UDC 004.032.26.93'12 (045)
DOI: 10.18372/2306-1472.74.12286

Volodymyr Kharchenko[1]
Iurii Chyrka[2]

TYPED DIGITS RECOGNITION USING SEQUENTIAL PROBABILITY RATIO TEST

National Aviation University
1, Kosmonavta Komarova Avenue, Kyiv, 03058, Ukraine
E-mails: [1]kharch@nau.edu.ua; [2]yurasyk88@ukr.net

Abstract

Purpose: *the represented research results are aimed to better understanding of computer vision methods and their capabilities. Both the statistical classifier and an artificial neural network allows processing of typical objects with simple descriptors.* ***Methods:*** *considered methods are grounded at probabilistic theory, optimization theory, kernel density estimation and computer-based simulation as a verification tool.* ***Results:*** *the considered artificial neural network architecture for digits recognition has advantage in comparison with statistical method due to its better classification ability. Presented results of experimental verification prove that advantage in both single observation and sequential observation scenarios.* ***Discussion:*** *the approach can be implemented in a variety of computer vision systems that observe typed text in difficult noisy conditions.*

Keywords: Bayesian classifier; digits recognition; neural network; sequential test.

1. Introduction

Character recognition (generally, pattern recognition) is a part to the problem of classifying input data into some categories. [1]. One of its subdomain is optical character recognition (OCR). It has been an active research field for a long time, because this technology is widely applied in many areas where is necessity to process text information produced as or transformed to a graphic image. For example it is used for books and documents digitization, automatic sorting of postal letters, the car license plate recognition, etc. The need for character recognition has increased much since the expansion of the Internet. OCR is used for recognizing both printed and handwritten letters and many researches have been conducted to reach higher precision and make recognition faster. [2]

The accurate recognition printed text with Latin letters is considered nowadays mostly as a solved problem when clear imaging is available, for example, with printed documents [3]. Usual accuracy on these scenarios exceeds 99%.

The character recognition software breaks the image into sub-images, each containing a single character. The sub-images are then transformed to a special format of data that is fed to the classifier that has been trained to make the association between the character image data and some numeric value that corresponds to the character. The output from the classifier is then translated into ASCII text and saved as a file. [1]

Characters recognition is a quite complex problem. The characters could have different font, size, orientation, thickness, format or introduction of different types of noises [4]. This gives us infinite variations. In these cases, it is extremely difficult to identify similar characters because each variation has unique shape. Therefore, it needs a robust OCR system that is able to recognize characters in various conditions. The most of existing OCR methods are not able to recognize characters with different fonts rather than in the training samples and need to adapt.

2. Analysis of the research and publications

There are some common methods to recognize printed and handwritten characters. In [5] the next list of typical text classifiers is presented.

The most basic way to recognizing the patterns using statistical methods in which we use Bayesian classifiers for characters recognition. Statistical methods include estimation of the distributions of single character feature vectors, and then apply a classifier or some detector to discover presence of some character. In this way, the invariance to some unavoidable conditions is achieved. Specifically to statistical methods, the supervised classification assumes the knowledge of a probability density function (PDF) for each class of objects. Of course, in most real situations those PDFs are unknown and have to be estimated from a set of training samples with a correct label of each class.

Linear classifier is one of the simplest existing classifiers. It adds each input pixel values with some weights for each known character. The output with the highest sum indicates the input character.

Another simple method is a K-nearest neighbor classifier with a Euclidean distance measure between input images. This classifier has such advantages as no need of training and specific participation of the designer. However, it requires huge memory size and big computational costs.

The linear classificator with many inputs can be simplified using principal component analysis. It requires the preprocessing stage, which computes the projection of the input pattern on the row of principal components of the set of training vectors. To compute the principal components, the mean of each input component was first computed and subtracted from the training vectors. The covariance matrix of the resulting vectors was then computed, and diagonalized using Singular Value Decomposition. The resulted multidimensional feature vector is used as the input of a second degree polynomial classifier.

Artificial Neural Networks (ANN) belongs to a class of simple classifier combinations. They combine multiple linear classifiers into the multilayer net and bring additional classification possibilities with introduction of nonlinear activation function for each classifier.

In order to find a compromise between small networks that cannot learn the training set, and large networks that can be overloaded with lots of parameters, there were introduced specialized network architectures that are designed to recognize two-dimensional shapes such as characters, while eliminating irrelevant distortions and variability. Such ANN are called as convolutional networks (CNN). Complete networks are formed of multiple convolutional layers, extracting features of increasing complexity and abstraction. During the training process, a convolutional network automatically synthesizes those features by itself.

