Method of Parallel-Hierarchical Network Self-Training and its Application for Pattern Classification and Recognition

Leonid TIMCHENKO, Nataliya KOKRIATSKAIA, Victor MELNIKOV, Roman MAKARENKO, Nikolay PETROVSKYI
Department of Telecommunication Technologies and Automatics, State University for Transport Economy and Technologies, 19 Lukashevich Street, Kyiv, 03049, Ukraine
timchen@list.ru, kokriatskaia@rambler.ru, mscon@yandex.ru, mrv_detut@bigmir.net, pettrovski@voliacable.com

***Abstract*—Propositions necessary for development of parallel-hierarchical (PH) network training methods are discussed in this article. Unlike already known structures of the artificial neural network, where non-normalized (absolute) similarity criteria are used for comparison, the suggested structure uses a normalized criterion. Based on the analysis of training rules, a conclusion is made that application of two training methods with a teacher is optimal for PH network training: error correction-based training and memory-based training. Mathematical models of training and a combined method of PH network training for recognition of static and dynamic patterns are developed.**

***Index Terms*—parallel-hierarchical network, training, population coding, preparation, face recognition.**

I. Introduction

An algorithm of back propagation of error assumes the presence of a certain external link supplying the neural network, in addition to input patterns, with targeted output patterns as well. Algorithms based on such concept are called training algorithms with a teacher. For their successful functioning, experts are needed that would specify a reference output pattern for each input pattern at the preliminary stage. As the creation of intelligent systems is based, to a great extent, on biological prototypes, up till now, the disputes do not end on whether to consider training algorithms with a teacher natural or completely artificial.

For instance, training of the human brain, in appearance, takes place without a teacher: information arrives to visual, acoustic, tactile and other receptors from without, and certain self-organization takes place in nervous systems. However, one can hardly deny the fact that people have many teachers in their life (in literal and in metaphorical sense), who coordinate external impacts. At the same time, whatever the result of the dispute between adherents those concepts of training is, they both have a right to existence.

In their classical work McCulloch and Pitts [1] describe the logic of computations in neural networks, based on the mathematical logic and on results obtained in neurophysiology, when a formalized model of the neuron corresponds to the principle "all or nothing". The network with such paradigm is composed of a great number of elementary computing units connected by the correctly configured and operating synchronically synaptic weight, and is able to perform any computations. This result became a real breakthrough in the sphere of nervous system modeling.

One of the modern directions in this area are parallel-hierarchical (PH) networks [2-5] that possess good computational performance due to the homogeneous organization of the parallel distributed structure, as well as due to ability to learn and generalize. Generalization means elaboration of a substantiated result on the basis of input data, which was not formed at the training stage. These data processing properties enable PH networks to solve complex problems, for which purpose they should be integrated into more complex systems.

II. Premises for Development of PH Network Training Methods

The most important property for any neural network is its ability to be trained, receiving data from the environment, and to improve its efficiency as a result of the training process . The neural network is trained by means of an interactive process of synaptic weights and thresholds adjustment.

Ideally, the neural networks obtain knowledge about the environment after each iteration of the training process. We may define training as a process, where free parameters of the network are adapting through a process of stimulation by the environment, to which the network itself is incorporated. A type of the training is determined by the application of a training method with varying parameters. This definition of the training process implies the following sequence of events:

1. The neural network is stimulated by the environment.
2. Free parameters of the neural network are changed as result of the stimulation.
3. The neural network responds to the stimuli of the environment in a new manner due to changes that occurred in the inner structure of the network.

A predetermined set of strict rules of training problem solving is called a training algorithm. As is it known, the universal training algorithm does not exist, since all training algorithms differ from each other by an adjustment method of synaptic weights of neurons. Another specific distinction

Digital Object Identifier 10.4316/AECE.2012.04006

of the neural networks created from interconnected neurons is a mode of network connection with the environment. In this context we may speak about an environment model-related training paradigm.

As is it known, there are three training paradigms: training with a teacher, without a teacher and a mixed one. In the first case, there is a required answer for every input pattern. Weights are adjusted in such a way that network outputs are as close as possible to the required answers. For the training without a teacher, there are no required answers for each input pattern. In this case the pattern clusterization occurs. In case of the mixed training, weight coefficients of one group of neurons are adjusted by means of training with a teacher, and another group is being self-trained.

The main feature, which makes training without a teacher attractive, is its independence. The training process, as in the case of training with a teacher, consists in adjusting synapse weights. Some algorithms also change the PH network structure as well. Such transformations of the structure are called self-organization. Adjustment of synapses weights can be performed only on the basis of information regarding the state of neurons and weight coefficients already available. The method of PH network organization is described in [2, 3].

