Available online at: https://ijact.in

Date of Submission	17/10/2019
Date of Acceptance	03/12/2019
Date of Publication	31/12/2019
Page numbers	3514-3524 (11 Pages)

Cite This Paper: Fazilov S, Mirzaev N, Radjabov S, Dadakhanov M, Asraev M, Shamsiev F. State of the art of writer identification, 8(12), COMPUSOFT, An International Journal of Advanced Computer Technology. PP. 3514-3524.

ISSN:2320-0790

An International Journal of Advanced Computer Technology

STATE OF THE ART OF WRITER IDENTIFICATION

Shavkat Kh. Fazilov[1], Nomaz N. Mirzaev[1], Sobirjon S. Radjabov[1*],
Musokhon Kh. Dadakhanov[2], Mukhammadmullo A. Asraev[3], Farkhodbek M. Shamsiev[4]

[1]Scientific and innovation center of information and communication technologies, Tashkent university of information technologies named after Muhammad al-Khwarizmi, 17A, Buz-2, Tashkent, 100125, Uzbekistan
[2]Namangan State University, 316, Uychi, Namangan, 160136, Uzbekistan
[3]Ferghana branch of Tashkent University of information technologies named after Muhammad al-Khwarizmi, 185, Mustaqillik, Ferghana, 150118, Uzbekistan
[4]Tashkent University of information technologies named after Muhammad al-Khwarizmi, 105, Amir Temur, Tashkent, 100200, Uzbekistan
*s_radjabov@yahoo.com

Abstract: The paper analytically reviews the methods and algorithms for solving problems arising in the development of writer identification offline systems. The main applied problems that are solved on the basis of the processing and analysis of handwritten text are considered; the classifications of writer identification systems, as well as the structure of offline systems are given. The paper also reviews handwritten text image databases and shows the tasks to which these databases are aimed. An attempt is made to systematize the algorithms for the preliminary processing of handwritten text images, depending on the task they are tackling at the stage under consideration. Based on the analysis of the most common algorithms of forming the handwritten text feature space, these algorithms are classified into ones for extracting geometric, structural, topological, statistical, and spectral features. A review of the algorithms for selecting each category from the listed features is conducted. Recognition methods used to identify the writer are also reviewed. The results of applying these methods, as well as their benefits and drawbacks are presented.

Keywords: online and offline writer identification systems; handwritten text database; handwritten text pre-processing; binarization; line segmentation; word segmentation; feature extraction; writer identification

I. INTRODUCTION

The issue of formalizing and automating the process of identifying the features of a handwritten text to tackle the tasks of recognition thereof was raised as far back as the earliest stages of the development of computer technologies for image processing. The complexity of this task is determined by its falling under the class of problems that are easily formulated verbally, but poorly formalized and difficult to solve.

Postal services, banking chains and manufacturers of computing equipment are considered the main institutions requiring that the tasks of handwritten text analysis should be achieved and allocating significant funds to finance relevant work.

Postal services are interested in automating the process of sorting and delivering correspondence, i.e. in the automatic reading and processing of addresses on envelopes and post packages. A number of countries have so-called postal banks, which operate with postal money orders, which are filled in by customers manually. These documents circulating in large quantities also need to be read and processed automatically [1,2].

A variety of needs related to the problems of handwritten text analysis arise in the activities of banks. It is the automatic processing of banker's checks, policies, and other handwritten payment documents. This includes both reading the contents of documents per se and identifying customers based on the validation of their signatures [3].

In recent years, large computer manufacturers trying to create new markets for their products have shown an increased interest in the analysis of handwritten text. With the advent of the pen technology (instead of the keyboard, a user has only a light pen to write on the screen or on a special panel) and computers that work entirely on this principle, the leading manufacturers of computing machinery began to allocate large funds for research in handwritten text analysis.

In view of the above, it can be said that the computer processing and analysis of handwritten text is now on the rise as a scientific discipline and information technology due to the great interest in this field shown by the world of commerce, computer companies, and the scientific community.

Most of the tasks tackled on the basis of handwritten text analysis are considered in forensic science [4], but there are also those related to image processing [5,6] and graphology [7]. That being said, handwritten text recognition is now the most popular and widely studied problem [8,9]. It is not completely solved, and this is determined by the fact that the existing algorithms for the recognition of a continuous handwritten text provide less accurate results than in recognizing a handwritten 'printed' text. Higher performance can only be achieved with the use of additional contextual and grammatical information.

The next task to be solved on the basis of processing and analyzing a handwritten text is to determine the character of its writer. The connection between a person's handwriting and his/her character is studied by graphology [7]. The issues of automating the process of identifying the features of a handwritten text in order to determine the writer's character are understudied.

The analysis of a handwritten text makes it possible to determine the gender and age group of the writer thereof [10,11]. Also, after analyzing the previously recognized words of a handwritten text, it is possible to create a rather comprehensive picture about the writer of the letter (level of education, field of activity, etc.) [4].

The change in handwriting due to injury, illness, and emotions experienced at the time of writing the text allows determining the mental and physical state of the person.

In addition to the above, the automation of the processing and analysis of handwritten texts make it possible to create tools for the restoration of ancient manuscripts [12], allowing researchers to use electronic copies of the manuscripts, thus protecting them from possible damage.

It should be noted that, according to many experts, the task of the writer identification is the most frequent and important task to be solved based on the analysis of handwritten texts [13,14].

The importance of the writer identification is growing due to the tendency of an increase in the number of crimes committed with the use of handwritten texts, as well as due to the fact that recent developments in this field and the effectiveness of the systems give the right to consider it as one in competition with reliable physiological methods of identification, such as DNA fingerprinting or fingerprint identification [15]. In addition, the solution to the problem of the writer identification can be widely used in various fields of human activity [16,17].

The purpose of this paper is to conduct an analytical review of methods and algorithms for solving problems arising in the development of writer identification offline systems.

The structure and content of this paper is composed as follows.

In the second section, a classification of the writer identification systems, as well as the structure of offline systems are presented.

The third section of the paper is devoted to handwritten text databases. Creating a database is a time-taking process, so their availability is a very important factor in developing and evaluating the operation of systems in all research areas. This section will be useful to researchers in choosing the most appropriate databases for evaluating the algorithms and systems developed by them.

Algorithms for preprocessing handwritten text images are covered in the fourth section. There are a great number of publications on the algorithms of this stage of development of identification systems. However, in this section, an attempt is made to systematize these algorithms depending on the problem they are solving at the stage under consideration.

The fifth section is devoted to the problem of the formation of the handwritten text feature space. The considered algorithms are classified into five categories: geometric, structural, topological, statistical and spectral ones.

The sixth section provides an overview of the recognition methods used in the writer identification. The results of the application of the methods, as well as their advantages and disadvantages are presented.

The conclusion contains the key findings of the analysis of the state of the art of the problem under consideration and a set of tasks that, in our opinion, need to be solved when developing a writer identification offline system, taking into account the specific features of the style of national characters.

