ᲡᲐᲥᲐᲠᲗᲕᲔᲚᲝᲡ ᲛᲔᲪᲜᲘᲔᲠᲔᲑᲐᲗᲐ ᲔᲠᲝᲕᲜᲣᲚᲘ ᲐᲙᲐᲓᲔᲛᲘᲘᲡ ᲛᲝᲐᲛᲑᲔ, ტ. 13 №1, 2019
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 13, no. 1, 2019

Informatics

Handwritten Character Recognition Using some (Anti)-Diagonal Structural Features

Jose Manuel Casas*, Nick Inassaridze, Manuel Ladra§, Susana Ladra#**

*Departamento de Matemática Aplicada I, Universidad de Vigo, EUIT Forestal, 36005 Pontevedra, Spain
**A.Razmadze Mathematical Institute, Ivane Javakhishvili Tbilisi State University; Georgian Technical University; Tbilisi Centre for Mathematical Sciences, Tbilisi, Georgia
§Departamento de Matemáticas, Instituto de Matemáticas, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
#Departamento de Computación, Universidad de A Coruña, Campus de Elviña, 15071 A Coruña, Spain*

(Presented by Academy Member Hvedri Inassaridze)

ABSTRACT. In this paper, we present a methodology for off-line handwritten character recognition. The proposed methodology relies on a new feature extraction technique based on structural characteristics, histograms and profiles. As novelty, we propose the extraction of new eight histograms and four profiles from the 32x32 matrices that represent the characters, creating 256-dimension feature vectors. These feature vectors are then employed in a classification step that uses a *k*-means algorithm. We performed experiments using the NIST database to evaluate our proposal. Namely, the recognition system was trained using 1000 samples and 64 classes for each symbol and was tested on 500 samples for each symbol. The accuracy obtained varied from 81.74% to 93.75%, showing better results than other methods from the state of the art also based on structural characteristics. *© 2019 Bull. Georg. Natl. Acad. Sci.*

Key words: OCR, English handwritten character recognition, structural feature, histograms, diagonal profiles

Character recognition, popularly referred as optical character recognition (OCR), has been one of the interesting, fascinating and challenging fields of research in pattern recognition, artificial intelligence and machine vision in the last years [1, 2]. It has multiple applications in the real life, such as verifying signatures, recognizing bank check account numbers and amounts, or automating the mail sorting process in postal services; thus, much research has been focused on designing accurate handwritten character recognition systems [3-5]. As both industry and academy have paid attention to this attractive field, there have been numerous previous attempts for recognizing handwritten characters, such as those methods for English handwritten characters included in [6, 1, 5] or [7]. One can find information about recent trends and tools in OCR in [7].

The generic handwritten recognition process includes several phases: preprocessing the handwritten text, segmenting the writing into isolated characters, extracting feature vectors from the individual characters, and finally classifying each character using the features previously extracted, such that it can be assigned to the most likely letter [4].

In this work, we will focus on the feature extraction step, as it is vital for obtaining good results in the recognition process in terms of accuracy. Its main goal is to extract and select a collection of features that maximizes the recognition rate using the least amount of elements as possible [8]. Moreover, the feature extraction method should be robust, that is, it should obtain similar feature sets from a collection of instances of the same symbol. This property makes the subsequent classification step less difficult [9].

According to [10], we can classify features in different categories: global transformation and series expansion, statistical features, and geometrical and topological features. This last approach allows encoding some knowledge about the structure of the object and the components that make up that object [8]. In addition, it is closer to the human way of recognition [11]. The structural features consider different properties of the characters, such as extreme points, maxima and minima, reference lines, ascenders, descenders, cusps above and below a threshold, isolated dots, cross points, branch points, direction of a strokes, inflection between two points, etc. [8]. Some of these structural features were already proposed by cognitive psychologists [12], when studying the visual and cognitive mechanisms involved in visual object recognition.