III. Development of Mathematical Models of Parallel-Hierarchical Network Training

Based on the PH network structure and an analysis of training rules [6], a conclusion can be drawn that the most optimal solution for the PH network training will be an application of two methods of training with a teacher: an error correction-based training and memory-based training.

A parallel-hierarchical network consists of a number of layers, with each layer consisting of parallel branches. Starting from the second layer, each following branch being formed is shifted in time in relation to the previous branch. An incoming pattern is sent to the elements of parallel branches of the first layer, and elements of branches of the second and following layers are formed consequently in time from elements of branches of the first and branches of the previous layers respectively. A length of the algorithm of the PH network and, correspondingly, a number of elements in branches, a number of branches and a number of layers is not constant values, but is determined by a character of distribution of incoming data.

At the initial stage of the PH network self-training, the experience, presented by input patterns, entering the PH network, is accumulated in large quantities of correctly classified samples of "input-output" type. (1):

$$\left\{\left(x_i, d_i\right)\right\}_{i=1}^{N}, \tag{1}$$

where x_i is an input vector, d_i is a corresponding required initial signal.

As we develop a training model without a teacher, i.e. a desired output signal of the network is unknown, let us assume $d_i = 1$. This way input patterns are taken as reference once in the formation of the training sample.

After the training sample is formed (for static patterns it will contain 1000-1500 patterns, for dynamic – 10000 patterns), certain classes should be formed and clusterization of patterns, accumulated in the database performed. Ten percent of the images were used as the training sample for each type of images. This clusterization can lead to the situation that classes formed, to which input patterns fall, will either merge or, on the contrary, will split into additional classes, in case, if patterns of the same class turn out to be close.

For this purpose, a method of convex combination [7] should be applied, whose essence is in subjecting input patterns to transformation (2):

$$x_i = \alpha(t)x_i + \left(1-\alpha(t)\right)\frac{1}{\sqrt{n}}, \tag{2}$$

where x_i is an input vector, n is the dimensionality of the vector, $\alpha(t)$ is a coefficient that in the process of training varies from zero to unit.

Transformations of the input patterns consist of sending practically similar patterns to the network input and in the course of time t patterns come closer and closer to reference ones.

After that the system operates on the principle of training with a teacher. Operation of the PH network in such mode is described in [7].

We will briefly describe those methods of PH network training.

The signal vector $x(n)$ is an input signal of the PH network. Discrete time, or a step number of the iterative training process of the PH network, is denoted by n. An output signal of the PH network $y_k(n)$ is compared with the desired output $d_k(n)$; as a result, an error signal $e_k(n)$ is formed. By definition:

$$e_k\left(n\right) = d_k\left(n\right) - y_k\left(n\right). \tag{3}$$

The error signal will initialize a control mechanism, whose aim is an application of the sequence of corrections of the PH network error signal. These changes are aimed at step-wise approximation of the initial signal $y_k(n)$ to the desired signal $d_k(n)$. It is achieved at the expense of minimization of the cost function or of the index of productivity $E(n)$, which is calculated by formula (4):

$$E\left(n\right) = \frac{e_k^{\,2}\left(n\right)}{2}, \tag{4}$$

where $E(n)$ is a current value of error energy.

The step-wise correction of PH network parameters takes place until system parameters reach an absolutely balanced state (+1) in case of the correct recognition, or an absolutely non-balanced state (-1) – in case of the incorrect recognition. At this point the training process stops.

Minimization of the cost function $E(n)$ occurs according to the Delta Rule or Widrow-Hoff Rule [8, 9]. Let us denote a current value of the PH network parameter, corresponding to the input vector $x(n)$, the j[th] level of the PH network at the discretization step n. In accordance with the delta rule $\Delta\sigma_j(n)$, the parameter change at this step of discretization is defined by relation (5):

$$\Delta\sigma_j\left(n\right) = \eta e_k\left(n\right) x_j\left(n\right), \tag{5}$$

where n is a positive constant that determines a training process rate and is used when transiting from one step to another.

Having computed a value of the PH network parameter change $\Delta\sigma_j(n)$, we can determine its new value for the next discretization step:

$$\sigma_j(n+1) = \sigma_j(n) + \Delta\sigma_j(n). \tag{6}$$

Thus, they can be considered as an old and a new value of PH network parameters. Hence, we can write:

$$\sigma_j(n) = z^{-1}\left[\sigma_j(n+1)\right], \tag{7}$$

where z is a single delay operator. This element is a memory cell.

