FACE RECOGNITION BASED ON THE PROXIMITY MEASURE CLUSTERING

V.B. Nemirovskiy[1], A.K. Stoyanov[1], D.S. Goremykina[1]
[1]Institute of Cybernetics of Tomsk Polytechnic University, Tomsk, Russia

Abstract

In this paper problems of featureless face recognition are considered. The recognition is based on clustering the proximity measures between the distributions of brightness clusters cardinality for segmented images. As a proximity measure three types of distances are used in this work: the Euclidean, cosine and Kullback-Leibler distances. Image segmentation and proximity measure clustering are carried out by means of a software model of the recurrent neural network. Results of the experimental studies of the proposed approach are presented.

Keywords: featureless comparison, clustering, one-dimensional mapping, neuron, Kullback-Leibler distance, image.

Introduction

At the present time content-based image retrieval is very relevant. It is used in Internet search engines, in the systems of technical vision and biometric identification, in digital image libraries, in archives, databases, etc. [1 – 3].

The range of tasks to be solved in such a search, include the face recognition. At the moment, the face recognition has been widely discussed, but in total, the problem is still far from being resolved.

In a wide variety of algorithms designed for this task we can distinguish three groups.

In the first from them recognition is performed by comparing of the characteristic facial features [4]. The general structure of the algorithms used to this include two stages.

At the first stage the detection and localization of the face in the image are being produced. The process of the second stage includes face alignment (geometric and brightness), feature extraction and actually recognition - features matching with the etalons from the data base.

Task feature extraction is laborious and requires a significant investment of time and computational resources. The metric algorithms used in this group for comparison the image are clearly presented as a feature vector.

The second group includes the neural network algorithms [5 – 7]. Neural Networks (NN) are being trained on a set of training examples. In a process of NN training an automatic extraction of key features occurs, as well as the definition of their importance and building relationships between them.

Convolutional NN produce the best results in the face recognition [7]. They provide partial resistance to changes in scale, to shifts, to rotations, to foreshortening change, and to others distortions.

Neural network algorithms have a major disadvantage: the addition of a new reference face into database requires a complete retraining of network on the all existing set [8]. Besides, there are problems associated with training, as well as difficulties associated with the choice of the number of neurons, layers, etc.

In the third group of algorithms features are not used, which significantly reduces the complexity of recognition.

The approach based on this principle is called featureless recognition. General statement of the featureless recognition problem is formulated in [9, 10]. Solution of this problem is based on the hypothesis of compactness. This hypothesis is based on the assumption that objects with similar properties more often are in one class than in different classes [11].

According to this statement it is supposed that each object of recognition can be presented by results of paired comparisons with basis objects. For comparison any arbitrary real-valued function called by distance can be used. This function not necessarily has to be a metric. Further, in space of distances any problem of featureless recognition is actually reduced to a problem of metric classification. The decision on belonging of an image to this or that class in classical statement usually made on the basis of excess of the threshold established by the chosen decisive rule.

In other words, the solution is as follows. In some set of distances, the subset of the distances is singled out. In this subset distances correspond to images most similar to the sample.

It is easy to see that such a statement of recognition problem is equivalent to solving the problem for clustering of distances. Indeed, let we have a set of distances from the sample object to another objects. Obviously, these distances will differ by its values. These values can be grouped. The smallest distances that close to zero, will define a group of objects the most similar to the sample. The other groups will include the distances corresponding to different similarity degrees of the presented object with a sample.

The described procedure, in fact, is the clustering of distances. According to the hypothesis of compactness, all distances which are a part of one cluster will correspond to similar images. Then the problem of face recognition according to the presented sample will be resolved.

In this paper we consider the possibility of application the clustering for solution of the featureless face recognition problem.

1. Use of clustering for featureless image recognition

The featureless approach was used in [12 – 14] for the recognition of near-duplicates of the presented image. In fact, near-duplicate recognition is an integral part of the

face recognition. As a distance function in these studies the cosine distance was used. With it the proximity of distributions for the cardinalities of brightness clusters was evaluated.

Obtaining a decision rule for featureless recognition in case of the use Euclidean metric as well as Manhattan metric as a function for pairwise comparison of the objects is shown in [10].

To implement the featureless images recognition, it is necessary to answer a few questions. The first of them - what characteristic of the image need to select for comparison. The next question – what function need to select for the pairwise comparison of images.