There are many works in OCR using structural feature extraction models. [13] proposed a method for the recognition of multi-font printed characters giving special emphasis to structural features. The structural description of the shape for each character considered convex arcs and strokes, singular points and their spatial interrelations. [14] also developed a structural feature set for recognition of printed text. They included different information for a character, such as the location and number of their holes, the concavities in their skeletal structure, characteristics of their bounding boxes, among others. Kuroda, Harada, and Hagiwara implemented a recognition system for the on-line identification of handwritten Chinese characters based on structural patterns [15]. [16] proposed a method for recognizing numeral characters, also based on structural features. Their technique considered topological characteristics like the number of cavities, the crossing sequences, the intersection with the principal and secondary axes, and the distribution of pixels. [17] proposed a structural approach for the analysis of handwritten mathematical expressions. This problem is even more complicated than recognizing individual characters or symbols, as the components of the mathematical expression are normally arranged as a complex two-dimensional structure, and have different sizes. [18] focused on printed Arabic text, which obtains lower recognition rates than those of disconnected characters such as printed English. He used seven types of global structural features such as number of sub words, number of peaks of each word, number of loops of each peak, number and position of complimentary characters, or the height and width of each peak. [19] proposed an integrated analysis system for unconstrained handwriting. The last module of this system includes a handwritten character recognition technique that uses a structural approach [20]. More concretely, it extracts a 280-dimensional feature vector for each character, consisting of the horizontal, vertical and radial histograms and the out-in and in-out radial profiles, and uses the k-means algorithm for the classification. [21] proposed the combination of structural and statistical features, in addition to BP networks for the classification step, to solve interferences of external noise.

In this paper, we propose a new algorithm for isolated English handwritten character recognition based on some structural features, using eight new histograms and four new profiles. Thus, we extract a 256-

dimensional integral vector for each character and then employ the k -means clustering algorithm for the classification step. We compare our results to those given in [20], as the methods for feature extraction and classification are the most similar to ours. Our illustrative tables show that we reduce the dimension of the feature vectors and improve the accuracy of recognition.

The rest of the paper is organized as follows. Section 2 describes the proposed technique in detail. Section 3 includes the experimental evaluation, and finally Section 4 contains the conclusions and future work.

Proposed Algorithm

Preliminaries. A handwritten character recognition system usually requires a preprocessing phase before the feature extraction and classification steps [4]. The main goal of this preprocessing phase is to obtain isolated characters and represent them conveniently for the following steps. In most cases, this includes a segmentation stage and a binarization stage to get the isolated characters in the form of $m \times n$ binary matrices. These matrices are then generally normalized by reducing the size and removing the redundant information from the image without loosing any important information.

Then, the feature extraction is applied over these matrices. This step can be considered as the heart of the system, as the feature selection is usually the most important factor to achieve high accuracies in the recognition process. After the normalization of the character images, the objective of the feature extraction is to represent the isolated characters as unique feature vectors. The key is to maximize the recognition rate using as few features as possible.

Finally, the classification stage is the main decision-making stage of the system, and it uses the feature vectors to identify the text segment according to preset rules. In this stage, the basic task is to design a decision rule that is easy to compute and that maximizes the certainty of the misclassification relative to the power of the feature extraction scheme employed.