Under the memory-based training, all previous experience is accumulated in a large amount of correctly classified samples of the "input-output" type (8):

$$\left\{(x_i, d_i)\right\}_{i=1}^{N}, \tag{8}$$

where x_i is an input vector, and d_i is a corresponding desired initial signal. If a certain unknown vector x_i needs to be classified, an initial signal that corresponds to the input signal, close to x_i, is selected from the database.

A memory-based training algorithm of the PH network includes two main components:

- a criterion, used for determination of x_i vector surrounding;

- training rules, applied to a sample from the test vector surrounding.

According to the nearest neighbour rule [8], a sample nearest to the test one, is included into the neighbors.

Let us consider a vector (9):

$$x'_N \in \left\{x_1, x_2, \ldots x_N\right\} \tag{9}$$

This vector is considered to be the nearest neighbor of vector x_i, if the condition (10) is satisfied:

$$\min d\left(x_j, x_i\right) = d\left(x'_N, x_i\right), \tag{10}$$

where $d(x_i, x_j)$ is the Euclidean distance between vectors x_i and x_j [10]. This training rule does not depend on the distribution law used in samples generation in the course of training.

A variation of the nearest neighbour-based classifier is a classifier of K-nearest neighbors, which can be described in the following way.

A set of k classified neighbors nearest to the input test vector x_i, where k is certain integer, are found. Vector x_i is referred to the class, being met among more frequently than others among k- closest neighbors.

Therefore, a classifier based on the k nearest neighbors rule operates similarly to the averaging unit. For instance, it may not take into account single outliers (i.e., observations differing from the nominal model), as shown in Fig 1:

Figure 1 Graphic interpretation of the memory-based PH network training method

In Fig.1, the area in dash circle contains two points, belonging to class 1 and one point belonging to class 0; point d corresponds to test vector x_i. If k=3, the classifier based on k nearest neighbors rule, will refer point d to class 1, despite the fact, that it lies closer to the class 0 outliers.

IV. Development of the Parallel-Hierarchical Network Training Methods

Results of the correlative comparison of images of different faces carried out in [7], show that an informative indicator is a correlation factor at the zero level of the PH network. We will use it to verify pattern correspondence, as well as to determine a desired answer in the error correction-based training method.

Results of the correlative comparison of images of different faces by the PH network levels [6, 11] indicate that informative indicators for face recognition are correlation factors at the first, middle, the last but one and third from the end levels. Those indexes will be auxiliary when using the PH network training method for recognition of static patterns.

For PH network training for static patterns recognition we will use a combined method, containing "self learning" technology-based training methods. A detailed description of the combined method is provided below.

A sequence of the combined method is based on errors correction and application of PH network memory in recognizing static patterns:

1. Construction of the PH network structure of a selected image.
2. Application of the memory-based training method. A number of the PH network levels is a classifier based on k-nearest neighbors used in the given method.
3. Construction of the PH network structure of the reference image.
4. Application of the error correction-based training method. A criterion used for verification in this method, is a correlation factor at the zero level of the PH network.
5. Error correction for correlation factors is performed at the zero, first, middle, last but one and third from the end levels of PH network. The correction is performed until network parameters reach an absolutely balanced state (+1), or exceed it. If the value of parameters leads to state (-1) (or an absolutely unbalanced state), then the conclusion is drawn that this image does not belong to the given pattern.

Results of processing of dynamic patterns of laser beam spot images [12] indicate that the informative indicator is an average correlation factor by PH network levels for formation of the lower boundary of the tunnel. A problem of formation of tunnel boundaries was solved on the basis of tunneling by averaged correlation coefficients when comparing a reference pattern with a separate fragment of the route. The PH network training included realization of the following stages:

1) find a point with maximum brightness $T_{\max}$ for the route fragment;

2) select a necessary number N of brightness values $T_{\max-w}$, $w = 1, 2, \dots, N$, starting with the maximum one (extreme points with brightness T_{max-w} form N boundary lines);

3) for each boundary contour, determine preliminary centers separately for each coordinate:

$$x_c = \frac{1}{M}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} w\big(f(x,y)\big)\cdot x,$$
$$y_c = \frac{1}{M}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} w\big(f(x,y)\big)\cdot y,$$
$$M = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} w\big(f(x,y)\big), \tag{11}$$

where $w = 1, 2, \dots, N,\ f(x,y) -$ is the image brightness function;

4) find averaged coordinates for the whole fragment of the route (data for averaging are coordinates of each boundary contour center) [10];

5) find a reference pattern (for the training sample, averaged brightness values of decomposition elements comprising the route fragment, are found for each route fragment);