Often an image is described by a vector in a multidimensional space of the values of considered characteristic, for example the brightness values of its pixels. In this case, for analysis of brightness characteristics often the histograms are used. Comparison of presented this way images more often are carried out by means of Euclidean metric. Use of the vector of all brightness values in the image description is redundant. Really, in any image there exist pixels with similar brightness values. These values may be represented in the histogram by the several adjacent columns. Such columns may be combined almost without loss of informativeness. This consolidation can significantly reduce the dimension of the image description. But combining of the histogram's columns essentially means the original image is needed to segment by brightness. In other words, featureless comparison of images easier to perform for the images segmented by brightness.

Any histogram can be interpreted as an estimate of a probability density of random value by its empirical values. For an image the pixel brightness is such random value.

A similar estimate of the probability can be obtained for the segmented image. For simplicity, we consider the images in a gray scale.

Let us suppose, for simplicity of reasoning, that image segmentation is performed by a brightness pixels clustering. If N_i is a number of brightness values of pixels for i-th cluster and N is a total number of pixels in the image, then the relative frequency of brightness of i-th cluster $p_i = N_i/N$ can be considered as an estimate of probability density. Indeed, for any i value $p_i < 1$ and $\sum p_i = 1$, i.e., the frequency p_i can be considered as the estimated probability of presence the pixel brightness value in the i-th cluster.

Thus any image is associated with the probability distribution of the pixel brightness in the clusters. In [12], the relative frequency p_i is called the relative cardinality of the cluster. This term we will use further.

In this case, the task of two images comparison is reduced to a comparison of two probability distributions. Quantitative estimation of such a comparison is accepted to perform by using the Kullback-Leibler distance. From the information point of view, the Kullback-Leibler distance is a measure of the loss of information about the reference distribution $\Phi(x)$ if to submit it by the distribution $G(x)$. Thus, this measure allows to estimate the difference in the information of any two distributions $\Phi(x)$ and $G(x)$. That is, the distribution of clusters cardinality provides the ability to compare the fields of measurement.

The Kullback-Leibler distance d between the two images can be calculated by the formula

$$d = \sum_i p_i \cdot \ln(p_i/q_i), \qquad (1)$$

where p_i is cardinality of the i-th cluster of the reference image and q_i is cardinality of the same cluster for comparing image.

2. Clustering of the Kullback-Laibler distance

Different values of the distance d determine a degree of proximity to the sample image. A zero value of d corresponds to an exact copy of presented image. The greater the distance d, the greater the difference between images. In other words, it means either finding of an exact copy of the sample, or finding image similar to sample. In the second case it may be near-duplicates of sample conditioned by change in the imaging conditions, such as illumination, shooting angle or zooming the image when copying.

For featureless recognition we used method of clustering that does not require a priori knowledge of the number of clusters. This property has the data clustering with help of recurrent neural network that was considered in [12 – 16]. This clustering method allows you simple enough to select the subsets (the clusters) of data having similar properties from data set.

Let us consider the model of functioning of a single neuron in the specified recurrent neural network [15 – 16] to find out which property allows you to combine data into one cluster.

Operating of the neuron with activation function $f(x)$ is modeled by the one-dimensional mapping on it of input signal value x. In our case, such function is a sigmoid. For sigmoid the mapping $x_{n+1} = f(x_n)$ is the contraction mapping (here n is the current iteration number). Mapping is the contraction mapping if there exists a constant $K < 1$ such that for any two points x and y the inequality

$$\rho(f(x), f(y)) \leq K\rho(x, y), \qquad (2)$$

where ρ is distance between points x and y.

Observance of this inequality leads to the fact that due to mapping the any value reaches a stable fixed point x^* with a given accuracy for a certain number of iterations. A fixed point is the point for which have the equality $x^* = f(x^*)$.

Accuracy of approximation to the x^* in a result of n iterations is determined by the relation [17]:

$$\rho(f^n(x), x) \leq K^n d / (1 - K), \qquad (3)$$

where $d = \rho(x, f(x))$

From (3) it is easy to get:

$$\frac{\rho(f^n(x), x^*)}{\rho(x, f(x))} \leq \frac{K^n}{(1 - K)}, \qquad (4)$$

where $f^n(x)$ is a value of mapping on its n-th iteration, $n = 0, 1, 2, \ldots$.