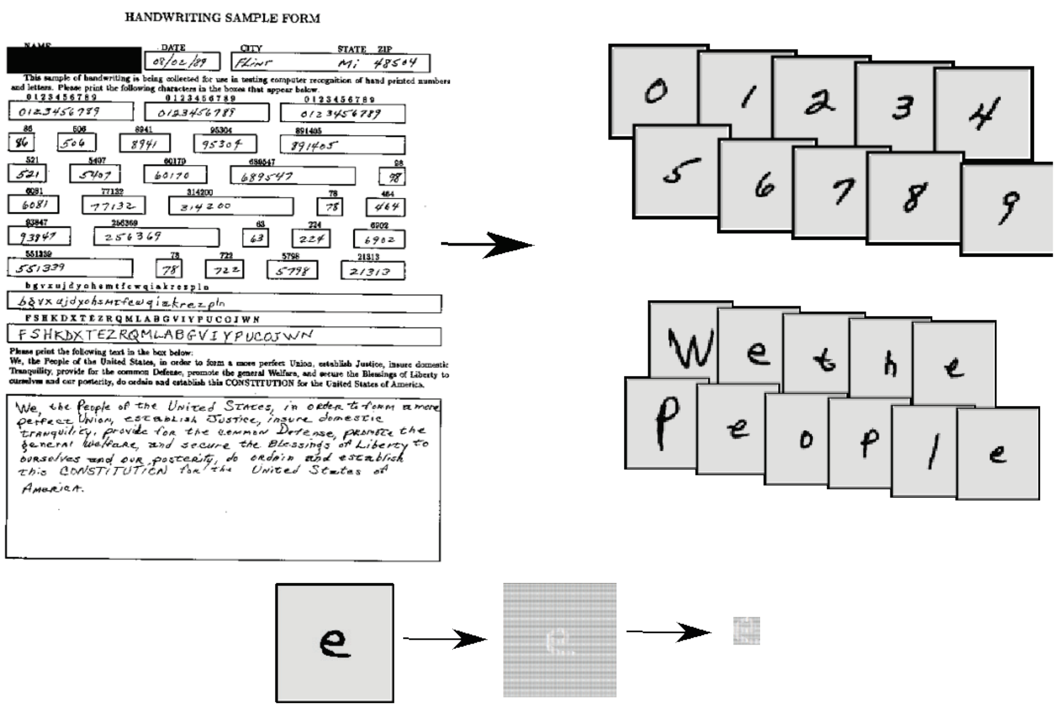

Fig. 1. Preprocessing steps done during the handwritten character recognition process: segmentation phase to obtain isolated characters (top), binarization phase (bottom-left), and normalization (bottom-right).

Preprocessing. For the experimental evaluation, we plan to use NIST database [22], which contains 128×128 BMP files for isolated handwritten English characters. Thus, before extracting the features, the preprocessing step must binarize and then normalize each original image data file to obtain a 32×32 matrix with entries in $\{0,1\}$, such that 0 -s stand for white pixels while 1 -s for black pixels.

We illustrate this preprocessing phase in Fig. 1. Using as example a handwritten sample form from the NIST database, we first illustrate the segmentation step that produces the isolated characters (top part of figure). Then, these isolated characters are further preprocessed using binarization (bottom-left part of the figure) to obtain binary 128×128 matrices, which are finally normalized into binary 32×32 matrices removing unwanted white regions.

Features Extraction. As we have already mentioned, in this paper we focus on structural characteristics for feature extraction. Instead of the well-known horizontal and vertical histograms, we introduce new horizontal left and right histograms and vertical upper and lower histograms. We also employ new orthodiagonal and orthoantidiagonal histograms and profiles. All these features are used for the first time in the optical character recognition research. We will study if these new features improve the accuracy of the handwritten character recognition algorithm in comparison with the algorithm proposed by [20].

Now, we give the formal definition of these features. We need a map $f:[32]\times[32]\rightarrow\{0,1\}$ defined as follows: $f(l,m)$ is the value of the element in the l -th row and m -th column of the character matrix, and $[32]=\{1,\ldots,32\}$.