II. The Structure of Writer Identification Offline Systems

There are two main approaches to identifying the writer of a handwritten text: identification in the mode of current character input (online) and identification using a previously written document (offline). The task of the writer online identification is considered less complicated than that of offline identification, since the online mode involves combining the process of forming the input images and entering them into the system. This allows the system to 'observe' the process of styling of the input characters, which in turn allows it to receive not only graphical information but also data on the style of entering input images, such as the dynamics of the pen movement and, in some cases, the pressure force that are not available in the off-line mode [9,15,18-20].

Further, the paper reviews methods for solving the problems of the development of writer identification offline systems only.

Fig. 1 shows the standard structure of a writer identification system [20]. At the first stage, the system receives handwritten text samples as images. At the second stage, the received data are pre-processed. Further, necessary features are extracted from the handwritten text images. Then, the features are used to calculate the degree of belonging of the text to various writers. The writer with the highest similarity rating is given by the system. It should be noted that the choice of the method for tackling the task of each stage depends heavily on the methods used at the previous stages.

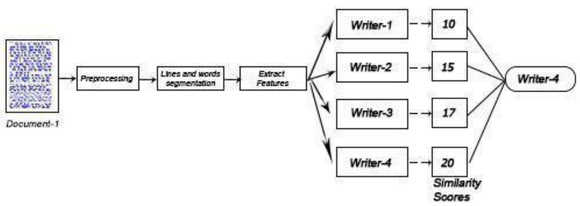


Figure 1: The standard structure of offline writer identification system [20]

In addition to the above classification, identification systems can be divided into text-dependent and text-independent ones, depending on whether the system knows the text to be entered by the user and whether the system uses this information. If the text-dependent identification is used, both fixed texts and those generated by the system and proposed to the user can be used. For example, the verification of identity using the signature widely employed when using credit cards is a special case of text-dependent identification. Text-independent systems are designed to process a free text and do not require the same characters for comparison. The text-independent approach uses a set of features whose components describe global statistical features extracted from the entire text image. Therefore, it can be called the texture analysis approach.

As noted previously, the task of the writer online identification is considered less complex than that of offline identification. One of the reasons is the fact that at the stage of obtaining the original image in case of offline identification, there is a high probability of its distortion. Such distortions may originate from technical reasons (artifacts and other defects that occur when scanning an image). In addition, the human factor is also of great importance (interferences during scanning caused by the user's fault), and the state of the scanned document may be far from ideal (spots, blots, bends, cuts, etc.).

However, we think that the most important factor that makes it difficult to tackle the task of identification using a handwritten text is the high variability of the same person's handwriting. Moreover, the characteristics of a text depend heavily on:

– the conditions under which it was written;
– the writer's emotional state;
– the writer's age and profession;
– the writer's intentions (deliberate change in his/her handwriting).

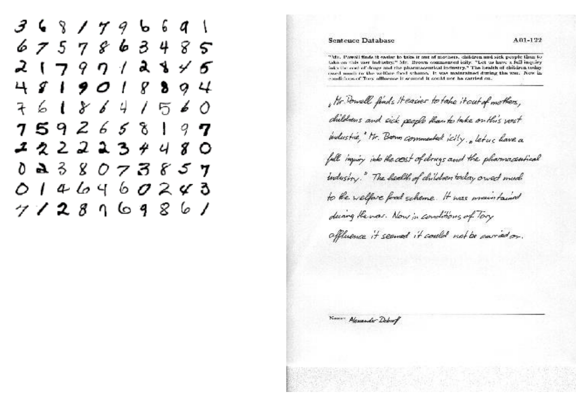
III. HANDWRITTEN TEXT DATABASES

Collecting samples to create databases is a long, labor-intensive process, since it involves obtaining the maximum possible diversity of samples. The availability of standard databases not only solves the problem of the source data for the researcher, but also allows him/her to objectively evaluate the effectiveness of the developed algorithms and the system as a whole, as well as to compare them with the existing ones.

So far, a large number of databases that hold the samples of handwritten texts in various languages have been created throughout the world. Most of them, however, require additional processing to verify the functionality of the developed algorithms. Further, there is a description of the databases that are the most popular with researchers.

The MNIST database [21] contains 60,000 handwritten numerals for learning and 10,000 ones for testing. Moreover, the images in the database are normalized in size and centered inside the image. In the database, the samples are stored in the form of grayscale images of handwritten numerals of 20x20 pixels inscribed in a 28x28 square (Fig. 2.a). The object in the image is centered by searching for the center of mass. The database is ideal for researchers who need to try out learning methods on real data without the extra effort to be spent on formatting and pre-processing.

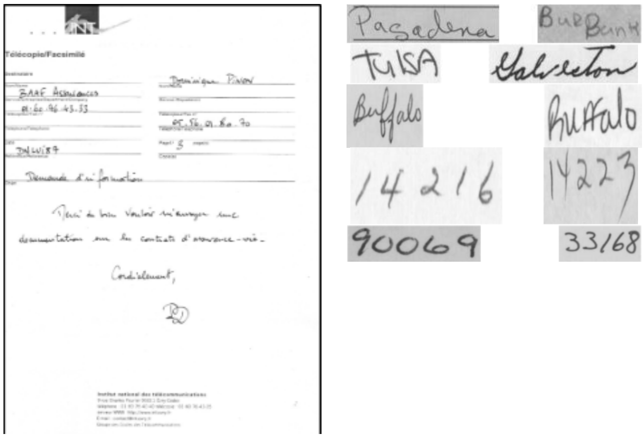
The IAM database [22] is a set of grayscale images of English handwritten text, scanned with a resolution of 600 dpi. The database contains 1,539 scanned handwritten pages of 657 writers (Fig. 2.b). This text database can be used to test the methods of segmentation, handwritten text recognition, as well as the writer identification.



a b

Figure 2: Samples of handwriting from MNIST (a) and IAM (b)

The Centre for Analysis and Recognition of Documents at the University of Buffalo has created the CEDAR [24] mainly intended to conduct research in the field of automatic processing of postal addresses on envelopes. The samples contain 5,632 handwritten samples of city names, 4,938 ones of state names and 9,454 zip codes (Fig. 3.b). The data are divided into separate subsets for learning and testing. This database can be used to evaluate the results of tackling a number of tasks of handwritten text analysis, including the tasks of text and word segmentation, the text recognition, and the writer identification.