6) perform tunneling by averaged correlation coefficients when comparing a reference pattern with an individual route fragment (1 is selected as the upper boundary of the tunnel, and an averaged correlation factor is selected as a bottom boundary);

7) determine "good" or "bad" images (a route fragment is considered to be "good" if the average correlation factor hits the tunnel of a reference pattern, otherwise it is considered to be "bad");

8) perform weights optimization for the "good" image by formula (12):

$$x_{ec} = \frac{1}{M}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} w\big(f(x,y)\big)\cdot\big(x+\delta_X\big)$$
$$y_{ec} = \frac{1}{M}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} w\big(f(x,y)\big)\cdot\big(y+\delta_Y\big) \tag{12}$$

where $w = 1, 2, \dots, N,\ f(x,y) -$ is brightness, $\delta_x\ \delta_y$ are center shifts by X- and Y-axes correspondingly.

The shift of centers on X- and Y-axes are determined by formulas (13) and (14) respectively:

$$\delta_X = \frac{1}{N}\sum_{i=0}^{N-1}\left(\bar{x} - x_i\right), \tag{13}$$

$$\delta_y = \frac{1}{N}\sum_{i=0}^{N-1}\left(\bar{y} - y_i\right), \tag{14}$$

where $\bar{x}, \bar{y}$ are average values of the center of the sample of images by X- and Y-axes respectively. x_i, y_i are center coordinates of each ith "good" image.

The obtained values x_{ec}, y_{ec} are the final result of the measurement of the energy center coordinates of laser beam spot images.

This indicator will be used to verify pattern correspondence, as well as to determine a desired response in the error correction-based training method.

Proceeding from the task of parameter control of laser beam radiation, we will use x and y coordinates of route fragment energy centers as basic indicators for recognition of laser beam spot images. Correlation factors at the first, middle, last but one and the third from the end levels will be auxiliary in applying the PH network training method for dynamic pattern recognition.

To train the PH network for the dynamic pattern recognition, the combined method will be used that contains error correction-based and memory-based training methods.

A sequence of the combined method of the PH network training for the dynamic pattern recognition comprises the following:

1. Construction of the PH network structure of the selected laser route fragment.

2. Application of the memory-based training method. A number of common levels of the PH network is a classifier based on the k-nearest neighbors rule, used in this method.

3. Construction of the PH network structure of the reference pattern.

4. Application of the error correction-based training method. The criterion used for verification in this method, is an average correlation factor by the PH network levels.

5. Error correction for X and Y coordinates of route fragment energy centers and for correlation factors at the first, middle, last but one, and the third from the end PH network levels. The correction is performed until network parameters reach an absolutely balanced state. If the value of parameters leads to an absolutely unbalanced state, the conclusion is made that the laser route fragment does not belong to the given pattern.

V. Results of the Experimental Research of Parallel-Hierarchical Network Self-Training under Recognition of Static and Dynamic Patterns

1000 images of human faces from the Manchester database of human faces were used for classification and recognition of static patterns during the PH network self-training [13].

To verify the quality of pattern classification and recognition, we selected image AGJ-08.

After the procedure of static pattern training sample formation, we obtain a database, whose fragment is presented in Table I.

A comparative analysis of models of learning PH network is presented in Table II.

Having applied a verification of image correspondence with the correlation factor at the zero level of the PH network, the following picture will be obtained (Table III).

TABLE I. Formation of the static pattern training sample during PH network self-training

Test image AGJ-02: 

Selected patterns: 1) AL-01  2) AGJ-01 

Training step, n	Image	$\sigma_0(n)$	$\sigma_1(n)$	$\sigma_{AVR}(n)$	$\sigma_{L-1}(n)$	$\sigma_{L-2}(n)$
1	AGJ-02	0,61	0,8	0,57	0,5	0,7
2	AGJ-02	0,75	0,82	0,77	1	0,9
3	AGJ-02	0,89	0,84	0,97	-	1
4	AGJ-02	1,03	0,86	1,17	-	-

TABLE II. A comparative analysis of the efficiency of PH network training and self-training models

