of the elaborated representation of the scanned document image, as well as the MRC model, is too generalized modeling of images in the form of layers (1), actually representing a "black box". Such representation makes it difficult to decompose an image into layers in the process of segmentation and compression, which is reflected in the literature [2, 5, 18], and to formalize a test for the adequacy of this representation to the scanned document images. Therefore, in the future, the research will be directed to improving the decomposition of the scanned document images into layers.

7. Conclusions

1. A mathematical model of the image of a scanned document in the form of a structural representation of the regions of the original image is proposed. This model differs from those known in the literature as follows. The image of the scanned document is presented in the form of layers containing only one class of region, as well as in the form of separate segments that contain regions of interest on a uniform background: either text, or graphics, or a photo image. The proposed model for the representation of a scanned document image allows elaborating a method of identification of these regions to solve the problem of these images segmentation for the purpose of further archiving and storage.

2. Verification of the proposed model showed that this model is adequate to the scanned document image regions and can be used for structural representation of homogeneous regions on each image layer. Therefore, the proposed model for the representation of scanned document images is recommended to be used in segmentation problems of such images to increase the performance speed.

References

1. De Queiroz R. L., Buckley R. R., Xu M. Mixed Raster Content (MRC) Model for Compound Image Compression // Visual Communications and Image Processing '99. San Jose, 1999. Vol. 3653. doi: https://doi.org/10.1117/12.334618

2. Banupriya D., Sundaresan M. Enhanced hybrid compression method for compound images // International Journal of Computational Inteligence and Informatics. 2016. Vol. 6, Issue 1. P. 18–24.

3. Maheswari D., Radha V. Enhanced layer based compound image compression // Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India – A2CWiC '10. 2010. doi: https://doi.org/10.1145/1858378.1858418

4. Huttenlocher D., Felzenszwalb P., Rucklidge W. Digipaper: a versatile color document image representation // Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348). 1999. doi: https://doi.org/10.1109/icip.1999.821601

5. Melnakshi P., Jabanesh P. An Efficient Compound Image compression by exploiting spatial correlation // International Journal of Engineering Research&Technology. 2014. Vol. 3, Issue 1. P. 2962–2967.

6. High efficiency video coding (HEVC) text specification draft 10 / Bross B. et. al. // Doc. JCTVC-L1003. Switzerland, 2013.

7. Overview of the High Efficiency Video Coding (HEVC) Standard / Sullivan G. J., Ohm J.-R., Han W.-J., Wiegand T. // IEEE Transactions on Circuits and Systems for Video Technology. 2012. Vol. 22, Issue 12. P. 1649–1668. doi: https://doi.org/10.1109/tcsvt.2012.2221191

8. Xu J., Joshi R., Cohen R. A. Overview of the Emerging HEVC Screen Content Coding Extension // IEEE Transactions on Circuits and Systems for Video Technology. 2016. Vol. 23, Issue 1. P. 50–62. doi: https://doi.org/10.1109/tcsvt.2015.2478706

9. Juliet S. E., Sadasivam V., Florinabel D. J. Effective layer-based segmentation of compound images using morphology // Journal of Real-Time Image Processing. 2011. Vol. 9, Issue 2. P. 299–314. doi: https://doi.org/10.1007/s11554-011-0223-8

10. Haneda E., Bouman C. A. Text Segmentation for MRC Document Compression // IEEE Transactions on Image Processing. 2011. Vol. 20, Issue 6. P. 1611–1626. doi: https://doi.org/10.1109/tip.2010.2101611

11. Minaee S., Wang Y. Screen content image segmentation using least absolute deviation fitting // 2015 IEEE International Conference on Image Processing (ICIP). 2015. doi: https://doi.org/10.1109/icip.2015.7351413

12. Raspoznavanie lic na gruppovyh fotografiyah s ispol'zovaniem algoritmov segmentacii / Sherstobitov A. I., Fedosov V. P., Prihodchenko V. A., Timofeev D. V. // Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki. 2013. P. 66–73.

13. Haralik R. Statisticheskiy i strukturnyy pohody k opisaniyu tekstur // TIIER. 1979. Vol. 67, Issue 5. P. 98–120.

14. Dunn D., Higgins W. E., Wakeley J. Texture segmentation using 2-D Gabor elementary functions // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1994. Vol. 16, Issue 2. P. 130–149. doi: https://doi.org/10.1109/34.273736

15. Polyakova M. V., Krylov V. N., Ishchenko A. V. Elaboration of the transform with generalized comb scaling and wavelet functions for the image segmentation // Eastern-European Journal of Enterprise Technologies. 2014. Vol. 5, Issue 2 (71). P. 33–37. doi: https://doi.org/10.15587/1729-4061.2014.27791

16. Polyakova M. V., Krylov V. N. Opredelenie granic segmenta uporyadochennoy tekstury na izobrazhenii s odnorodnym fonom v prostranstve preobrazovaniya Gil'berta // Elektronika i svyaz'. 2007. Issue 5. P. 54–61.

17. Gonsales R. S., Vuds R. E., Eddins S. L. Cifrovaya obrabotka izobrazheniy v srede MATLAB. Moscow: Tekhnosfera, 2006. 616 p.

18. Combined method for scanned documents images segmentation using sequential extraction of regions / Polyakova M., Ishchenko O., Volkova N., Pavlov O. // Eastern-European Journal of Enterprise Technologies. 2018. Vol. 5, Issue 2 (95). P. 6–15. doi: https://doi.org/10.15587/1729-4061.2018.142735