3. Aim of the paper

In this paper, first we analyze performance of the presented earlier object recognition method with Bayesian classifier in the task of printed digits recognition. Additionally we compare its performance with the 2-layer fully connected ANN.

4. Feature extraction

Similarly to principal component analysis, the main idea of feature extraction is reduction of dimensionality and introduction of some robustness to observation conditions. The feature vector for the character is extracted on the preprocessing stage by some predefined algorithm and then is used for classification by a preferable method.

Extracting features is the key process and has direct influence on the final recognition performance. Therefore, the extracted feature must describe each character accurately. Among more complicated feature descriptors, there are ‘classic’ ones such as SIFT, SURF, and HOG. They have been known as the most popular and successful feature descriptors, which can be applied in many applications, from image registration to object recognition [6], [9]. Regarding CNN, as it was mentioned before, they synthesize their own features based on input images in the training set. However, in many situations they are too excessive and can be replaced by a set of simple statistical parameters. In this paper, we consider such parameters as the eigenvalues of the covariance matrix and its first eigenvalue in particular, that describe the character’s shape. In addition, such statistical values by each axis as a skewness and an excess kurtosis are defined in order to describe shapes of brightness distribution. They form our five-component feature vector in the next order the first eigenvalue, the skewness by x, the kurtosis by x, the skewness by y, the kurtosis by y.

5. Classification algorithms

We compare two classification methods: Bayesian classifier and ANN in two scenarios: single image recognition and consequent detection and recognition by the set of images with the same character.

Firstly, let us briefly describe each of methods. We assume that N classes of objects are given. N=10 in the case of single digit recognition. They correspond to prior probabilities $p_1,\dots,p_N$, with

$$p_i > 0,\ i = \overline{1,N},\ \sum_{i=1}^{N} p_i = 1$$

Bayesian classifier builds the separation plane based on maximum a posteriori probability. It requires knowledge of prior probabilities p_i, of each i-th class and likelihood functions i.e. a probability distribution function (PDF) $\rho_i(\mathbf{x})$ of feature vector $\mathbf{x} = (x_1,\ldots,x_D) \in \Re^D$ for the i-th class.

Posterior probabilities can be evaluated as follows:

$$q_k = \frac{p_k \rho_k(\mathbf{x})}{\sum_{i=1}^{N} p_i \rho_i(\mathbf{x})},$$

with $k = \overline{1, N}$. In a case with equal prior probabilities $p_i = 1/N$, $\forall i = \overline{1, N}$ this equation is simplified to $q_k = \rho_k(\mathbf{x})$. In a single observation scenario the answer is chosen as

$$\hat{k} = \arg\max_k q_k .$$

We do not know PDFs of each class and consider their nonparametric kernel density estimations [7], [8]. For this purpose, we use the Epanechnikov kernel that is preferable over the Gaussian one due to absence of exponent and possible extremely large or small numbers.

The kernel density estimator for any particular class is evaluated by observations of the training set of T objects and corresponding set of feature vectors $\mathbf{X}_1,\ldots,\mathbf{X}_T$ for each class as

$$\hat{f}(\mathbf{x}) = \frac{1}{Th_1 \ldots h_D} \sum_{i=1}^{T} \prod_{j=1}^{D} K\left(\frac{x_j - X_{ij}}{h_j}\right),$$

where; h_j, $j = \overline{1, D}$, is a bandwidth for jth parameter of the feature vector, $K(\cdot)$ is the Epanechnikov kernel function.

For classification with ANN there was constructed the network with fully connected two layers of 11 neurons (10 for digits plus one for a background). For the first layer there were used the sigmoid activation function and the second layer has a simple linear activation function. The network structure is shown in the Fig. 1.

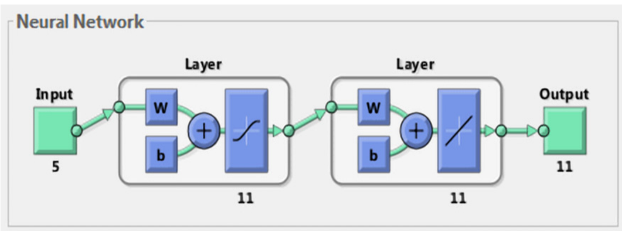

Fig. 1. The artificial neural network structure

Preliminary research has shown that a single layer of 11 neurons with the sigmoid activation function is not enough for correct classification. As the input, we use the mentioned feature vector with 5 elements and output we interpret as a set of probabilities of each class of digits.