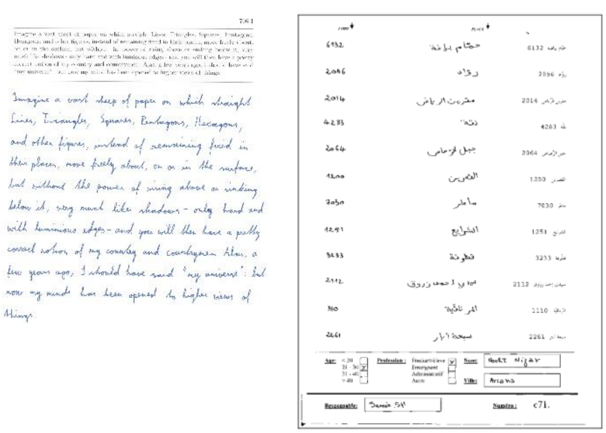
a b

Figure3. Samples of handwriting from RIMES (a) and CEDAR (b)

The CVL [25] is a database of handwritten samples designed to evaluate the performance of systems of handwritten text recognition, word segmentation, and the writer identification. The database consists of sets of handwritten texts by 310 writers, representing one page of a text in German and six pages in English (Fig. 4.a).

A large number of image databases: IFN/ENIT [26], ARABASE [27], CENPARMI-Arabic [28], Al-Isra [29], KHATT [30], QUWI [31], etc. are available for testing the systems for the identification of the Arabic text writer.

The IFN/ENIT database [26] contains images of the names of cities and villages of Tunisia, together with a zip code, written by hand. The image database consists of specimen of handwriting of 411 writers. The total number of words (names of cities/villages, zip code) in the database is more than 26,400, consisting of 210,000 letters and numbers. All the data are stored in the database as binary images with a resolution of 300 dpi (Fig. 4.b). The database is designed to evaluate pre-processing algorithms, recognition of Arabic handwritten words, as well as the writer identification.



a b

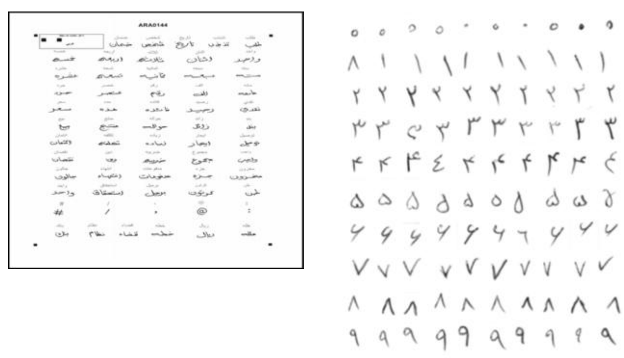
Figure4. Samples of handwriting from CVL(a) and IFN/ENIT(b)

The ARABASE database [27] is intended for both the writer online and offline identification and the handwritten text online and offline recognition. The database holds segmented images of paragraphs, words, letters, numerals, and signatures. It is also notable that a toolkit is attached to this database in the form of a program that allows tackling traditional tasks of analysing documents in image databases.

To create the CENPARMI-Arabic database [28], samples of Arabic handwritten text by 328 writers were used. It holds segmented images of numerals, characters, lines and words (Fig. 5.a). The database is divided into three groups: the first group of images consists of handwritten samples by 100 writers, the second one contains handwritten samples of 228 writers, and the third group is a mixture of samples from the first and second groups. The database can be used for recognition of Arabic characters and numerals, as well as for word segmentation.

The Al-Isra database [29], created by researchers at the University of British Columbia, is a large collection of handwritten samples containing words, numerals (Fig. 5.b), signatures, and sentences. The samples are provided by 500 students of Al-Isra University (Jordan). The database holds 500 Arabic sentences, 37,000 words, 10,000 numerals, and 2500 signatures. The database can be used for the handwritten text recognition and the writer identification.

KHATT [30] is an Arabic handwritten text database that consists of 1000 samples of the same number of writers from different countries (Fig. 6.a). Each sample text is scanned in three different resolutions: 200, 300 and 600 dpi. The database holds 2,000 paragraphs of text. It is provided with tools that make it possible to segment the images of texts contained in the database into lines and paragraphs. The database is designed to assess the performance of the system when tackling tasks of pre-processing, segmentation, as well as the writer identification.



a b

Figure5. Samples of handwriting from CENPARMI-Arabic(a) and Al-Isra(b) databases

The QUWI database [31] stands out of the other Arabic handwritten text databases. The database is an extensive collection of text samples written by 1017 writers of different cultural and educational backgrounds. Its unique feature lies in the fact that it is a bilingual database holding four pages of handwritten text by each writer, two of which are written in English, and the other two are written in Arabic. This allows using the database in solving a number of interesting cases of the writer identification. Another specific feature thereof is that 2 pages in different languages contain free text, and the other two contain predefined text (in Arabic and English). This allows using the database for testing both text-dependent and text-independent systems for the writer identification.

Chinese, Japanese and Korean are the most common Asian languages that partially or fully use Chinese characters. To conduct research in the field of analysis and processing of such characters, a huge number of handwritten text databases have been created. Let us consider some of the most common ones.

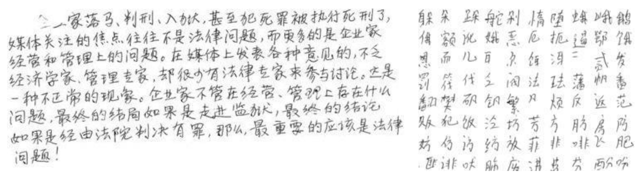
PE92 [32] is a unique database holding 100 handwritten text images consisting of 2,350 letters of the Korean phonetic alphabet (Hangul). 70 pages are written by more than five hundred writers, and 30 ones are written by one person. The writers filled the pre-defined fields with characters. The database can be used for the character recognition and the writer identification.

The HCL-2000 database [33] consists of a set of frequently used Chinese cursive characters by 1000 writers (Fig. 6.b). In addition to the information about the 3755 characters that make up the database, there is information related the writer's age and gender. This allows using the database in tackling the tasks of the writer identification, determining his/her age and gender based on the analysis of the character images.

a b

Figure6. Samples of handwriting from KHATT (a) and HCL-2000 (b)

Currently, the CASIA database [34] is one of the largest Chinese character databases and includes data for both the offline and online writer identification. However, for the offline identification, the database was created as follows. Each writer (2,039 people in total) wrote five pages of handwritten text. The pages were scanned with a resolution of 300 dpi, and the characters were segmented (Fig. 7). Each image was also divided into the background and the text. The division makes it possible to get a high quality binary image using a simple transformation.



a b

Figure7. Samples of handwriting text (a) and segmented hieroglyphs (b) from CASIA

In addition to the above bases, there are a large number of open access handwritten text databases [35-41], which can be used to evaluate the solution of various problems arising in the development of writer identification offline systems. A detailed description of the databases is given in [42].

IV. IMAGE PREPROCESSING

The preprocessing of the original image is an important and integral part of image processing and analysis systems. At this stage, the following tasks are tackled:

– improving the quality of the original image;
– the image binarization;
– line and word segmentation in the image;
– the image normalization.

The results of the stage significantly affect the quality of the system as a whole. The following are the most frequently used algorithms that allow tackling the above tasks.