By (4) it follows that from the set of input signal values **x** we can select the subsets of values that satisfy (4) for a given **n** (i.e. the clusters of values). United in one cluster the values have a common property – an equal number of iterations required to achieve the stable fixed point. Should be noted that not for every *n* in the input signal there are the values **x** satisfying of (4). The clusters for such *n* we call empty clusters.

Thus, the process of mapping, which implemented by the neuron allows to select the clusters on the set of input signal values. As such, the set we may be considered set of distances between the sample and the other images, which compare with the sample.

If we will include in clustering process in addition the distance sample – sample, then all distances in the same cluster with sample will be correspond to images closest to sample. Clusters neighboring with this cluster will be correspond to images having more differences with the sample. At the same time although such images are more different from sample, they are similar to each other.

3. Experimental results

To check presented above the reasoning, we carried out a number of experiments. Experiments were conducted by the following procedure.

Two samples from several face images were formed. The first of these samples consisted of 45 images (sets of 9 images for five persons). Each set includes the original face image (let us call it further by its number, e.g. image 1) taken from a database of images provided by Yandex (see Fig. 1), and the various distortions of the original image obtained in Adobe Photoshop.
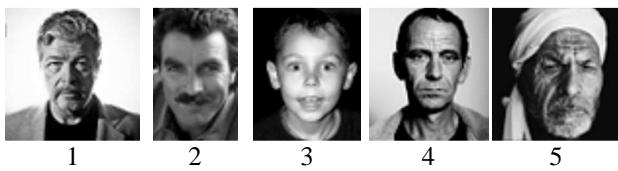

Fig. 1. The original images of recognizable faces

Thus, in each set there were used:
1 – an original image;
2 – Gaussian blur by radius of 2 pixels;
3 – Gaussian blur by radius of 4 pixels;
4 – Gaussian blur by radius 6 pixels;
5 – a spot noise;
6 – a mirror reflection;
7 – ripples;
8 – a reduced image;
9 – wind.

The second of these samples included a set of images from the collection [18]. The collection includes 20 face images of 375 people in different lighting conditions. Part of the images has a significant deviation from frontal view and expresses different emotions. Total in our experiments from this database has been used 40 images of two people (the files from 9338462.1.jpg to 9338462.20.jpg and the files from hensm.1.jpg to hensm.20.jpg of collection [18]).

For each image the distribution of relative cardinality of brightness clusters was formed. Clustering of brightness

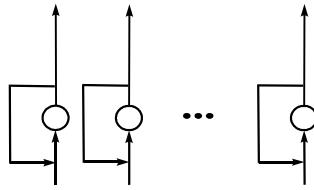was performed by the recurrent neural network (Fig. 2) with the parameters calculated in accordance to [13 – 16].

Fig. 2. The structure of the neural network

To ensure equal conditions of brightness clustering for all images, in our case, unlike [13 – 16], one-step clustering without optimization of the parameter μ in the expression of neuron activation function was used:

$$f(x) = \mu / [1 + \exp(-\alpha x + \beta)],$$

where α is a coefficient of inclination and β is the amount of displacement.

The first series of experiments was implemented by the following algorithm:
• a sample image is chosen alternately from the existing originals;
• the distance from the sample to each of all other images is calculated;
• clustering of a set of obtained distances is performed (including zero distance).

For clustering distances, the same neural network was used, but with the optimization of μ. Evaluation of the images proximity to the sample was carried out using three distances: Euclidean distance, cosine distance and Kullback-Leibler distance. The results of this series of experiments are given in Tables 1 and 2.

Table 1. The distribution of correctly recognized faces by the clusters

Clusters	Distance		
	Euclidean	Cosine	Kullback-Leibler
1-2	30	38	42
1-3	43	43	43

By (4) it is follows that clustering may reveal several clusters for specific values of the number of iterations n. Some of them will be filled by the distance values; part of them will be not filled. Table 2 shows the content of only a few filled clusters. This content is sufficient to perform the analysis.

The content of the vast majority of unspecified in Table 2 clusters includes the distances to the images, which do not similar to the sample and therefore are not of interest for further discussion.

In Table 1 for all five sets of collection (45 images total) the distribution by clusters of number of all correctly recognized images (to be more exact corresponding to them distances) is shown. Results is given for the first two clusters and for the first three clusters for each metric. We can see that the first two clusters contain 93 % of all correctly recognized faces (42 face images) only for the Kullback-Leibler distance. When using the first three clusters this percentage increases to 96 % for all metrics. Therefore, for analyzing of clustering results it is sufficient to consider the content of the first three clusters.