6. Consequent object detection and recognition using Wald criterion

During the consequent analysis of the image set with the same character, its presence should be firstly detected and then classified.

We check the null hypothesis H_0: a painted object belonging to one of the abovementioned classes is observed, vs. an alternative hypothesis H_1: a background or other object is observed. Consider a test statistic on the n-th observation

$$T_n = \prod_{k=1}^{n} \frac{\bar{\rho}(\mathbf{x}_k)}{\rho_0(\mathbf{x}_k)} .$$

Here

$$\bar{\rho}(\mathbf{x}) = \sum_{i=1}^{N} p_i \rho_i(\mathbf{x})$$

denotes a weighted PDF of the feature vectors with digits. We suppose that α and β are given levels for a Type I error (the probability to reject H_0 under true H_0) and a Type II error (the probability to accept H_0 under true H_1).

We choose thresholds as

$$A = \frac{1-\beta}{\alpha}, \; B = \frac{\beta}{1-\alpha}.$$

We continue observations until one of the inequalities holds true:

$$T_n \geq A \text{ or } T_n \leq B.$$

In the first case, H_0 is accepted (a digit is detected), and in the second case, H_0 is rejected (a digit is not detected).

Suppose that an object has been detected. We continue the observations and receive new data

$$\mathbf{x}_n,\ldots,\mathbf{x}_m, \quad m \geq n.$$

Here we use the observation $\mathbf{x}_n$ which was received at the moment of the detection algorithm termination, and new observations as well.

Posterior probabilities can be evaluated as follows:

$$q_k^{(m)} = \frac{p_k \rho_k(\mathbf{x}_n) \rho_k(\mathbf{x}_{n+1}) \ldots \rho_k(\mathbf{x}_m)}{\sum_{i=1}^{N} p_i \rho_i(\mathbf{x}_n) \rho_i(\mathbf{x}_{n+1}) \ldots \rho_i(\mathbf{x}_m)},$$

with $k = \overline{1, N}$.

The observations can be terminated if the following condition holds true:

$$\max_{1 \leq k \leq N} q_k^{(m)} \geq P_C .$$

Here P_C is a probability of correct recognition. After the termination, the answer is chosen similarly to a single observation as

$$\hat{k} = \arg\max_k q_k .$$

7. Simulation results

The performance of the considered classifiers was tested on the set of digits 0-9 that are shown in Fig. 2. For test purposes, they were typed using the Arial font separately and then imported to MATLAB where they were transformed accordingly to specified parameters and polluted by simulated noise. Hereafter they were processed as two-dimensional arrays. We considered digits with the single font but polluted with randomized Gaussian noise.

All digits have the same extents, placed in the center of the observation window.

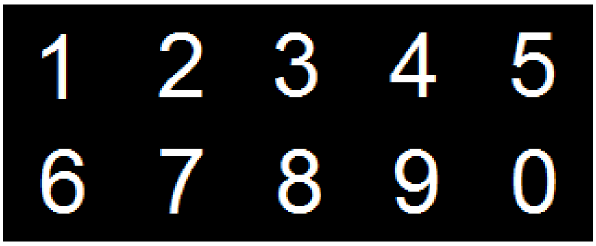

Fig. 2. The set of simulated objects shapes

Firstly, let us consider recognition performance of algorithms in a single shot case without multiple observations. The most representative parameters in this case is probability of correct classification. Then the consequent set of data will be supplied to each classifier and compare by the number of required observation that is needed to reach predefined detection probability and recognition levels.

The simulations were carried out via Monte-Carlo method with the next general parameters and conditions: number of independent iterations 1,000; there were simulated grayscale images with size 64x64 pixels with the bit depth 8bit per pixel; ten classes of digits have been randomly placed in the center of each image with equal probabilities, signal-to-noise ratio SNR=10dB, training sample size for each digit N=100, digit brightness is set to 127 with the maximum brightness value 255, digit to background brightness ratio is 1/4. In the sequential processing the maximal number of available observations for detection and recognition is limited by the training sample size, probabilities of false alarm and missing the character are α, β =0.01; classification threshold $P_c = 0.99$.

The network was trained using the MATLAB Neural Networks Toolbox with next typical parameters: optimization algorithm is error back propagation; quadratic cost function; number of epochs 1000; gradient threshold value 1e-7.