The horizontal left and right histograms, H_{hl} and H_{hr} , of the character matrix are the number of black pixels in the even rows of the left half of the matrix and the odd rows of the right half of the matrix, respectively (i.e. 32 features):

$$H_{hl}(n)=\sum_{m=1}^{16} f(2n,m) \quad \text{for all } 1\le n\le 16, \text{ and}$$

$$H_{hr}(n)=\sum_{m=16}^{32} f(2n-1,m) \quad \text{for all } 1\le n\le 16.$$

The vertical upper and lower histograms, H_{vu} and H_{vl} , of the character matrix are the number of black pixels in the even columns of the upper half of the matrix and the odd columns of the lower half of the matrix, respectively (i.e. 32 features):

$$H_{vu}(n)=\sum_{m=1}^{16} f(m,2n) \quad \text{for all } 1\le n\le 16, \text{ and}$$

$$H_{vl}(n)=\sum_{m=16}^{32} f(m,2n-1) \quad \text{for all } 1\le n\le 16.$$

Besides above-given histograms, we introduce several other new histograms. We start with the upper and lower diagonal histograms, H_{ud} and H_{ld} , given by the number of black pixels according to the odd and even orthogonal lines to the diagonal of the character matrix in the upper and lower triangles, respectively (i.e. 32 features in total):

$$H_{ud}(n)=\sum_{\substack{k\ge 0\\ 2n-1-k\ge 1\\ 2n-1+k\le 32}} f(2n-1-k,2n-1+k) \quad \text{for all } 1\le n\le 16, \text{ and}$$

$$H_{ld}(n)=\sum_{\substack{k\geq 0\\ 2n-k\geq 1\\ 2n+k\leq 32}} f(2n+k,2n-k) \quad \text{for all } 1\leq n\leq 16.$$

Symmetrically, the upper and lower antidiagonal histograms, H_{uad} and H_{lad}, are defined as the number of black pixels according to the even and odd orthogonal lines to the antidiagonal of character matrix in upper and lower triangles, respectively (i.e. again 32 features in total):

$$H_{uad}(n)=\sum_{\substack{k\geq 0\\ 2n-k\geq 1\\ 33-2n-k\geq 1}} f(2n-k,33-2n-k) \quad \text{for all } 1\leq n\leq 16, \text{ and}$$

$$H_{lad}(n)=\sum_{\substack{k\geq 0\\ 2n+k-1\leq 32\\ 34-2n+k\leq 32}} f(2n-1+k,34-2n+k) \quad \text{for all } 1\leq n\leq 16.$$

Additionally, we introduce the out-in and in-out diagonal and antidiagonal profiles for each normalised character. Namely, the *out-in upper diagonal profile* P_{oiud} and the *out-in lower diagonal profile* P_{oild} are defined at the index $1\leq n\leq 16$ as the position of the first black pixel found in the $(2n-1)$-th orthogonal line to the diagonal of the character matrix, starting from the periphery in the upper triangle going down; and the $2n$-th orthogonal line to the diagonal of the character matrix, starting in lower triangle going up, respectively (i.e. 32 features in total):

$$P_{oiud}(n)=\left\{ I \,\middle|\, \begin{array}{l} \sum_{\substack{k\geq I+1\\ 2n-1-k\geq 1\\ 2n-1+k\leq 32}} f(2n-1-k,2n-1+k)=0 \\ f(2n-1-I,2n-1+I)=1 \end{array}\right. \quad \text{for all } 1\leq n\leq 16, \text{ and}$$

$$P_{oild}(n)=\left\{ J \,\middle|\, \begin{array}{l} \sum_{\substack{k\geq J+1\\ 2n-k\geq 1\\ 2n+k\leq 32}} f(2n+k,2n-k)=0 \\ f(2n+J,2n-J)=1 \end{array}\right. \quad \text{for all } 1\leq n\leq 16.$$