The cells content of Table 2 shows number of distances fallen into specified cluster for each metrics. Image number of original from collection of Fig. 1, to which distance relates, is indicated in brackets. For example, notation 9(4) mean 9 distances corresponding to images from the set of 4-th original (see Fig. 1). Several distances corresponding to different originals are separated by semicolon.

Table 2. Images in obtained clusters

Sample – image 1 from Fig. 1			
Clusters	Distance		
	Euclidean	Cosine	Kullback-Leibler
1	**1 (1)**	**3(1)**	**6(1)**
2	**5(1)**	**6(1)**	**3(1); 3(4)**
3	3(1)	9(4)	6(4)
Sample – image 2 from Fig. 1			
Clusters	Distance		
	Euclidean	Cosine	Kullback-Leibler
1	**1(2)**	**3(2)**	**5(2)**
2	**3(2)**	**2(2)**	**4(2)**
3	5(2)	4(2)	8(1); 9(3); 1(4); 9(5)
Sample – image 3 from Fig. 1			
Clusters	Distance		
	Euclidean	Cosine	Kullback-Leibler
1	**1(3)**	**1(3)**	**2(3)**
2	**6(3)**	**5(3)**	**6(3)**
3	1(3)	2 (3)	1 (3); 6 (4); 6(5)
Sample – image 4 from Fig. 1			
Clusters	Distance		
	Euclidean	Cosine	Kullback-Leibler
1	**1 (4)**	**1 (4)**	**1 (4)**
2	**8 (4)**	**8 (4)**	**7 (4)**
3	1(3)	1(3)	9(1); 9(3); 6(5)
Sample – image 5 from Fig. 1			
Clusters	Distance		
	Euclidean	Cosine	Kullback-Leibler
1	**2 (5)**	**2 (5)**	**1 (5)**
2	**2 (5)**	**2 (5)**	**7 (5)**
3	4 (5)	4 (5)	1 (5); 1(3)

Table 3 shows the distribution by clusters of the average image recognition error depending on the applied metric. It is seen that when using two clusters for recognition it is more profitable to calculate the Kullback-Leibler distance. In this case, we obtain average errors 1 and 2 types no more 0.07. When using three clusters it is more profitable to use the Euclidean distance which gives average errors 1 and 2 types no more 0.04.

Table 3. Average error of image recognition

Clusters	Distance		
	Euclidean	Cosine	Kullback-Leibler
The average error of the first type			
1-2	0.33	0.27	0.07
1-3	0.04	0.04	0.03
The average error of the second type			
1-2	0	0	0.05
1-3	0.02	0.12	0.53

Figure 3 shows incorrectly recognized images when the first original of Fig. 1 was taken as sample. From the collection of Fig. 1 it is not difficult to discover that the first and the fourth faces are markedly different from the other faces.

At that, they have similar visual brightness distribution. Perhaps therefore the faces of Fig. 3 were attributed to the first original. Here, as in other cases of incorrect recognition, we meet with the semantic contradiction inherent to the present method. The contradiction is that each class used in face recognition, includes images of faces of the same person. Our method compares the brightness distributions, which are not sensitive to the semantics of the images. However, just this feature of the method allows detecting near-duplicates of one face.
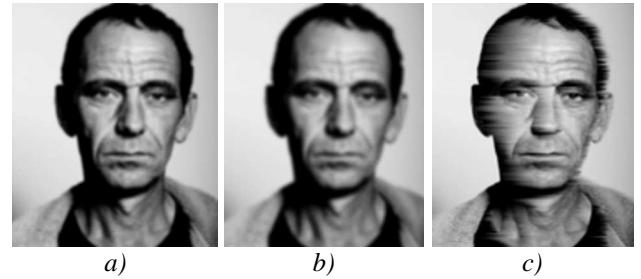

Fig. 3. Incorrectly recognized image: a) radius of the Gaussian blur is 4 pixels; b) radius of the Gaussian blur is 6 pixels; c) wind distortion

In the next experiment, the Gaussian blur of face 1 (Fig. 1) by radius of 6 pixels was taken as a sample. Proximity of images to sample was determined by the Kullback-Leibler distance. The set of obtained distances (including zero distance the sample-sample) have been clustered as described earlier. By a result of the clustering of these distances, in the first cluster was only zero distance. The remaining 8 distances for face images were included in the second cluster. This is consistent with the above conclusions. It is noteworthy that in this, as in the previous experiments, mirror reflection of faces as well as reduced faces were recognized unmistakably. All of them, depending on the used metric, fall into the first two or three filled cluster.