Fig. 3 shows plots of probabilities of correct recognition in a single run scenario. Plots behavior is quite predictable and both methods improve their performance when SNR grows. It shows that the both classifiers works well in a wide range of SNR. That can be explained by the fact that the distance between classes in the space of features is quite big. Still, the good recognition probability 0.9 is reached at about 10dB. It must be noted that ANN has well visible advantage in recognition probability, particularly with low SNR.

Also there has been simulated the case for dependencies by the number of used features. Corresponding plots are shown in Fig. 4. It is well visible that performance has significant dependence on number of used features and reach acceptable levels with five proposed features. For statistical criterion the maximum is reached for 4 features. That can indicate that for five and more features in these conditions the model starts to overlearn.

Additionally it has been estimated the complication of classification of each digit. Fig. 5 represents corresponding probabilities of recognition. One can see that the hardest digits for recognition are “6”, “9”, “0” and “8”. Generally, both methods have similar performance level.

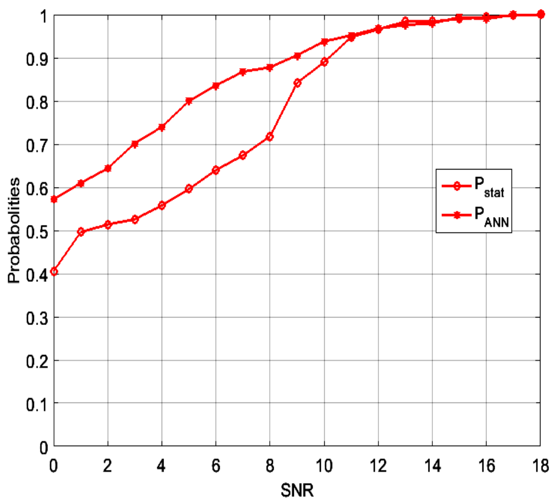

Fig. 3. Correct classification probability at various SNR

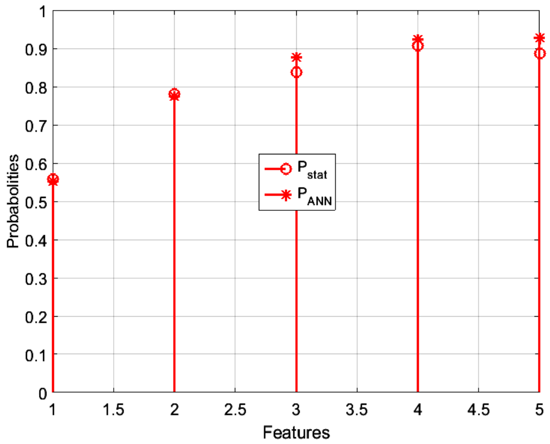

Fig. 4. Correct classification probability at various features numbers

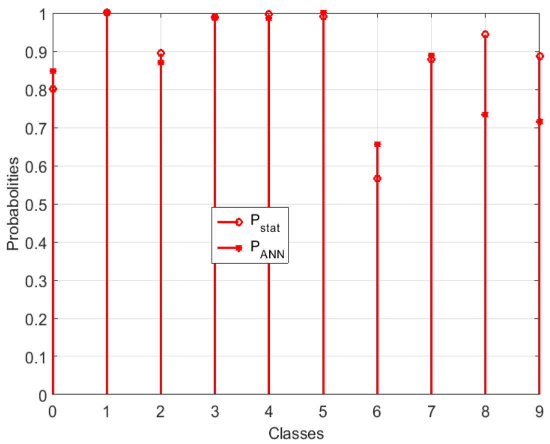

Fig. 5. Correct classification probability for each digit

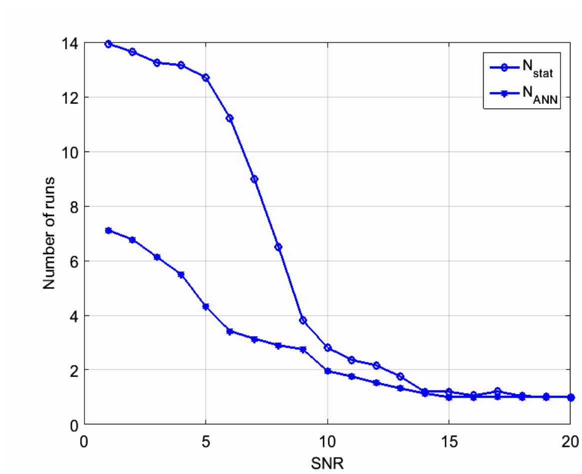

Fig. 6. Average number of observations at various SNR

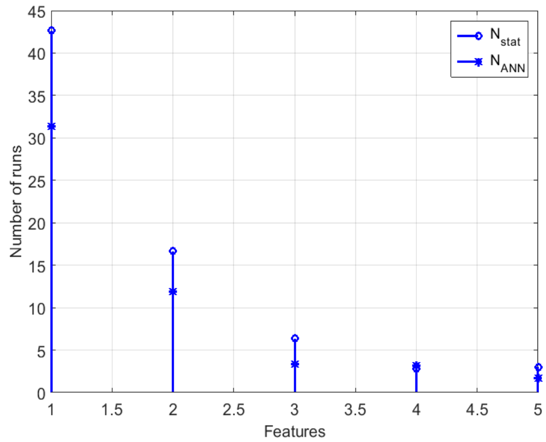

Fig. 7. Average number of observations at various features numbers

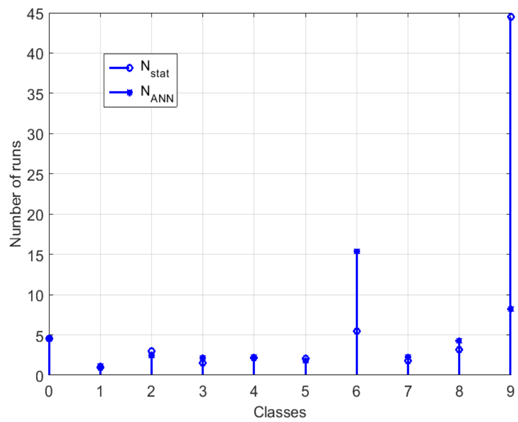

Fig. 8. Average number of observations required for classification of each class of objects

Now let us consider the scenario with multiple observations of the same object and performance of mentioned classifiers with a sequential probability ratio test. Fig. 6 shows plots of average number of observation at which classification have been done for various signal to noise ratios.

Results for detection are not presented because they have shown that the detector works well in a wide range of SNR and even in severe noise conditions a single observation was usually enough for detection. That can be explained by the fact that the distance between features classes with background only and background plus digit is big enough for brief detection of a digit presence. On the other hand, the recognition procedure significantly drops at 5dB that correlates with results for a single observation scenario.