A. Image Enhancement

The original image enhancement is carried out using image processing methods in order to eliminate defects arising in scanning, as well as improving the quality of the work of other pre-processing algorithms.

Such actions as blurring, contrast enhancement, brightness histogram equalization, etc. are often applied to the original image. [5,6].

B. Binarization

The image binarization is conducted in order to significantly reduce the amount of processed information by transforming the original image (usually a grayscale one) to a binary one. The smallest loss of the information important to recognition and the maximum elimination of noise in the image is an important condition for the transformation.

This stage is closely related to the subsequent stages of the image processing and analysis used in recognition, and, from a practical standpoint, the binarization algorithm performance is evaluated based on the operation of the system as a whole.

The binarization image transformation algorithm is extremely simple, and the main problem is to choose the binarization threshold value based on some a priori information about the image. The choice of the binarization threshold is based on the assumption of a pronounced bimodality of the image brightness histogram.

Binarization algorithms are classified into global and local by the type of the binarization threshold. Binarization with a global threshold implies dividing the original image into the background and objects using the threshold defined for the entire image [43-47]. Local thresholds are defined in the local block (window) of the image, and the transformation using the threshold is performed only in the block under consideration [47-52]. Note that transformation with a global threshold is much faster than transformation with a local one. However, in practice, the latter algorithms are more often used, since they give better results with great differences in brightness in various parts of the text.

The binarization problem can also be considered as a clustering problem with an a priori known number of clusters equal to 2. This implies that known clustering methods can be used to solve this problem [53,54].

C. Segmentation

The algorithms of this stage allow selecting areas of interest in the image for their subsequent analysis.

Proper segmentation is an essential component of offline identification systems. Recently, a large number of segmentation algorithms for handwritten objects in an image have been developed. The algorithms can be classified depending on their purpose: line segmentation algorithms, word segmentation algorithms.

***Line Segmentation*:** Line segmentation in a handwritten text is a difficult task due to the fact that lines can be non-parallel, bend, be close to each other, the text elements of different lines can be superimposed on each other [55]. All the factors, which are not found in machine-printed texts, do not allow using directly known algorithms for machine-printed line segmentation.

Algorithms of the handwritten text line segmentation are divided into three main groups: algorithms based on the analysis of horizontal projections; blur algorithms; algorithms based on the Hough transformation. In addition to the mentioned ones, there are also algorithms that cannot be included in any of the listed groups or be combined into one group due to the absence of a generalizing feature.

The algorithms based on the analysis of horizontal projections segment the lines by searching for the minimum ‘islands’ on the calculated horizontal projection of the original image. The algorithms are often used for machine-printed line segmentation, but they can be adapted to handwritten line segmentation [56,57]: the image is divided into vertical bars, and an analysis of horizontal projections is conducted for each of them.

In blur algorithms [58,59], in a binary image, successive black pixels are blurred along the horizontal direction. If the distance between the white space is within the specified threshold, it is filled with black pixels. The areas of related components in the blurred image are taken as lines.

As is known, the Hough transformation can be used to find straight lines on images [5,6]. The slant of the handwritten lines in the image can be determined by applying the Hough transformation to the center of gravity of each related component of the analysed image [60]. Using the proximity criteria and the directions of continuity of related components, handwritten text lines are segmented.

Intersecting components are a problem for algorithms based on the analysis of horizontal projections (as they increase the value of the projection profile in the areas where its minimum should be) and blur algorithms (as they use connected components of text pixels to build lines), but have little effect on some algorithms of the third group, in particular, [61].

To search for intersecting elements from different lines, it is possible to use such features as the size of the text cohesion components, the fact that one component is assigned to several lines or, on the contrary, that it is not assigned to any line. After such questionable components have been found, it is necessary to determine whether they belong to a certain line or whether they should be decomposed into elements belonging to different lines. Such vertical decomposition of components is a difficult task.

A simple solution to the task involves cutting the component into parts by horizontal lines [62], but more nuanced approaches, such as extracting individual strokes, can be applied [63].

Word Segmentation: Normally, word segmentation is performed after the line segmentation stage. However, there are algorithms [64] that first allow segmenting words, and then segmenting lines based on an assessment of the proximity criterion and the direction of continuity of the connected components.

The difficulty of the handwritten text word segmentation lies in the fact that, unlike a machine-printed text, where the distance between words is more or less constant, and the distances between characters inside a word is much less than those between words, in a handwritten text, the distance between words can vary greatly.

This task is tackled as follows. The words are formed from connected components of the text of the line under consideration based on the analysis of the distance between the components. The problem of attributing this distance to the distance between the components of a word or the distance between words is solved by classification methods.

Given that the distance between connected components is sensitive to the forms of the components, a large number of methods for calculating the distances have been proposed. Thus, for example, in [65], it was proposed to calculate the distance as the Euclidean distance between the nearest points of adjacent components or between the rectangles into which the components are inscribed. When cross-cutting such rectangles, it is recommended to use the minimum distance determined among the distances between points of adjacent connected components located on the same horizontal line.

[66] Proposes a different approach to determining the distance: each connected component is inscribed into a convex polygon, and the center of gravity of the polygon is determined. Further, the points of the centers of gravity of the adjacent connected components are connected by a right line, and the points of intersection of the convex polygon lines and the right line are determined. The

distance between adjacent connected components is taken as the distance between the intersection points.

D. Normalization

The stage of pre-processing is aimed at bringing the images of handwritten objects to a certain standard form without significant loss of individual features of their style.

The most commonly used methods are the correction of a word or a line in general [67,68]. Correction algorithms for the slant angle, the width normalization and vertical scaling of a word or part thereof are less commonly used. The influence of applying the algorithms on the writer identification using hidden Markov models was studied in [69]. The results of the studies showed that the pre-processing conducted in the form of word vertical scaling turned out to be the most effective for extracting features using this method.

Before using some methods of feature extraction, methods of skeletonization of handwritten objects are used [5,64,70,71]. The methods are used in order to extract handwritten text features that are resistant to changing the width of the text strokes [72,73].

V. Feature Extraction

The object feature extraction is one of the main tasks to be tackled by many systems for processing and analysing visual information. This task has been the focus of attention of many researchers for a rather long time, and in many publications, in particular [74-76], the importances of issues related to the extraction of characteristic features are emphasized.

An analysis of publications devoted to image processing and computer vision, in particular [5,6,74-79], shows that various authors consider the task of extracting the features of objects given as images and offer various classifications of the methods used to tackle this task. For example, in [74] all the features used in the description and recognition of images are divided into three categories: physical, mathematical and structural ones. In [75], the features used in the description and recognition of images are classified into four groups: geometrical, spectral, statistical, and topological ones. A similar classification of the features of objects represented as images and the corresponding methods for the extraction of characteristic features with a certain level of detail is given in many sources of literature. However, in order to facilitate the analysis of the task under consideration, let us divide all the methods for feature extraction and, accordingly, all the features used in the description of recognizable objects represented as images into five categories: geometrical, structural, topological, statistical and spectral ones.