Symmetrically, we introduce the *out-in upper antidiagonal profile* P_{oiuad} and the *out-in lower antidiagonal profile* P_{oilad}, which are defined at the index $1\leq n\leq 16$ as the position of the first black pixel found in the $2n$-th orthogonal line to the antidiagonal of the character matrix, starting from the periphery in upper triangle going down; and the $(2n-1)$-th orthogonal line to the antidiagonal of the character matrix, starting in lower triangle going up, respectively (i.e. 32 features in total):

$$P_{oiuad}(n)=\left\{ I \,\middle|\, \begin{array}{l} \sum_{\substack{k\geq I+1\\ 2n-k\geq 1\\ 33-2n-k\geq 1}} f(2n-k,33-2n-k)=0 \\ f(2n-I,33-2n-I)=1 \end{array}\right. \quad \text{for all } 1\leq n\leq 16, \text{ and}$$

$$P_{oilad}(n)=\left\{ J \,\middle|\, \begin{array}{l} \sum_{\substack{k\geq J+1\\ 2n-1+k\leq 32\\ 34-2n+k\leq 32}} f(2n-1+k,34-2n+k)=0 \\ f(2n-1+J,34-2n+J)=1 \end{array}\right. \quad \text{for all } 1\leq n\leq 16.$$

Moreover, the *in-out upper diagonal profile* P_{ioud} and the *in-out lower diagonal profile* P_{iold} are defined at the index $1\leq n\leq 16$ as the position of the first black pixel found in the $(2n-1)$-th and in the $2n$-th orthogonal lines to the diagonal of character matrix starting from the diagonal going to the periphery in the upper and lower triangles, respectively (i.e. 32 features in total):

$$P_{ioud}(n)=\left\{ I \;\middle|\; \begin{array}{l} \sum\limits_{\substack{k=0\\2n-1-k\geq 1\\2n-1+k\leq 32}}^{I-1} f(2n-1-k,2n-1+k)=0 \\ f(2n-1-I,2n-1+I)=1 \end{array}\right. \quad \text{for all } 1\leq n\leq 16, \text{ and}$$

$$P_{iold}(n)=\left\{ J \;\middle|\; \begin{array}{l} \sum\limits_{\substack{k=0\\2n-k\geq 1\\2n+k\leq 32}}^{J-1} f(2n+k,2n-k)=0 \\ f(2n+J,2n-J)=1 \end{array}\right. \quad \text{for all } 1\leq n\leq 16.$$

Symmetrically, we introduce the *in-out upper antidiagonal profile* P_{iouad} and the *in-out lower antidiagonal profile* P_{iolad}, which are defined at the index $1\leq n\leq 16$ as the position of the first black pixel found in the $2n$-th and in the $(2n-1)$-th orthogonal lines to the antidiagonal of the character matrix, starting from the antidiagonal going to the periphery in upper and lower triangles, respectively (i.e. 32 features in total):

$$P_{iouad}(n)=\left\{ I \;\middle|\; \begin{array}{l} \sum\limits_{\substack{k=0\\2n-k\geq 1\\33-2n-k\geq 1}}^{I-1} f(2n-k,33-2n-k)=0 \\ f(2n-I,33-2n-I)=1 \end{array}\right. \quad \text{for all } 1\leq n\leq 16, \text{ and}$$

$$P_{iolad}(n)=\left\{ J \;\middle|\; \begin{array}{l} \sum\limits_{\substack{k=0\\2n-1+k\leq 32\\34-2n+k\leq 32}}^{J-1} f(2n-1+k,34-2n+k)=0 \\ f(2n-1+J,34-2n+J)=1 \end{array}\right. \quad \text{for all } 1\leq n\leq 16.$$

Classification. In the previous step, a 256-dimensional feature vector have been extracted from each isolated handwritten character image. These feature vectors are then used in the classification step, where we use the k-means clustering algorithm to train and create a classification model.

To better illustrate the idea behind this process, we show in Fig. 2 some examples of 32×32 binary matrix used as input of the feature extraction phase. In Fig. 3, we plot (using gray dashed lines) the values of the 256-dimensional feature vectors for some characters preprocessed from the image of Fig. 1, more concretely, some instances of characters e (top), h (center) and l (bottom). In addition, we also plot the 256-dimensional vector with the mean values for each character, using black solid lines, which correspond to the centroids that represent those handwritten characters in the classification step. As expected, we can see that the values of the 256-dimensional feature vectors are similar when the input matrices correspond to the same handwritten character, but are significantly different to those corresponding to other handwritten characters.