Also there has been simulated the case for dependencies by the number of used criterions. Corresponding plots are shown in Fig. 7. It is well visible that sufficient performance is achieved when

4 features are used for the classification. These results are similar to the Fig. 4 but demonstrate advantage of ANN classifier over the statistical one.

Additionally it has been estimated the complication of classification of each digit. Fig. 8 represents average number of observations for recognition of each class of objects

It shows that such digits like “6” and “9” are hard for recognition by the statistical approach when the considered ANN has only slight worsen of performance on these digits. For better handling of this situation, some additional information or slightly modified criterions must be used.

8. Conclusions

The considered Bayesian approach as well as proposed architecture of the Artificial Neural Network are suitable for detection and recognition of typed digits after the preprocessing stage that extracts some statistical features. Presented simulation results prove that the considered methods can be applied for detection and classification of digits with various shape.

References

[1] Vasudeva N., Parashar H. J., Vijendra S. (2012) Offline Character Recognition System Using Artificial Neural Network. *International Journal of Machine Learning and Computing*, No. 4(2), pp. 449–452.

[2] Samadiani N., and Hassanpour H. (2015) A neural network-based approach for recognizing multi-font printed English characters. *Journal of Electrical Systems and Information Technology*, No. 2, pp. 207–215.

[3] Dong Xiao Ni Seidenberg, “Application of Neural Networks to Character Recognition”, CSIS, Pace University, School of CSIS, Pace University, White Plains, NY, 2007

[4] Coates A., Carpenter B., Case C., Satheesh S., Suresh B., Wang T., Wu D.J., Ng A.Y. (2011) Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning. Proc. 2011 Int. Conf. on Document Analysis and Recognition (ICDAR). Beijing, China, doi: 10.1109/ICDAR.2011.95

[5] LeCun Y., Jackel L., Bottou L., Brunot A., Cortes C., Denker J., Drucker H., Guyon I., Müller U., Säckinger E., Simard P., Vapnik V. (1995) Comparison of learning algorithms for handwritten digits recognition. Proc. 1995 Int. Conf. on Artificial Neural Networks. Paris, France, pp. 53-60.