As it is known, there are many languages in the world, and each of them, due to its uniqueness, requires a specific approach to the task of the writer identification. Therefore, it is obvious that the task of the writer identification varies depending on the language. So, textural features obtained using a multi-channel Gabor filter and gradation adjacency matrices are generally accepted for the Chinese language. As to the English language, the range of such features is very wide: from micro-level features to macro-level ones and edge distribution. Moreover, studies have been conducted for other languages (Arabic, Persian), too. Combinations of some textural and allographic features, as well as of hybrid spectral and statistical measurements, multichannel filters (Gabor, XGabor) are used to obtain the individual characteristics of writers' handwritings. It follows that the features should be selected based on the distinguishing characteristics of each language.

A. Geometrical Features

The geometrical approach includes those methods that are based on determining the geometrical characteristics of objects represented as images [5,6,75]. This group includes methods related to the representation of the geometry of the object contour, which are quite specific, but widely used in the field of image processing and computer vision [5,6].

[80,81] describe algorithms for generating a set of geometric features of the English text images, which consists of the height of the three main zones of a handwritten text, the distance between the connected components, and the angle of the style slant. This set of features made it possible to achieve 90% identification accuracy using the k nearest neighbours' method in a sample of 100 images by 20 writers.

In [82], eleven features were identified for tackling the task of identifying the writer of documents containing the numbers: aspect ratio, the number of end points, the number of transitions, the size of shape and the number of loops, width and height distribution, the angle of style, shape, average curvature, and gradient. To make comparison, the authors used the Hamming distance.

B. Structural Features

The structural approach is used in extracting the characteristics and features that display the characteristic elements (primitives) and fragments, as well as the relationship between them [5,77,83]. Structural features of images are resistant to various kinds of changes, and they are most preferable in analysing images with great a priori uncertainty.

In [84], images of handwritten texts are segmented into characters, and so-called micro-features are extracted for each of them. The micro-features consist of 512 bits: gradient features (192 bits), structural features (192 bits), and concave surface features (128 bits). Using this set of features, the accuracy of identification by the k nearest neighbours' method has reached 97.71%.

The authors of [85] used four different methods for feature extraction. The first one extracts gradients, structural features, and concave surface features; the second one extracts the features describing the distribution of directions in each segment of the word; the third one extracts the features that contain information about the curvature of the segmented handwritten object; and the fourth one extracts the features that measure the similarity between the contours of objects.

In [86], the writer of a letter was identified on the IAM and RIMES bases using the chain code of the contour of handwritten objects. The recognition accuracy on the first one was 86%, and the same obtained on the other one was 79%. It should be noted that the use of such a feature as a chain code that is easy to extract allowed obtaining good results, therefore, in our opinion, being used in combination with other features, it will significantly increase the quality of recognition.

C. Statistical Features

The statistical approach is based on the stochastic nature of the model used to describe the brightness function in the image [5,6,77,87]. In this approach, the brightness function is considered as the realization of a (stationary) random process (or processes for colour images). In this case, the numerical characteristics of a random process will be the features of the object. This group includes methods of stochastic geometry. Using these methods, a number of geometrical features associated with some random variables are determined.

A review of methods for the extraction of statistical features of handwritten objects is given in [88]. The authors determined that estimating the distribution of the loop contour of handwritten objects provides better recognition results as compared to other statistical features. This estimation made it possible to identify the Firemaker base writers with an accuracy of 63%.

D. Spectral Features

The methods for feature extraction developed as part of the spectral approach are based on the spectral model of image transformation. They differ in the type of function called the kernel of transformation [5,6,77,87]. The kernels of transformation used in the Hadamard, Karhunen-Loève, Radon, Fourier, Haar transforms can be indicated as typical representatives of such functions.

[89] presents a method for spectral feature extraction based on the fast Fourier transform. This method has been tested on 200 Chinese texts by 100 writers. The method showed an accuracy of 98% for the first 10 candidates for the authorship of the letter and 64% for the first candidate using Euclidean and weighted distance classifiers. This scheme is stable, and the possibility to use it for large volumes of data is another advantage thereof, however, it takes large computational effort to extract features using this method.

VI. Recognition

A. Separation Surfaces-Based Algorithms

The earliest pattern recognition algorithms are the ones based on building surfaces that separate classes from each other. In many tasks, including the task of person identification using the handwritten text images, descriptions of objects are specified by sets of values of numerical features (objects can be represented as points in n-dimensional Euclidean space). Such objects (points) can be divided into classes by hypersurfaces of a rather simple form.

The support vector method (SVM) and linear discriminant classifiers (LDC) are the most common algorithms for assessing the similarity of handwritten images from the class of algorithms under consideration. For example, the LDC was successfully applied to classify images of handwritten text in [90], and accuracy of 90% to 100% was achieved in various test bases.

B. Neural Networks

These algorithms are based on the use of various types of artificial neural networks (ANNs) for pattern recognition.

In a multilayer perceptron with backward propagation of error, which is one of the most common architectures, the work of neurons is imitated within a hierarchical network, where each neuron of a higher level is connected by inputs to the outputs of neurons of a downstream level. The values of the input parameters, based on which it is necessary to make certain decisions, predict the development of the situation, etc., are fed to the neurons of the lowest layer. These values are considered as signals transmitted to the next layer, attenuating or amplifying depending on the numerical values (weights) attributed to interneuronal connections. As a result, a certain value, which is considered as a response - the reaction of the entire network to the entered values of the input parameters, is generated at the output of the neuron of the uppermost layer. To apply the network in the future, it should first be trained on previously obtained data, for which both the values of the input parameters and the correct responses to them are known. Learning consists in the selection of weights of interneuronal connections.

To tackle the task of assessing the similarity of handwritten texts, neural networks are often used [91,92]. For example, high identification accuracy (98.9%) was achieved on the ICDAR13 test base using convolutional neural networks [93].

C. Metric Algorithms

Of this class of algorithms, the k nearest neighbour’s algorithm in various modifications was most widely used in tackling the task of identifying a person using text images. In addition, we note the use of the classification algorithm based on the potential method in the development of these systems.

In [94], a weighted Euclidean distance was used to assess the similarity of handwritten Chinese characters of texts, and an accuracy of 95.7% was achieved when checking on the Chinese HBPI base. The authors of [95] argue that when checking the k nearest neighbour’s algorithm using the Hamming distance on the Firemaker handwritten texts base to determine the writer, a better result (97%) was obtained against the use of Chi-square, Minkowski, Euclidean, and Bhattacharyya distances.

The k nearest neighbour’s algorithm has the following advantages:

– it is easy to implement, and introducing various modifications is possible;

- it is possible to interpret the classification of an object by presenting the nearest object or several ones to the user;
- the precedent logic of the work of the algorithm is well understood by experts, etc.