Fig. 2. Examples of some 32×32 binary matrices that are used as input of the feature extraction phase. These examples correspond to characters e (left), h (center), and l (right).

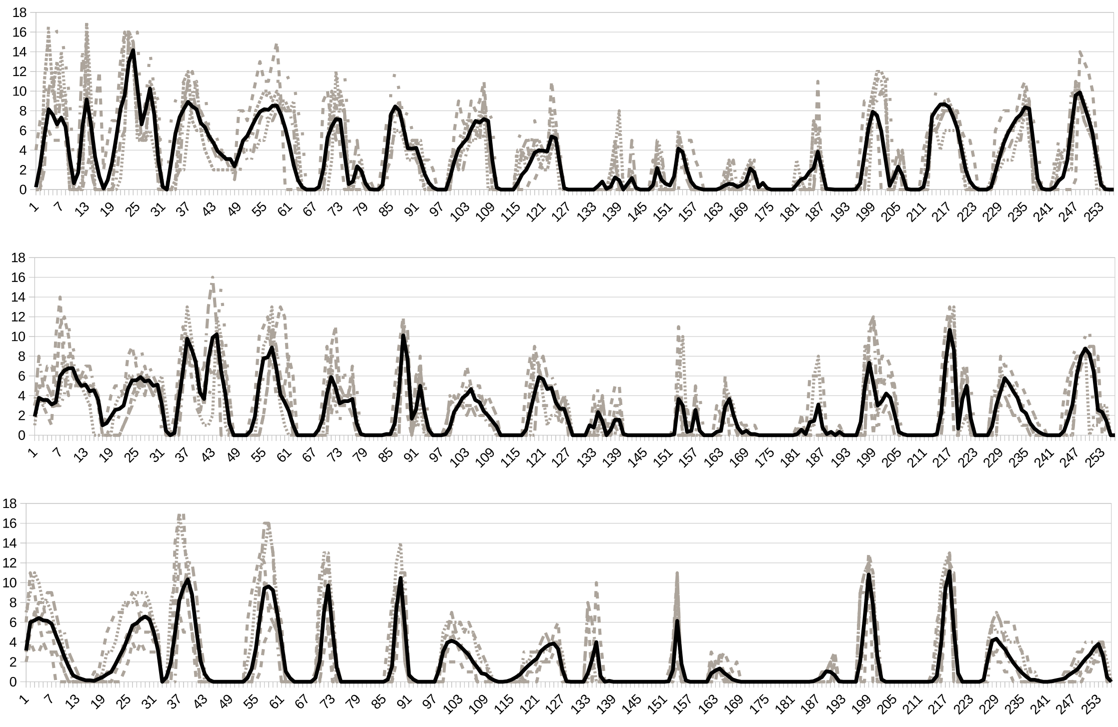

Fig. 3. Values for the 256-dimensional feature vector for different handwritten characters. These plots include some samples of characters e (top), h (center), and l (bottom). Gray dashed lines correspond to values for real handwritten characters, whereas the black solid lines correspond to the 256-dimensional vector with the mean values. These centroids are then used in the classification step.

Conclusion. In this paper, we present a technique for English handwritten character recognition based on the extraction of new structural features. More concretely, we introduce eight new histograms and four new profiles, which have been proven to successfully represent the handwritten characters.

We have tested our approach using the NIST database and obtained recognition accuracies varying from 81.74% to 93.75%, depending on the difficulty of the character category. These results outperform previous attempts of using just structural features, in addition to being fast and simple to compute.