[6] Pham C.H., Yaguchi Y., and Naruse, K. (2016) Feature Descriptors: a Review of Multiple Cues Approaches. Proc. 2016 IEEE Int. Conf. on Computer and Information Technology. Fiji, pp. 310-315. doi: 10.1109/CIT.2016.61

[7] Wasserman L. (2006). *All of Nonparametric Statistics*. Springer Texts in Statistics. New York, NY, Springer, 268 p.

[8] Kharchenko V.P., Kukush A.G., Kuzmenko N.S., and Ostroumov I.V. (2017) Probabilistic Approach to Object Detection and Recognition for Videostream Processing. *Proceedings of the NAU*, No. 2(71), pp. 8–14. doi: 10.18372/2306-1472.71.11741

[9] Kharchenko V., Pawęska M., Bugayko D., Antonova A., Grigorak M. (2017) Theoretical Approaches for Safety Levels Measurements– Sequential Probability Ratio Test (SPRT). *Logistics and Transport,* no. 2(34), p.p. 25-32.

Received 01 December 2017

В.П. Харченко[1], Ю.Д. Чирка[2]

Розпізнавання друкованих цифр з використанням послідовного ймовірнісного критерія

Національній авіаційний університет, просп. Космонавта Комарова, 1, Київ, 03058, Україна
E-mails: [1]kharch@nau.edu.ua; [2]yurasyk88@ukr.net

Мета: представлені результати досліджень спрямовані на краще розуміння методів комп’ютерного зору та їхніх можливостей. Статистичний класифікатор та штучна нейронна мережа дозволяють обробляти типові об’єкти з простими дескрипторами. **Методи дослідження:** розглянуті методи базуються на теорії ймовірності, теорії оптимізації, ядерній оцінці щільності ймовірності та комп'ютерному моделюванні як засобі апробації. **Результати:** розглянута архітектура штучної нейронної мережі має перевагу у порівнянні зі статистичним методом завдяки кращій здатності до класифікації. Представлені результати експериментальної перевірки доводять цю перевагу і у випадку одиничного спостереження, і при послідовному сценарії. **Обговорення:** підхід може бути реалізований у багатьох системах комп’ютерного зору, що оглядають друкований текст в складних шумових умовах.

Ключові слова: Байєсівський класифікатор; нейрона мережа; послідовний критерій; розпізнавання цифр.

В.П. Харченко[1], Ю.Д. Чирка[2]

Распознавание печатных цифр с помощью последовательного вероятностного критерия

Национальный авиационный университет, просп. Космонавта Комарова, 1, Киев, 03058, Украина
E-mails: [1]kharch@nau.edu.ua; [2]yurasyk88@ukr.net

Цель: представленные результаты исследований направлены на лучшее понимание методов компьютерного зрения и их возможностей. Статистический классификатор и искусственная нейронная сеть позволяют обрабатывать типовые объекты с простыми дескрипторами. **Методы исследования:** рассматриваемые методы базируются на теории вероятности, теории оптимизации, ядерном оценивании плотности вероятности и компьютерном моделировании как средстве апробации. **Результаты:** рассматриваемая архитектура искусственной нейронной сети имеет преимущество по сравнению с статистическим методом благодаря лучшей способности к классификации. Представленные результаты экспериментальной проверки подтверждают это преимущество как в случае одиночного наблюдения, так и при последовательном сценарии. **Обсуждение:** подход может быть реализован во многих системах компьютерного зрения, которые наблюдают цифры в сложных шумовых условиях.

Ключевые слова: Байесовский классификатор; нейронная сеть; последовательный критерий; распознавание цифр.

Kharchenko Volodymyr. Doctor of Engineering. Professor.
Vice-Rector on Scientific Work, National Aviation University, Kyiv, Ukraine.
Editor-in-Chief of the scientific journal Proceedings of the National Aviation University.
Winner of the State Prize of Ukraine in Science and Technology, Honored Worker of Science and Technology of Ukraine.
Education: Kyiv Institute of Civil Aviation Engineers, Kyiv, Ukraine.
Research area: management of complex socio-technical systems, air navigation systems and automatic decision-making systems aimed at avoidance conflict situations, space information technology design, air navigation services in Ukraine provided by CNS/ATM systems.
Publications: 540.
E-mail: kharch@nau.edu.ua

Chyrka Iurii (1988). Candidate of Engineering Sciences. Senior Researcher.
National Aviation University.
Education: National Aviation University, Kyiv, Ukraine (2011).
Research area: control systems, radar signals processing, acoustic holography, and applied statistics.
Publications: 45.
E-mail: yurasyk88@ukr.net