Krasilenko V.G., C. Sc., Lazarev A.A., C. Sc., Nikitovich D.V.
Vinnytsia National Technical University

MODELING OF SELF-LEARNING EQUIVALENT-CONVOLUTIONAL NEURAL STRUCTURES FOR IMAGE FRAGMENTS CLUSTERING AND RECOGNITION

Introduction. For creation of biometric identification and machine vision systems necessary solve the problem of object recognition in images. The basis of most known methods is to compare two different images of object, or its fragments, one of which is a benchmark and the second image is a set of images that belong to different classes. Discriminant measure of the reference fragment with the current image fragment, the coordinate offset is often two-dimensional correlation function. In our work [1] it was shown that to improve accuracy and recognition qualities of distorted and correlated images, it is desirable to use methods based on mutual equivalently two-dimensional spatial functions, nonlinear transformations of adaptive-correlation weighting. For the recognition and clustering of images, for modeling associative memory, biometric identification and robotic devices equivalence models (EMs) of auto-associative memory (AAM) and hetero-associative memory (HAM) were proposed [2, 3]. The simulation results of such EMs [2-4] were showed and confirmed that the EM has advantages. These EM HAM studies have shown that models allow the recognition of large-size vectors and a significant percentage (up to 25-30%) of damage, at a network power that is 3 to 4 times higher than the number of neurons. At the same time, knowing the significant advantages of EM when creating on their basis improved NNs, multiport AAM and HAM [3, 4], there was a suggestion about the possibility of modifying EM, MHAM for parallel cluster image analysis [4, 5]. Hardware implementations of these EMs are based on structures, including multipliers, equivalentors with spatial, time integration [3]. In previous work [5], these questions were considered for bitmaps of multi-level images. Therefore, in this paper, we want to generalize and show that the self-learning concept works with directly multi-level images without processing the bitmaps. In addition, the previous work did not investigate the influence of the size of the filters on the convergence of the self-study method and no simulation was carried out for different types and dimensions of the images.

Presentation of the main material. We consider the based on MHAM idea of clustering, which can be used to simultaneously calculate the corresponding distances between all cluster-neurons (CN) and all training vectors. This approach allows the use of MAAM for parallel calculation of distances between a CN and learning neurons, to identify and mark all winners corresponding to each cluster learning vector (CV). As metrics, we use generalized normalized vector equivalence functions. This method gives good convergence, speed and describes an iterative learning process, that consists in computing of optimal weight vectors for all cluster-neurons using the training matrix TX (see Fig.1) [5]. Unlike previous works in this work, we should look for a set of optimal matrix templates rather than a set of vectors. We specify the number of clusters and their size. A visualization of optimal templates is formed by the corresponding iterative procedure on the basis of the definition of regularities in fragments that are in the images from the set of trainees.

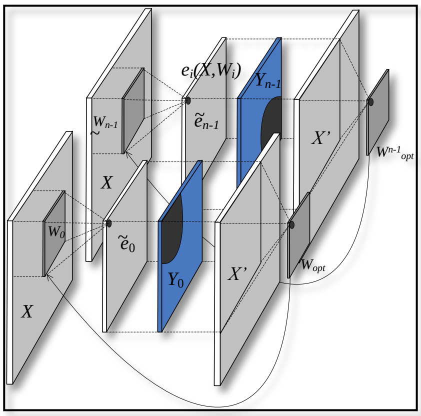

Fig. 1. The structure of the basic unit of the SL_EC_NS, which explains the principle of its functioning

Therefore interpretation method for spatially invariant case requires the calculation of spatial features convolution-type,

$\mathbf{E}^m = \mathbf{W}^m \overset{t}{\tilde{\otimes}} \mathbf{X}$, where

$E^m_{k,l} = 1 - mean\ \left| submatrix\ \overrightarrow{\mathbf{X}, k, k + r_0 - 1, l, l + r_0 - 1\ - \mathbf{W}^m} \right|$, nonlinear processing by the expression

$$EN^m_{k,l} = G\ \ a, E^m_{k,l}\ \ = 0{,}5\left[1 +\ \ 2E^m_{k,l} - 1\ ^a\right],$$

comparing each other to determine the winners for indexing expressions: $MAX_{k,l} = \max_{index\ m}\ EN^{m=0}_{k,l}, EN^{m=1}_{k,l} ... EN^{M-1}_{k,l}$ and $EV^m_{k,l} = f^{activ}_{nonlinear}\ \ EN^m_{k,l}, MAX_{k,l}$. The first algorithmic step defines all matrices $\mathbf{EQ}^m(a, \mathbf{W}^m, \mathbf{MAX}, \mathbf{X}) = \mathbf{EV}\left(\mathbf{EN}_{nonlinear}\left(\mathbf{W}^m \overset{t}{\tilde{\otimes}} \mathbf{X}\right)\right)$ and considering the second step (convolution of the latter with the matrix **X**) the model of proposed method will look like:

$\mathbf{W}^m\ \ t+1\ = \mathbf{EQ}^m(\mathbf{EQ}^m(a, \mathbf{W}^m, \mathbf{MAX}, \mathbf{X}), \mathbf{X}) = \mathbf{EV}\left(\mathbf{EN}\left(\mathbf{EV}\left(\mathbf{EN}\left(\mathbf{W}^m_t \overset{t}{\tilde{\otimes}} \mathbf{X}\right)\right)\overset{t}{\tilde{\otimes}} \mathbf{X}\right)\right)$. In Fig. 1 shows the block diagram of the main unit of the SL_EC_NS, offered by us. The idea is that by feeding an input multilevel image, which is one color image spectral component, the matrix **X**, to the unit input, we form a certain number of convolutions in the form of matrices $e_0 \div e_{n-1}$ using a set of defined filters-templates $\mathbf{W}_0 \div \mathbf{W}_{n-1}$ which, in our case, are multilevel values. As a measure of the similarity of the fragment of the matrix X and the filter the equivalent measures of proximity or other measures such as a histogram and other features can be used. Thus, we compare for each filter similar fragments in the matrix. Having obtained these convolutions, we per-pixel form a maxima map on the base of normalized and non-linearly transformed equivalence convolution maps. Thus, clusteral maps $\mathbf{Y}_0 \div \mathbf{Y}_{n-1}$ are formed in which single signals correspond to those neurons of the hidden layer that were winners and which are responsible for all possible shifted fragments in the matrix **X** which are closest to this filter. These maps $\mathbf{Y}_0 \div \mathbf{Y}_{n-1}$ must be disjoint sets. With each map **Y**, we again convolve with the original matrix **X** and again define the normalized equivalence measures and form output arrays as a set of new filters $\mathbf{W}_0 \div \mathbf{W}_{n-1}$ (t+1), which are designated in the Mathcad simulation as **WT0–WT7.** Each of them is a weighted average equivalent function. This process is repeated and after several steps, the optimal values of these filters are formed. The process of self-learning is combined with the recognition process. Similarly, when all the RGB components are at the input in sequence, auto-coding and decoding are realized by using the same basic modules. It turns out that a set of received maps and filters allows you to restore the image X. To restore the image, you can also apply several approaches. The number of filters will affect the accuracy of recovery (auto-decoding). But, as we have found, often, only with a small amount (8-15!) of small-sized filters all binary sections of a multilevel image can be decoded with an accuracy of 90-97% [5]. Below we will show that for multilevel images and filters, an insignificant number of iterative steps is also required. In the proposed self-learning model, only the number of filters, their size and the required number of units are specified from the outside taking into account the tasks to be solved. Sometimes the number of filters can be adaptively chosen taking into account the errors in auto-decoding of images. To model the base unit in the first experiment, we used a color image "eye" with a size of 360x362 elements and 8 filters of size 3x3, the elements of which were randomly selected for the first iteration. Simulation results for all components are shown in Fig. 2.

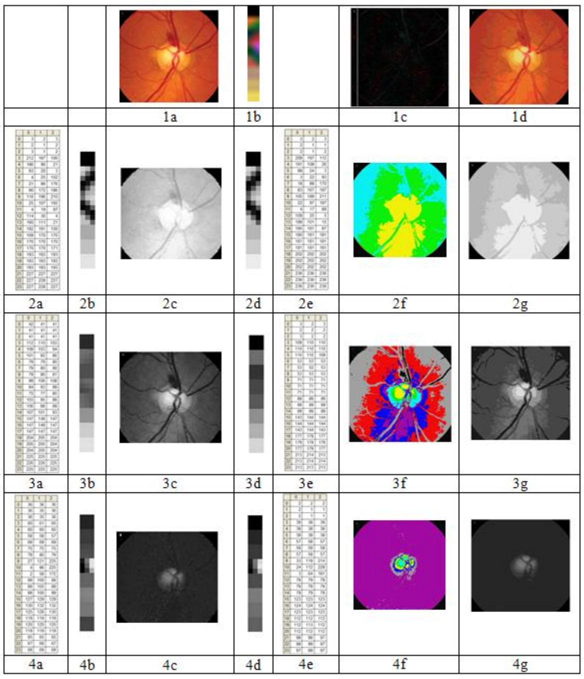

Fig. 2. Fragments of Mathcad windows obtained as a result of modeling which explain the principle of self-learning and adjusting filters and restoring the components of a color image: 1a - Input RGB color image, 1b - View of 8 color filters after self-learning, 1c - Color image recovery error, 1d - Restored image using 8 filters; 2a - Set of 8 filters (3x3) in the form of a matrix, 2b - Set of 8 filters (3x3) in the form of a multilevel image, 2c - R spectral component, 2d - Created filters after convolution and non-linear processing 17 iteration, 2e - Digital format of the created filters, 2f - The color indicates 8 clusters, 2g - Restored R component using 8 filters; 3a – 3g – same as for 2a-2g but for G component; 4a – 4g – same as for 2a-2g but for B spectral component

The analysis of the results shows that the number of iterative steps for self-learning of filters does not exceed 30-40, and can often be reduced by tuning the filters on the trained filters of the other spectral component. The type of concrete maps obtained in the color and gray formats and their changes for iteration steps and different filter sizes are shown in Fig. 3.