The algorithms have the following disadvantages:

- it is necessary to store the entire learning sample, which leads to inefficient memory consumption and excessive complexity of the decision rule. In the presence of errors (both in the initial data and in the similarity model), this can lead to a decrease in the classification accuracy near the class boundary. It is reasonable to select the minimum subset of reference objects that are really necessary for classification;
- searching for the nearest neighbor requires comparing the classified object with all the objects in the sample for $O(\ell)$ operations. For tasks with large samples, this can be disadvantageous. The problem is solved with the help of efficient nearest neighbor search algorithms, which require an average of $O(ln\ell)$ operations;
- in its original form, the model of the k nearest neighbor's algorithms is extremely poor. It has only a free parameter k, and even that one is discrete with a small number of reasonable alternatives. To enrich the model, it is necessary to enter the weights of objects and/or parameterize the metric calculation method.

D. Recognition Algorithm Based on the Potential Function Method

The idea of this recognition algorithm is quite simple: if the learning object x_i is classified incorrectly, then the y_i class potential is insufficient in the point x_i, and the weight γ_i increases by one.

The advantage of this algorithm lies in the fact that it is very effective when learning objects arrive in a stream, and there is no possibility or need to store them in memory. In those years when the method of potential functions was invented, storing the sample was really a big problem.

The algorithm, however, has many disadvantages:

- the result of training depends on the order of presentation of objects;
- slow convergence;
- the weights γ_i are too roughly tuned;
- the potential centers are too roughly tuned; generally speaking, it is not required that their optimal positions coincide with the learning objects;
- the task of minimizing the number of potentials (nonzero γ_i) is not set at all;
- the h_i parameters (window width) are not tuned at all, and as a result, the quality of classification is relatively slow.

The more modern method of tuning linear combinations of potential functions is based on the EM algorithm [96]. It allows optimizing both the width of each potential and the positions of the centers, and even the number of potentials.

VII. CONCLUSION

An analysis of literature on the theme allows us to draw the following conclusions. Unlike the algorithms of pre-processing of handwritten text images that are not related to the language of the text, the choice of the handwritten text features depends on the specific features of the language in which the text is written. It follows that the development of a system for Uzbek writer identification is a qualitatively new task for a researcher. It is necessary to use the experience of colleagues in the subject area.

To develop the above system, it is necessary to create a database of handwritten text images in the Uzbek language, as well as to develop and programmatically implement the algorithms of the stages of the pre-processing, feature extraction and recognition, taking into account the specific features of the style of the national characters.

REFERENCES

[1] Srihari S. N. Handwritten address interpretation: a task of many pattern recognition problems. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 14(05), 2000, pp. 663-674.

[2] Charfi M., Kherallah M., El Baati A., Alimi A.M. A New Approach for Arabic Handwritten Postal Addresses Recognition.*Int. Journal of Advanced Computer Science and Applications*, Vol. 3, No.3, 2012, pp. 1-7.

[3] Palacios R., Gupta A. A system for processing handwritten bank checks automatically. *Image and Vision Computing*, Vol.26 (10), 2008, pp. 1297-1313.

[4] Ishenko E.P., Toporkov A.A. Forensics. INFRA-M, Moscow, 2010.

[5] Gonzalez R.C., Woods R.E.Digital image processing. Pearson, New Jersey, 2008.

[6] ShapiroL., StockmanG. Computer vision. Pearson, New Jersey, 2001.

[7] Solomevich V.I. Handwriting and character. Harvest, Minsk, 2009.

[8] Patel M., Thakkar S.P. Handwritten Character Recognition in English: A Survey. *Int. Journal of Advanced Research in Computer and Communication Engineering*, Vol. 4, Issue 2, 2015, pp. 345-350.

[9] Plamondon R., Srihari S.N. Online and off-line handwriting recognition: a comprehensive survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, Vol. 22, Issue 1, 2000, pp. 63-84.

[10] Siddiqi I., Djeddi C., Raza A., Souici-meslati L. Automatic analysis of handwriting for gender classification. *Pattern Analysis and Applications*, Vol. 18, Issue 4, 2015, pp. 887-899.

[11] Al Maadeed S., Hassaine A. Automatic prediction of age, gender, and nationality in offline handwriting. *EURASIP Journal on Image and Video Processing*, Vol. 2014, No. 10, pp. 1-10.

[12] Antonacopoulos A., Downton A. Special issue on the analysis of historical documents.*Int. Journal on Document Analysis and Recognition*, Vol. 9, No. 2-4, 2007, pp. 75-77.

[13] Plamondon R., Lorette G. Automatic Signature Verification and Writer Identification – State art.*Pattern Recognition*, Vol. 22(2), 1989, pp. 107-131.

[14] Awaida S.M., Mahmoud S.A. State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic text. *Educational Research and Reviews*, Vol. 7(20), 2012, pp. 445-463.

[15] Schomaker L. Advances in Writer Identification and Verification. *In Proc. of 9th Int'l Conf. on Document Analysis and Recognition*, Vol. 2, pp. 1268-1273.

[16] Chaudhry R., Pant S.K. Identification of authorship using lateral palm print – a new concept. *Forensic Science International*, vol. 141, 2004, 49-57.

[17] Fornes A., Llados J., Sanchez G., Bunke H. Writer Identification in Old Handwritten Music Scores. *In Proc. of 8th IAPR Workshop on Document Analysis Systems*, 2008, pp. 347-353.

[18] Schlapbach A., Marcus L., Bunke H. A writer identification system for on-line whiteboard data. *Pattern Recognition Journal*, 41, 2008, pp. 23821-23897.

[19] Maarse F.J., Schomaker L.R.B., Teulings H.L. Automatic Identification of Writers in Human Computer Interaction: Psychonomic Aspects. Springer, Heidelberg, 1986.

[20] Gupta S. Automatic Person Identification and Verification using Online Handwriting. Master Thesis. International Institute of Information Technology Hyderabad, India, 2008.

[21] LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. *In Proc. of IEEE*, Vol. 86 (11), 1998, pp. 2278–2324.

[22] Marti U., Bunke N. A full English sentence database for off-line handwriting recognition.*In Proc. of the 5th Int. Conf. on Document Analysis and Recognition*, 1999, pp. 705-708.

[23] Grosicki E., Carré M., Brodin J-M., Geoffrois E. Results of the RIMES Evaluation Campaign for Handwritten Mail Processing. *In Proc. of ICDAR 2009*, pp.941-945.

[24] Hull J.J. A database for handwritten text recognition research. *IEEE Trans. Pattern Anal. Mach. Intell.*,16(5), 1994, pp.550-554.

[25] Kleber F., Fiel S., Diem M., Sablatnig R. CVL-database: An off-line database for writer retrieval, writer identification and word spotting. *In Proc. of the 12th Int. Conf. on Document Analysis and Recognition*, 2013, pp. 560-564.

[26] Pechwitz M., Maddouri S.S., Maergner V., Ellouze N., Amiri H. IFN/ENIT-database of handwritten Arabic words. *In Proc. of CIFED*, 2, 2002, pp. 127-136.