Parameter or property	Training paradigm	
	PH network training with a teacher	PH network self-training
1. Training methods and algorithms used in the given paradigm, their application order.	Error correction- and memory-based network training paradigms are used. Error correction is performed through the cost function minimization by the delta rule. Patterns are classified by the nearest neighbour rule [8].	Error correction- and memory-based network training paradigms are used. Patterns are optimized based on the convex combination method.
2. Problems solved by means of the model and its application areas	Pattern recognition and classification, forecast, function approximation, control.	Patterns clusterization, (in-class categorization), data analysis
3. Brief description of the model	The signal vector $x(n)$ is sent to the network input. The output signal of the PH network $y_k(n)$ is compared with the desired output $d_k(n)$, and an error signal $e_k(n)$ is formed as a result. The error signal will initialize a control mechanism, whose purpose is to apply a sequence of corrections of the PH network error signal. These changes are aimed at the step-wise approximation of the initial signal $y_k(n)$ to the desired one $d_k(n)$. The step-wise correction of PH network parameters takes place until the system parameters reach an absolutely balanced state (+1) in case of the correct recognition, or an absolutely unbalanced state (-1) in case of the incorrect recognition. At this point training process stops. An unknown vector x_i is classified by a sampling from a database of the initial signal corresponding to the input signal close to x_i. A variation of the nearest neighbour-based is a classifier of the *k*–nearest neighbors [8].	At the initial stage of the PH network self-training, all the experience presented by input patterns, arriving at the PH network input, is accumulated in the large amount of correctly classified patterns of the "input-output" type. During the formation of the training sample we take input patterns as reference ones. After the training sample is formed, formation of classes and clusterization of patterns accumulated in database takes place. This clusterization may lead to the situation when the classes formed, to which input patterns correspond, will either merge or, vice versa, split into additional classes in case, if patterns of the same class turn out to be close. In this case, the convex combination method should be applied [14]. The input pattern transformations of are performed in such a way that, first, practically identical patterns are sent to the network input, and in the course of time *t* the patterns more and more approximate the reference patterns. After that the system functions according to the principle of training with a teacher.
4. Advantages of the model	Hardware and software realization is simple and allows applying methods of parallel programming.	With application of the developed model of the PH network training. a situation with the *k*- nearest neighbors-based classifier rule is excluded. It is achieved due to the fact that at the initial stage of self-training various patterns are entered into the database, and, subsequently, outliers may belong to one of the formed classes.
5. Drawbacks of the model	During the PH network training with a teacher [8], a *k*-nearest neighbors-based classifier operates similarly to an averaging unit and does not take into account single outliers (observations that differ from the nominal model), which may influence a low percentage of "good" images in the process of dynamic pattern recognition	Loss of time as a result of the additional verification of coincidence of pattern vectors during formation of classes.

TABLE III. FORMATION OF CLASSES AND CLASSIFICATION OF STATIC PATTERNS

Image	Num. of the PH network levels	Class	σ_0 (n)	σ_1 (n)	σ_{AVR} (n)	σ_{L-1} (n)	σ_{L-2} (n)
AGJ-01	108	AGJ-08	1	1	0	1	1
AGJ-02	107	AGJ-01	1	1	0	1	1
AGJ-03	93	AGJ-08	1	1	0	1	1
AGJ-04	101	AGJ-08	1	1	0	1	1
AGJ-05	93	AGJ-05	1	1	0	1	1
AGJ-06	107	AGJ-05	1	1	0	1	1
AGJ-07	99	AGJ-02	1	1	0	1	1
AGJ-08	103	AGJ-08	1	1	0	1	1
AGJ-09	99	AGJ-09	1	1	0	1	1
AGJ-10	100	AGJ-08	1	1	0	1	1
AL-01	84	AL-01	1	1	0	1	1
AL-02	100	AGJ-08	1	1	0	1	1

The performed classification of static patterns may be presented as a diagram of the proportion of correctly classified patterns in the number of patterns in the class. The resulting diagram is shown in Fig 2 (for part of all images).

Resulting data of the static pattern classification (for different classes) are presented in Table IV.

Patterns correspondence is be verified by means of averaged correlation factors by the PH network levels.

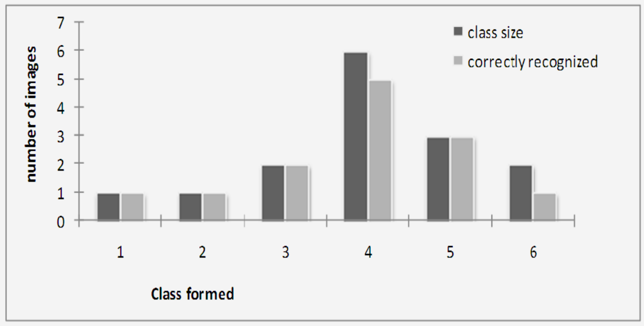

Figure 2 Classification of static patterns during the PH network self training

TABLE IV. PERCENTAGE OF CORRECT RECOGNITION OF STATIC PATTERNS

№	Pattern	Percentage of correct recognition, %
1.	AGJ-01	100
2.	AGJ-02	100
3.	AGJ-05	100
4.	AGJ-08	83,3
5.	AL-01	100
6.	AL-02	50
Average percentage of correct recognition, %		88,88

As we can see, an average percentage of correct recognition was 88,88%. Further classification is performed on the basis of the convex combination method.