The next series of experiments was carried out with the second set of images taken from [18]. As previously noted, many images of this collection were obtained with deviations from the frontal view and express various kinds of emotions. Such images are often difficult to recognize by the presented sample. In Fig. 4 some of the images that illustrate this feature of the collection are shown.

In these experiments the image shown in Fig. 4.1 was used as a sample. To determine the proximity of images Kullback-Leibler distance was calculated. The set of the obtained distances (including zero distance) was clustered using the same neural network as before.
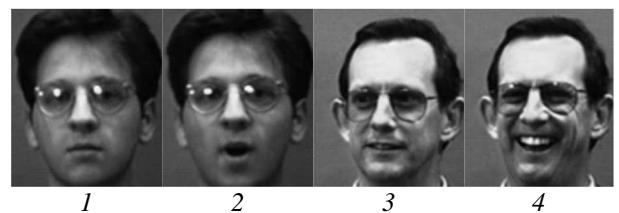

Fig. 4. Faces from collection [18]

Fig. 5 shows a diagram of distribution of the images by clusters as a result of the distances clustering. In the legend of Fig. 5 we use image file names from collection [18].
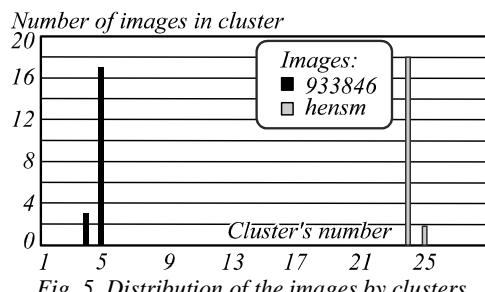
Number of images in cluster

Fig. 5. Distribution of the images by clusters

Previously, when discussing the inequality (4), it has been marked that the region of the input signal values **x** in the mapping process is divided into a number of sub-bands (i.e. clusters). Fig. 5 shows that the all of the images were distributed to 4 clusters. Besides the first two filled clusters are located close one to the other and contain all twenty distances corresponding to images belonging to a class of the sample. The remaining twenty distances are remoted from the first group by a large number of empty clusters. At the same time, they are also located in two neighboring clusters and correspond to the images of another face (of another class).

In this experiment, as in the previous ones, besides of images recognition by the presented sample, the effect of partial ordering of the unrecognized images into classes is observed. It consists in the fact that some clusters may include distances of a greater part or all of images of one or more classes. Table 4 shows the distribution of classes of the unrecognized images for clusters 4 – 5 when using the different metrics.

Table 4. Images in obtained clusters

Sample – image 1		
Clusters	**Distance**	
	Euclidean	Cosine	Kullback-Leibler
4	2(4)	1(3); 2(5)	9(3); 8(5)
5	5(4)	8(3)	8(5) ; 9(4)
Sample – image 2		
4	1(5)	1(3); 1(5)	1(1); 8(4)
5	1(4)	8(1); 8(3); 8(5); 9(4)	--
Sample – image 3		
4	1(3)	1(3); 1(5)	3(1); 2(4)
5	1(4)	1(4)	6(1)
Sample – image 4		
4	1(1); 8(3)	9(1); 7(3)	1(4)
5	8(1); 1(3); 4(5)	1(3); 1(5)	1(5)
Sample – image 5		
4	1(3)	1(3)	8(3)
5	1(5); 5(3); 1(4)	1(5); 1(3); 1(4)	1(4)

It is seen that this effect is mostly expressed when using the Kullback-Leibler distance.

Conclusion

1. Clustering of proximity measures of segmented images to the sample allows to implement featureless face recognition.
2. To analyze featureless face recognition, it is sufficient to use two first clusters in the case of the Kullback-Leibler measure and three first clusters in the case of the Euclidean measure.

3. Using clustering of the proximity measure to the sample for face recognition allows to provide in the experiments the average errors 1 and 2 types not above 0.07.
4. Clustering of the proximity measure to the sample causes the effect of the partial ordering of the unrecognized images by the classes. This can be used for a preliminary selection of the groups of the similar images.

References

[1] Wang JZ, Li J, Wiederhold G. SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001; 23(9): 947-963.