[27] Amara N.B., Mazhoud O., Bouzrara N., Ellouze N. ARABASE: a relational database for Arabic OCR systems. *Int. Arab J. Inf. Technol.* 2(4), 2005, pp. 259-266.

[28] Alamri H., Sadri J., Suen C.Y., Nobile N. A novel comprehensive database for Arabic off-line handwriting recognition. *In Proc. of the 11th Int. Conf. on Frontiers in Handwriting Recognition*, 2008, pp. 664-669.

[29] Kharma N., Ahmed M., Ward R. A new comprehensive database of handwritten Arabic words, numbers, and signatures used for OCR testing. *In IEEE Canadian Conf. on Electrical and Computer Engineering*, 2, 1999, pp. 766-768.

[30] Mahmoud S.A., Ahmad I., Al-Khatib W.G., Alshayeb M., Parvez M.T., Märgner V., Fink G.A. KHATT: An open Arabic offline handwritten text database. *Pattern Recognition*, 47(3), 2014, pp. 1096-1112.

[31] Al Maadeed S., Ayouby W., Hassaine A., Aljaam J.M. QUWI: An Arabic and English handwriting dataset for offline writer identification. *In Proc. of Int. Conf. on Frontiers in Handwriting Recognition*, 2012, pp. 746-751.

[32] Kim D., Hwang Y., Park S., Kim E., Paek S., Bang S. Handwritten Korean character image database Pe92.*In Proc. of the 2nd Int. Conf. on Document Analysis and Recognition*, 1993, pp. 470-473.

[33] Zhang H., Guo J., Chen G., Li C. HCL2000 - a large-scale handwritten Chinese character database for handwritten character recognition. *In Proc. of the 10th Int. Conf. on Document Analysis and Recognition*, 2009, pp. 286-290.

[34] Liu C.L., Yin F., Wang D.H., Wang Q.F.CASIA online and offline Chinese handwriting databases. *In Proc. of International Conference on Document Analysis and Recognition*, 2011, pp. 37-41.

[35] Bidgoli A.M., Sarhadi M. IAUT/PHCN: Islamic Azad university of Tehran/Persian handwritten city names, a very large database of handwritten Persian words. *In Proc. of the International Conference on Frontiers in Handwriting Recognition*, 2008, pp. 192-197.

[36] Sagheer M.W., He C., Nobile N., Suen C.Y. A new large Urdu database for off-line handwriting recognition. *Image Analysis and Processing Lecture Notes in Computer Science*, 2009, pp. 538-546.

[37] Jifroodian H.P., Nicola N., He C.L., Suen C.Y. A new large-scale multi-purpose handwritten Farsi database. *Image Analysis and Recognition Lecture Notes in Computer Science*, Vol. 5627, 2009, pp. 278-286.

[38] Safabaksh R., Ghanbarian A.R., Ghiasi G. HaFT: A handwritten Farsi text database. *In Proc. of the 8th Iranian Conference on Machine Vision and Image Processing*, 2013, pp. 89-94.

[39] Ma L., Liu H., Wu J. MRG-OHTC database for online handwritten Tibetan character recognition. *In Proc. of the Int. Conf. on Document Analysis and Recognition*, 2011, pp. 207-211.

[40] Kavallieratou E., Liolios N., Koutsogeorgos E., Fakotakis N., Kokkinakis G. The GRUHD database of Greek unconstrained handwriting. *In Proc. of the 6th Int. Conf. on Document Analysis and Recognition*, 2001, pp. 561-565.

[41] Saady Y.E., Rachidi A., Yassa M. AMHCD: A database for Amazigh handwritten character recognition research. *Int. Journal of Computer Applications*, Vol. 27(4), 2011, pp. 44-48.

[42] Hussain R., Raza A., Siddiqi I., Khurshid K., Djeddi C. A comprehensive survey of handwritten document benchmarks: Structure, Usage and Evaluation. *EURASIP Journal on Image and Video Processing*, Vol. 46, 2015, pp. 1-24.

[43] Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans. Systems Man Cybernet.*, Vol. 9 (1), 1979, pp. 62-66.

[44] Kittler J., Illingworth J. On threshold selection using clustering criteria. *IEEE Trans. Systems Man Cybernet.*, Vol. 15, 1985, pp. 652-655.

[45] Brink A.D. Thresholding of digital images using two-dimensional entropies. *Pattern Recognition*, Vol. 25 (8), 1992, pp. 803-808.

[46] Yan H. Unified formulation of a class of image thresholding techniques. *Pattern Recognition*, Vol. 29 (12), 1996, pp. 2025-2032.

[47] Sahoo P.K., Soltani S., Wong A.K.C. A survey of thresholding technique. *Computer Vision, Graphics Image Processing*, Vol. 41 (2), 1988, pp. 233-260.

[48] Ntirogiannis K., Gatos B., Pratikakis I. A performance evaluation methodology for historical document image binarization. *IEEE Transactions on Image Processing*, Vol. 22, 2013, pp. 595-609.

[49] Yang Y., Yan H. An adaptive logical method for binarization of degraded document images. *Pattern Recognition Letter*, Vol. 33, 2012, pp. 787-807.

[50] Gatos B., Pratikakis I., Perantonis S. Adaptive degraded document image binarization. *Pattern Recognition*, Vol. 39, 2006, pp. 317-327.

[51] Sauvola J., Pietikainen M. Adaptive document image binarization. *Pattern Recognition*, Vol. 33, 2000, pp. 225-236.

[52] Bataineh B., Abdullah S., Omar K. An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows. *Pattern Recognition Letter*, Vol. 32, 2011, pp. 1805-1813.

[53] Mandel I.D. Cluster analysis. Finance and Statistics, Moscow, 1988.

[54] Kaufman L., Rousseeuw P.J. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, New Jersey, 2005.

[55] Li Y., Zheng Y., Doermann D., Jaeger S. Script-independent text line segmentation in freestyle handwritten documents. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol.30 (8), 2008, pp. 1313-1329.

[56] Zahour A., Taconet B., Mercy P., Ramdane S. Arabic hand-written text-line extraction. *In Proc. of ICDAR'01*, 2001, pp. 281-285.

[57] Likforman-Sulem L., Zahour A., Taconet B. Text line segmentation of historical documents: a survey.*Int. Journal of Document Analysis and Recognition*, 9, 2-4, 2007, pp. 123-138.

[58] Li Y., Zheng Y., Doermann D., Jaeger S. A new algorithm for detecting text line in handwritten documents. *Int. Workshop on Frontiers in Handwriting Recognition*, 2006, pp. 35-40.

[59] Shi Z., Govindaraju V. Line Separation for Complex Document Images Using Fuzzy Runlength. *In Proc. of the Int. Workshop on Document Image Analysis for Libraries*, 2004, pp. 306-312.