In the process of PH network self-training, 10000 route fragments were used for classification and recognition of dynamic patterns.

To verify the quality of pattern classification, route fragment TV-030 will be selected (Fig.4).

After the procedure of formation of the training sample of dynamic patterns, a database is obtained, a fragment of which is presented in Table V.

TABLE V. CREATION OF CLASSES AND CLASSIFICATION OF DYNAMIC PATTERNS

No	Route fragment	Number of PH network levels	Class	X	Y	$\sigma_0(n)$	$\sigma_1(n)$	$\sigma_{AVR}(n)$	$\sigma_{L-1}(n)$	$\sigma_{L-2}(n)$
1	TV-001	35	TV-017	66,0	40,0	1	1	1	1	1
2	TV-003	25	TV-026	66,0	40,0	1	1	1	1	1
3	TV-004	31	TV-026	66,0	44,0	1	1	1	1	1
4	TV-005	42	TV-026	66,0	40,0	1	1	1	1	1
5	TV-006	25	TV-017	74,0	43,0	1	1	1	1	1
6	TV-007	26	TV-015	80,0	45,0	1	1	1	1	1
7	TV-008	30	TV-024	66,0	38,0	1	1	1	1	1
8	TV-009	34	TV-026	71,0	40,0	1	1	1	1	1
9	TV-010	27	TV-026	75,0	42,0	1	1	1	1	1
10	TV-011	37	TV-015	78,0	43,0	1	1	1	1	1
11	TV-012	31	TV-015	70,0	39,0	1	1	1	1	1
12	TV-013	37	TV-016	73,0	40,0	1	1	1	1	1
13	TV-014	31	TV-015	76,0	41,0	1	1	1	1	1
14	TV-015	31	TV-023	78,0	42,0	1	1	1	1	1
15	TV-016	25	TV-023	73,0	39,0	1	1	1	1	1
16	TV-017	46	TV-023	75,0	40,0	1	1	1	1	1
17	TV-018	46	TV-026	77,0	41,0	1	1	1	1	1
18	TV-020	37	TV-026	74,0	39,0	1	1	1	1	1
19	TV-021	38	TV-026	76,0	40,0	1	1	1	1	1
20	TV-022	39	TV-027	78,0	41,0	1	1	1	1	1
21	TV-023	36	TV-026	79,0	41,0	1	1	1	1	1
22	TV-024	40	TV-027	75,0	40,0	1	1	1	1	1
23	TV-025	29	TV-027	77,0	40,0	1	1	1	1	1
24	TV-026	47	TV-027	78,0	41,0	1	1	1	1	1
25	TV-027	45	TV-027	80,0	41,0	1	1	1	1	1