The results are promising and usable is some sort of applications. In the nearest future they will be implemented in a mobile (iOS) application. We also plan to apply the technique to characters from other languages, e.g. Georgian characters and use ANN for the classification stage. Moreover, due to the nature of our method, it is possible to reduce the current number of features, 256 in this paper, according to the needs of the application where the technique will be used.

This work was supported by the Agencia Estatal de Investigación, Spain (European ERDF support included, UE) [grant numbers MTM2016-79661-P, TIN2016-77158-C4-3-R]; and by the Consellería de Cultura, Educación e Ordenación Universitaria and the European Regional Development Fund (ERDF) [grant number ED431G/01]. The second author was partially supported by a joint project of Shota Rustaveli National Science Foundation and Science & Technology Center in Ukraine, Project N6321 [contract № STCU-2016-08].

ინფორმატიკა

ხელნაწერი სიმბოლოების ამოცნობა ზოგიერთი (ანტი)-დიაგონალური სტრუქტურული თვისების გამოყენებით

ხ. მ. კასასი*, ნ. ინასარიძე, მ. ლადრა§, ს. ლადრა#**

**ვიგოს უნივერსიტეტი, გამოყენებითი მათემატიკის დეპარტამენტი, პონტევედრა, ესპანეთი*
***ივანე ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტი, ა.რაზმაძის მათემატიკის ინსტიტუტი; საქართველოს ტექნიკური უნივერსიტეტი; თბილისის მათემატიკის მეცნიერების ცენტრი, თბილისი, საქართველო*
§სანტიაგო დე კომპოსტელას უნივერსიტეტი, მათემატიკის ინსტიტუტი, მათემატიკის დეპარტამენტი, სანტიაგო დე კომპოსტელა, ესპანეთი
#ა კორუნას უნივერსიტეტი, გამოთვლითი დეპარტამენტი, ა კორუნა, ესპანეთი

(წარმოდგენილია აკადემიის წევრის ხ. ინასარიძის მიერ)

მოცემულ სტატიაში წარმოდგენილია ხელნაწერი სიმბოლოების ამოცნობის ოფლაინ მეთოდოლოგია. შემოთავაზებული მეთოდოლოგია ეყრდნობა თვისების მოპოვების ახალ ტექნიკას, დაფუძნებულს სტრუქტურულ მახასიათებლებზე, ჰისტოგრამებსა და პროფილებზე. სიახლეა ჩვენ მიერ შეთავაზებული ახალი რვა ჰისტოგრამისა და ოთხი პროფილის მოპოვება სიმბოლოების წარმომდგენი 32×32 მატრიცებისგან, რომლებიც ქმნიან 256-განზომილებიან თვისებების ვექტორებს. აღნიშნული ვექტორები შემდეგ ექვემდებარება კლასიფიკაციის ეტაპს, რომელიც იყენებს k-საშუალოს ალგორითმს. ჩვენ შეთავაზების შესამოწმებლად ჩავატარეთ ექსპერიმენტები **NIST** მონაცემთა ბაზის გამოყენებით. კერძოდ, ამოცნობის სისტემა იქნა დატრენინგებული თითოეული სიმბოლოსათვის 1000 ნიმუშისა და 64 კლასის გამოყენებით, ხოლო ტესტირებული თითოეული სიმბოლოსათვის 500 ნიმუშზე. მიღებულია ამოცნობის სიზუსტე $81,74$-დან $93,75\%$-მდე, რაც წარმოადგენს უკეთეს შედეგს, დღეისთვის არსებულ სტრუქტურულ მახასიათებლებზე დაფუძნებულ სხვა მეთოდებთან შედარებით.