[60] Louloudis G., Gatos B., Pratikakis I., Halatsis K. A Block-Based Hough Transform Mapping for Text Line Detection in Handwritten Documents. *In Proc. of the Tenth Int. Workshop on Frontiers in Handwriting Recognition*, 2006.

[61] Kim G., Govindaraju V., Srihari S.N. An architecture for handwritten text recognition systems. *Int. Journal on Document Analysis and Recognition*, 1999, 2, 1, pp. 37-44.

[62] Likforman-Sulem L., Faure C. Extracting text lines in handwritten documents by perceptual grouping. *Advances in handwriting and drawing: a multidisciplinary approach*, 1994, pp. 117-135.

[63] Nakajima Y., Mori S., Takegami S., Sato S. Global methods for stroke segmentation. *Int. Journal on Document Analysis and Recognition*, 1999, 2, 1, pp. 19-23.

[64] MestetskiyL.M. Continuous morphology of binary images. Figures. Skeletons. Circulars. FIZMAT-LIT, Moscow, 2009.

[65] Seni G., Cohen E. External word segmentation of off-line handwritten text lines. *Pattern Recognition*, 27(1), 1994, pp. 41-52.

[66] Marti U.V., Bunke H. Text line segmentation and word recognition in a system for general writer independent handwriting recognition. *In Proc. of the 6th Int. Conf. on Document Analysis and Pattern Recognition*, 2001, pp. 159-163.

[67] Kapoor R., Bagai D., Kamal T.S. A new algorithm for skew detection and correction. *Pattern Recognition Letters*, Vol. 25, 2004, pp. 1215-1229.

[68] Kennard D.J., Barrett W.A. Separating Lines of Text in Free-Form Handwritten Historical Documents. *In Proc. of Int. Conf. on Document Image Analysis for Libraries*, 2006, pp. 12-23.

[69] Schlapbach A., Bunke H. Writer Identification Using an HMM-Based Handwriting Recognition System: To Normalize the Input or Not? *In Proc. of the 9th Int. Workshop on Frontiers in Handwriting Recognition*, 2004, pp. 38-142.

[70] Zhang T., Suen C. A fast parallel algorithm for thinning digital patterns.*Communications of the ACM*, Vol.27 (3), 1984, pp. 236-239.

[71] Wu R.-Y., Tsai W.-H. A new one-pass parallel thinning algorithm for binary images. *Pattern Recognition Letters*, Vol.13 (10), 1992, pp. 715-723.

[72] Siddiqi I., Vincent N. Stroke width independent feature for writer identification and handwriting classification. *In Proc. of the Eleventh Int. Conf. on Frontiers in Handwriting Recognition*, 2008, pp. 76-81.

[73] Hamdulla A., Naby G., Ubul K., Moydin K. Research on Uyghur Handwriting Identification Technology Based on Stroke Statistical Features.*Int. Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol.7 (1), 2014, pp. 415-424.

[74] Tou J., GonzalezR. Pattern Recognition Principles. Addison-Wesley, Reading, MA, 1974.

[75] Szeliski R. Computer Vision: Algorithms and Applications. Springer, New York, 2011.

[76] Nixon M., Aguado A. Feature Extraction and Image Processing. Elsevier, New York, 2008.

[77] Soyfer V. et al. Computer image processing methods. Fizmatlit, Moscow, 2003. – 784 c.

[78] Duda R., Hart P., Stork D. Pattern Classification. John Wiley & Sons, New York, 2001.

[79] Pratt W.K. Digital image processing. John Wiley & Sons, New York, 2001.

[80] Marti U.V., Messerli R., Bunke H. Writer identification using text line based features.*In Proc. of 6th Int. Conf.on Document Analysis and Recognition*, 2001, pp. 101-105.

[81] Hertel C., Bunke H. A Set of Novel Features for Writer Identification. *In Proc. of Fourth Int. Conf.on Audio and Video-Based Biometric Person Authentication*, 2003, pp. 679-687.

[82] Leedham G., Chachra S. Writer identification using innovative binarised features of handwritten numerals. *In Proc. of Seventh Int. Conf.on Document Analysis and Recognition*, 2003.

[83] SuyasovD.I. Isolation of structural features of symbol images based on cellular automata with labels. *Information and control systems*, 2010, Vol. 4, pp. 39-45.

[84] Zhang B., Srihari S., Lee S. Individuality of Handwritten Characters. *In Proc. of Seventh Int. Conf. on Document Analysis and Recognition*, 2003, pp. 1086-1090.

[85] Tomai C., Zhang B., Srihari S. Discriminatory power of handwritten words for writer recognition. *In Proc. of Int. Conf. on Pattern Recognition*, Vol. 2, 2004, pp. 638-641.

[86] Siddiqi I., Vincent N. A Set of Chain Code Based Features for Writer Recognition. *In Proc. of Tenth Int. Conf.on Document Analysis and Recognition*, 2009, pp. 981-985.

[87] ShlezingerM., GlavachV. Ten lectures on statistical and structural recognition. Naukovadumka, Kiev, 2004.

[88] Bulacu M., Schomaker L., Vuurpijl L. Writer identification using edge-based directional features. *In Proc. of Seventh Int. Conf.on Document Analysis and Recognition*, 2003, pp. 937-941.

[89] Yan Y., Chen Q., Deng W., Yuan F. Chinese Handwriting Identification Based on Stable Spectral Feature of Texture Images. Int.*Journal of Intelligent Engineering and Systems*, Vol.2, No.1, 2009, pp. 17-22.

[90] Al-Dmour A., Zitar R.A. Arabic writer identification based on hybrid spectral-statistical measures. Journal of Experimental & Theoretical Artificial Intelligence, 2007, 19(4), p. 307–332.

[91] Srihari S., Cha S.-H., Arora H., Lee S. Individuality of Handwriting. Journal of Forensic Sciences, 2002, vol. 47, no. 4, pp. 1-17.

[92] Srihari S., Cha S.-H., Lee S. Establishing Handwriting Individuality Using Pattern Recognition Techniques. *In Proc. of the Sixth Int. Conf. on Document Analysis and Recognition*, 2001, pp.1195-1204.

[93] Christlein V., Bernecker D., Maier A., Angelopoulou E. Offline Writer Identification Using Convolutional Neural Network Activation Features. *In Proc. of the German Conf. on Pattern Recognition*, pp. 540-552, 2015.

[94] Zhu Y., Tan T., Wang Y. Biometric Personal Identification Based on Handwriting. *In: Proc. 15th Int. Conf. on Pattern Recognition, 2000*, Vol. 2, pp. 801–804.

[95] Schomaker L., Bulacu M., Franke K. Automatic Writer Identification Using Fragmented Connected-Component Contours. In *Proc. of 9th Int. Workshop on Frontiers in Handwriting Recognition*, 2004, pp. 185–190.

[96] Schlapbach A. Writer Identification and Verification. Clifton VA, IOS Press, 2007.