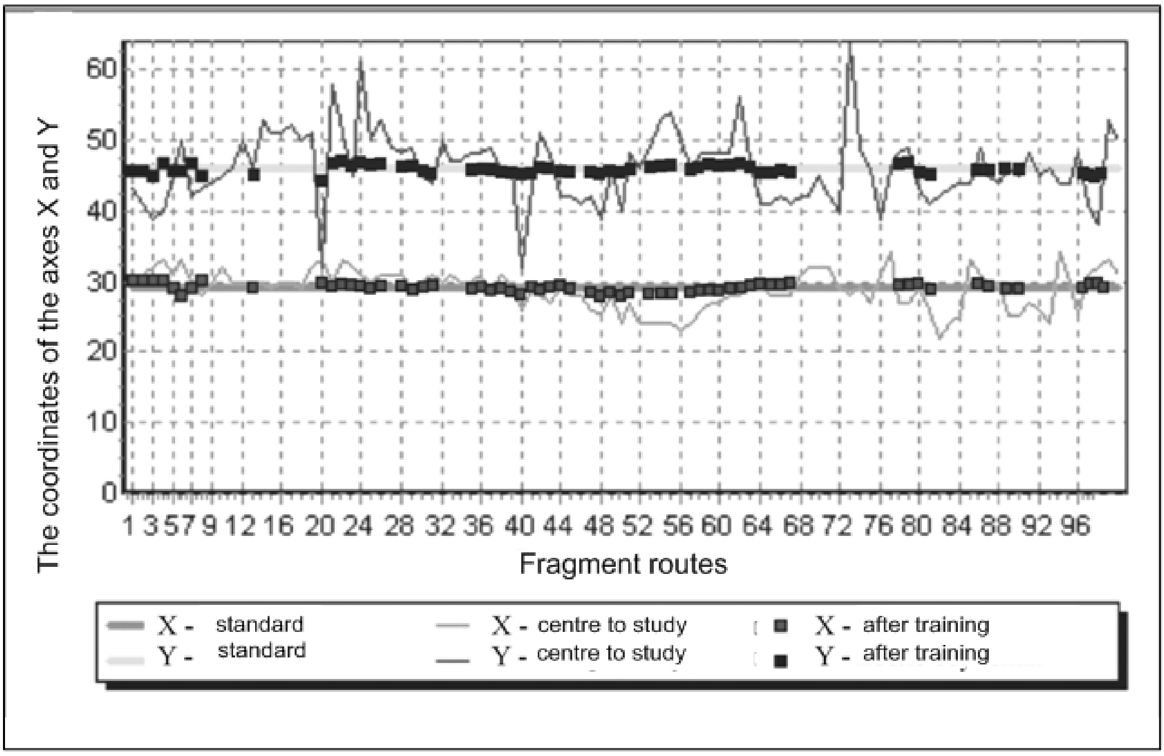

Figure 3. Classification of dynamic patterns during the PH network self-training

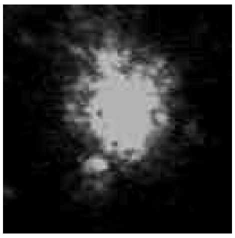

Figure 4. Test fragment of trace TV-30

Fig. 3 demonstrates a diagram representing the performed classification of dynamic patterns in the form of the proportion of correctly classified patterns in the total number of patterns in the class:

Obtained data of classification of dynamic images are shown in Table VI.

TABLE VI. CREATION OF CLASSES AND CLASSIFICATION OF DYNAMIC PATTERNS

No	Pattern	Percentage of correct recognition %
1.	TV_009	100
2.	TV_010	100
3.	TV_015	80
4.	TV_016	0
5.	TV_017	0
6.	TV_023	67,7
7.	TV_024	0
8.	TV_026	37,5
9.	TV_027	83,3
Average percentage of correction recognition,%		52,05

As we can see, an average percentage of correct recognition is 52.05%. Further classification is performed based on the convex combination method.

Now we will briefly describe technical advantages of considered methods of PH network self-training, proceeding from the results of experimental research (Table VII):

TABLE VII. COMPARATIVE CHARACTERISTIC OF INFORMATION TECHNOLOGIES OF THE PH NETWORK TRAINING

No	Parameters used in PH network	Information technology [15,16]	Developed information technology
Recognition of static patterns of human face images			
	Training method used	Method of PH network training with a teacher [8].	Method of PH network self-training
a)	Average percentage of correct recognition of static patterns at the zero level of PH network, %	93.75	88.88
b)	Average percentage of correct recognition of static patterns with construction of correlation curves by levels, %	84	90
c)	Average recognition time, sec	Depending on the level of the PH network training (from 0.5 to 12)	Depending on the analysis of correlation coincidence of patterns (from 0.1 to 2)
Recognition of dynamic patterns of laser beam spot images			
a	Average percentage of “good” route fragments %	38.4	40
b	Average percentage of correct recognition, %	92.5	52.05
c	Accuracy of reference point determination, not more, decomposition elements	0.01	0.01
d	Average time of recognition, sec	Depending on the level at the PH network training (from 0.5 to 60)	Depending on the analysis of correlation coincidence of patterns (from 0.1 to 1)

VI. Conclusion

Propositions necessary for development of the PH network training methods are discussed in the paper. Unlike already known structures of artificial neural networks, where non-normalized (absolute) similarity criteria are used for comparison, in the suggested method of PH network organization, a normalized criterion is used. Based on the analysis of training rules, a conclusion is made that the application of two training methods with a teacher is optimal for PH network training: error correction-based training and memory-based training. Mathematical models of PH network training and a combined method of PH network training for recognition of static and dynamic patterns were developed.

According to results of the experimental research of PH network training with recognition of the test pattern AGJ-02 and selection of the pattern #1 AL-01 from database, parameters of the PH network reached an unbalanced state at the third step. With recognition of test pattern AGJ-02 and selection of pattern #1 AGJ-01 from the database, PH network parameters reached a balanced state at the fourth step. That is why a conclusion is made that image AGJ-02 belongs to pattern AGJ-01, which means that the recognized object is classified correctly.

During the recognition of test fragment TV-029 and selection of the pattern from the database, correlation parameters of PH network reached a balanced state at the eleventh step, and coordinate parameters of the PH network – at the sixteenth step.