[2] Melnichenco A. Methods of search images by visual similarity and detection fuzzy duplicate images [In Russian]. Russian seminar on Information Retrieval Evaluation. ROMIP Proceedings 2009: 108-121.

[3] Pimenov V. Simple methods for content-based image retrieval [In Russian]. Russian seminar on Information Retrieval Evaluation. ROMIP Proceedings 2010: 69-79.

[4] Kuharev GA. Biometric systems: Methods and tools for the identification of the human person [In Russian]. Saint-Petersburg: Politechnika; 2001.

[5] Hemant SM, Harpreet K. Face Recognition using PCA & neural network. International Journal of Emerging Science and Engineering (IJESE) 2013; 1(6): 71-75.

[6] Varlamov A, Sharapov R. Machine Learning of Visually Similar Images Search. CEUR Workshop Proceedings 2012; 934: 113-120.

[7] Taigman Y, Yang M, Ranzato M, Wolf L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2014: 1701-1708. DOI: 10.1109/CVPR.2014.220.

[8] Ibragimov VV, Arsentjev DA. Algorithms and methods for person recognition in the conditions of modern information technologies [In Russian]. Bulletin MGUP 2015; 1: 67-69.

[9] Seredin OS. Linear methods of pattern recognition on the sets of the objects of arbitrary nature represented by pairwise comparisons. The general case [In Russian]. Izvestija TulGU: Natural Sciences 2012; 1: 141-152.

[10] Seredin OS, Mottl VV, Tatarchuk AI, Razin NA Convex selective criteria of relevant vectors method in the spaceof objects pairwise comparisons [In Russian]. Izvestija Tul-GU: Natural sciences 2013; 1: 165-176.

[11] Vorontsov KV. Lectures on the metric classification algorithms [In Russian]. Moscow: MFTI; 2007.

[12] Nemirovskiy VB, Stoyanov AK. Near-duplicate image recognition based on the rank distribution of the brightness clusters cardinality. Computer Optics 2014; 38(4): 811-817.

[13] Nemirovskiy VB, Stoyanov AK. Near-duplicate image recognition. Mechanical Engineering, Automation and Control Systems (MEACS). Proceedings of the International Conference, Tomsk, 2014, Institute of Electrical and Electronics Engineers (IEEE) 2014; 4. DOI: 10.1109/MEACS.2014.6986916.

[14] Nemirovskiy VB, Stoyanov AK. Application of recurrent neural network for multi-step full-color image segmentation. The 8th International Forum on Strategic Technology (IFOST-2013), 2013, Ulaanbaatar, Mongolia. Mongolian University of Science and Technology 2013; 2: 221-224, DOI: 10.1109/IFOST.2013.6616891.

[15] Nemirovskiy VB, Stoyanov AK. Multi-step segmentation of images by means of a recurrent neural network. The 7th International Forum on Strategic Technology (IFOST-

2012), September 18-21, 2012, Tomsk: [proceedings], National Research Tomsk Polytechnic University (TPU), [S. l.]: IEEE, 2012, 4, DOI: 10.1109/IFOST.2012.6357619.

[16] Nemirovskiy VB, Stoyanov AK. Image segmentation by recurrent neural network. [In Russian]. Bulletin of the Tomsk Polytechnic University 2012; 321(5): 205-210.

[17] Danilov VI. Lectures on fixed points [In Russian, Electonic Preprint]. Moscow: Russian Economic School; 2006. Source: ⟨https://www.nes.ru/dataupload/files/programs/econ/pre-prints/2006/danilov_fixed_points.pdf⟩.

[18] Collection of Facial Images. Source: ⟨http://cswww.es-sex.ac.uk/mv/allfaces/index.html/⟩.

Authors' information

Victor Borisovich Nemirovskiy (b. 1945) Candidate of Physical and Mathematical Sciences, associate professor, Department of Informatics and System Design of National Research Tomsk Polytechnic University. Research interests are neural networks, image processing, image recognition. E-mail: *vbn@tpu.ru* .

Alexander Kirillovich Stoyanov (b. 1946) Candidate of Technical Sciences, associate professor, Department of Informatics and System Design of National Research Tomsk Polytechnic University. Research interests are neural networks, image processing, image recognition. E-mail: *sto@tpu.ru* .

Darya Sergeevna Goremykina (b. 1994) postgraduate student of Department of Informatics and System Design of National Research Tomsk Polytechnic University. Research interests are image processing, image recognition. E-mail: *goremykina2008@mail.ru* .