REFERENCES

1. Mori S., Suen C. and Yamamoto K. (1992) Historical review of OCR research and development. *Proceedings of the IEEE* **80**(7): 1029–1058.
2. Impedovo S., Ottaviano L. and Occhinegro S. (1991) Optical character recognition - a survey. *International Journal Pattern Recognition and Artificial Intelligence* **5**(1): 1–24.
3. Bahlmann C., Haasdonk B. and Burkhardt H. (2002) Online handwriting recognition with support vector machine a kernel approach. *Proceeding of the 8th International Workshop on Frontiers in Handwriting Recognition* (IWFHR), pp. 49–54.
4. Beigi H. (1993) An overview of handwriting recognition. *Proceedings of the 1st Annual Conference on Technological Advancements in Developing Countries*, pp. 30–46. Columbia University, New York.
5. Nadal C., Legault R. and Suen C. (1990) Complementary algorithms for recognition of totally unconstrained handwritten numerals. *Proceedings of the 10th International Conference on Pattern Recognition* (ICPR), pp. 434–449.
6. Trier O., Jain A. and Taxt T. (1996) Feature extraction methods for character recognition – a survey. *Pattern Recognition* **29**(4): 641–66.
7. Sharma O. P., Ghose M., Shah K. B., Thakur B. K. (2013) Recent trends and tools for feature extraction in OCR technology. *International Journal of Soft Computing and Engineering* **2**(6): 220–223.
8. Arica N. and Yarman-Vural F. (2001) An overview of character recognition focused on off-line handwriting, IEEE Transactions on Systems, Man, and Cybernetics, Part C: *Applications and Reviews* **31**(2): 216–233.
9. Fitzgerald J., Geiselbrechtinger F., Kechadi T. (2004) Application of fuzzy logic to online recognition of handwritten symbols. *Proceedings of the 9th International Workshop on Frontiers in Handwritten Recognition* (IWFHR), pp. 395–400.
10. Govindan V. and Shivaprasad A. (1990) Character recognition - a review. *Pattern Recognition* **23**(7): 671–683.
11. Bunke H. and Sanfeliu A. (1990) Syntactic and structural Pattern Recognition – Theory and Applications, World Scientific, Singapore.
12. Lindsay P. and Norman D. (1977) Human information processing: An introduction to psychology, Academic Press, New York.
13. Rocha J. and Pavlidis T. (1994) A shape analysis model with applications to a character recognition system. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(4): 393–404.
14. Kahan S., Pavlidis T. and Baird H. (1987) On the recognition of printed characters of any font and size. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**(2): 274–288.
15. Kuroda K., Harada K. and Hagiwara M. (1997) Large scale on-line handwritten Chinese character recognition using improved syntactic pattern recognition. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (ICSMC), pp. 4530–4535.
16. Lee L. and Gomes N. (1997) Disconnected handwritten numeral image recognition. *Proceedings of 4th International Conference on Document Analysis and Recognition* (ICDAR), pp. 467–470.
17. Chan K.-F. and Yeung D.-Y. (2000) An efficient syntactic approach to structural analysis of on-line handwritten mathematical expressions. *Pattern Recognition,* **33**(3): 375–384.
18. Amin A. (2000) Recognition of printed Arabic text based on global features and decision tree learning techniques. *Pattern Recognition,* **33**(8): 1309–1323.
19. Kavallieratou E., Fakotakis N. and Kokkinakis G. (2002b) An unconstrained handwriting recognition system. *International Journal on Document Analysis and Recognition,* **4**(4): 226–242.
20. Kavallieratou E., Fakotakis N. and Kokkinakis G. (2002a) Handwritten character recognition based on structural characteristics. *Proceedings of the 16th International Conference of Pattern Recognition* (ICPR), pp. 139–142.
21. Yang Y., Lijia X. and Chen C. (2011) English character recognition based on feature combination. Proc. Engineering 24: 159–164.
22. Wilkinson R., Geist J., Janet S., Grother P., Burges C., Creecy R., et al. (1992) The first census optical character recognition systems conference, Technical Report NISTIR 4912, The U.S Bureau of Census and the National Institute of Standards and Technology, Gaithersburg, MD.

Received October, 2018