The most important technical advantages of the suggested method of PH network training realized in the new software as compared with previous software and methods it used, are the following: measurement accuracy after correction of determination of route fragment energy centers was 0.01 decomposition element (as compared to 1,2 decomposition element). The percentage of "good" images was 38.4% (as compared to 50%), that increases the total percentage of correct recognition to 92.5% (84.8%). However, an indicator of the "average recognition time" at the initial level for PH network training is 60 sec (30 sec).

An average percentage of correct recognition at the zero level of the PH network increased 93.75% (80%) for recognition of static patterns. A considerable growth of the average percentage of correct recognition of static patterns in constructing correlation curves by levels was also noted: 84% (as compared to 5%). An indicator of "average recognition time" at initial level of PH network training is 10.73 sec (12 sec).

Acknowledgment

Authors thank to Yu.F. Kutayev for problem formulation at the initial stage of the research and for video data of fragments of laser rout images made available for the scientific research.

Authors express their sheer gratitude to the team of the Chair of the Telecommunication Technologies and Automatics of the State Economy and Technology University of Transport (Kyiv, Ukraine) for their valuable comments and recommendations expressed in the course of the experimental research. Authors also thank Professor M.V. Makarenko, Vice Rector on Scientific Work Associate PhD V.I. Danilevskiy, and Dean of the Department of Infrastructure and Rolling Stock of Railway Transport Professor A.D. Voznenko for provided facilities and equipment for scientific studies of this problem in the framework of the state-financed scientific research topics.

References

[1] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, Vol. 5, pp. 115-133, 1943.

[2] L.I. Timchenko, V.V. Melnikov, N.I. Kokryatskaya, Yu. F. Kutaev, I.D. Ivasyuk. A method of organization of a parallel-hierarchical network for image recognition. Journal Cybernetics and system analysis. , Vol.47 (1), pp. 140-151, 2011a.

[3] L.I. Timchenko, V.V. Melnikov, N.I. Kodryatskaya, Parallel-hierarchical network learning methods and their application to pattern recognition, Cybernetics and Systems Analysis, 47(6), 2011.

[4] M. Hirahara, N. Oka, T. Kindo. A cascade associative memory model with a hierarchical memory structure. Journal Neural Networks, Vol.13, Issue 1, pp. 41-50, 2000.

[5] J. Sacramento, A. Wichert. Tree-like hierarchical associative memory structures. Journal Neural Networks, pp. 143-147, 2010.

[6] L.I. Timchenko. A multistage parallel-hierarchic network as a model of a neurolike computation scheme. Journal Cybernetics and system analysis. - Vol.36(2), pp. 251-267, 2000.

[7] L.I. Timchenko, Y.F. Kutaev, S.V. Chepornyuk, M.A. Grudin, A.A. Gertsiy.A brain-like approach to multistage hierarchical image processing. Springer-Verlag Processing. - in Proc. Image Analysis and Processing, Florence, Italy, pp. 246 – 253, 1997.

[8] D.E Hinton. How do neural networks train? In the world of science, 11, 1992.

[9] B. Widrow, and M.A. Lehr. 30 years of adaptive neural networks: Perceptron, madaline and backpropagation. Proceedings of the Institute of Electrical and Electronics Engineers, Vol. 78, p. 1415-1442, 1990.

[10] T. Hastie, R. Tibshirani, J Friedman. The Elements of Statistical Learning. Springer, 2001.

[11] S. Gadat, L. Younes. A stochastic algorithm for feature selection in pattern recognition. Research Journal of Machine Learning Research (8), pp. 509-547, 2007.

[12] L.I. Timchenko, N.I. Kokryatskaya, A.A. Poplavskyy, A.A Poplavska, I.D. Ivasyuk. Method of reference tunnel formation for improvement of forecast results of laser beams spot images behavior. 18th International Conference IWSSIP-2011, pp. 1-3., 2011b.

[13] Manchester base of human faces. Access mode: http://peipa.essex.ac.uk/ipa/pix/faces/manchester, 2003.

[14] L.I. Timchenko, Y.F. Kutaev, V.P. Kozhemyako, et al. Method for Training of a Parallel-Hierarchical Network, Based on Population Coding for Processing of Extended Laser Paths Images. Proceedings of SPIE, Vol. 4790, pp. 465-479, 2002.

[15] Tom Mitchel. Machine Learning . McGraw Hill, 432p, 1997.

[16] V.P. Kozhemyako, E. I. Ponuraya, V. Belokonniy. Logic-temporal functions processing for object recognition. Selected papers from the International Conference on Optoelectronic Information Technologies. Bellingham, Wash., USA, SPIE,+ Vol.4425, pp. 35-40, 2001.