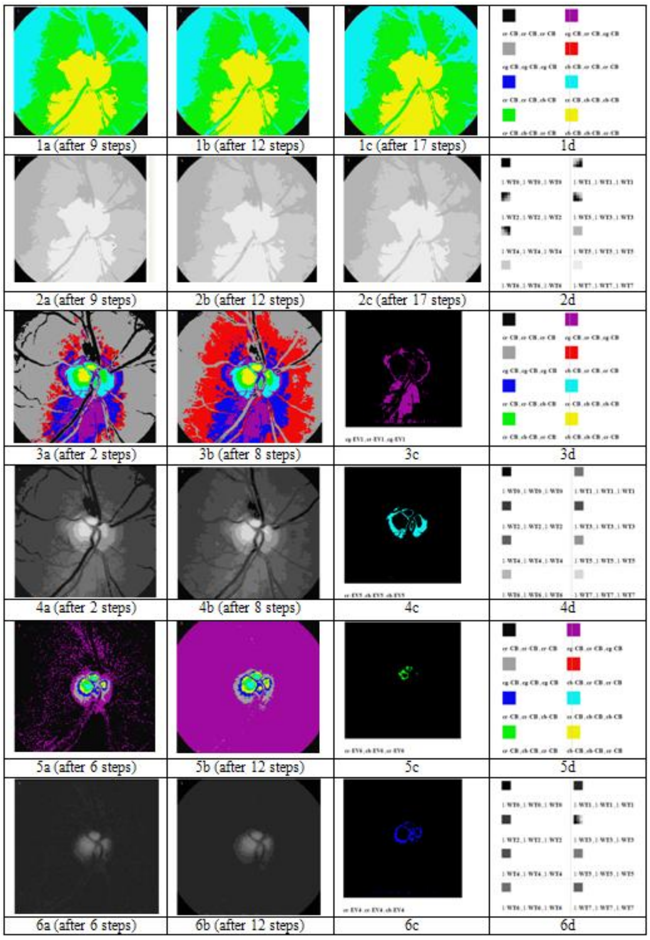

Fig. 3. Dynamics of the change of clusters of R, G, B components of color image for 3x3 filters size: 1a–1c – clusters R, 1d, 3d, 5d – color set of clusters; 2a–2c – gray clusters, 2d - set of filters R; 3a-3b - clusters G, 3c - Map EV1 view after 2nd iteration; 4d - set for G; 5c, 6c - EV6 view after 6th and EV4 (12th); 6d set of B.

The analysis of image transformations and its histogram, when it is encoded-compressed and decoded, shows that filters try to change their values so that they take the positions that are closest to the statistical distribution of the original histograms, as can be seen from Fig. 4. The proposed method of combined learning-recognition is corresponds to the optimal methods of coding, and also is effective for self-segmentation and archiving. The set of filters is adjusted to the contents of the analyzed images.

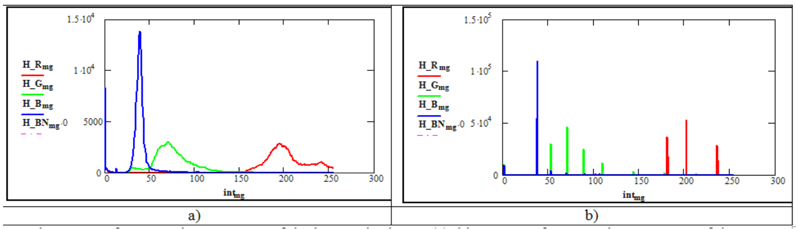

Fig. 4. Histograms of 3 spectral components of the input color image (a), histograms of 3 spectral components of the reconstructed color image (b)

Література:

1. Krasilenko V. G., Nikolskyy A. I., Bozniak Y. A., "Recognition algorithms of multilevel images of multicharacter identification objects based on nonlinear equivalent metrics and analysis of experimental data," Proc. of SPIE Vol. 4731, pp.154-163 (2002).

2. Krasilenko V. G., Magas A. T., "Multiport optical associative memory based on matrix-matrix equivalentors," Proc. of SPIE Vol. 3055, pp. 137 - 146.

3. Krasilenko V. G., Lazarev A., Grabovlyak S., "Design and simulation of a multiport neural network heteroassociative memory for optical pattern recognitions," Proc. of SPIE Vol. 8398, 83980N-1 (2012).

4. Krasilenko V. G. , Lazarev A., Grabovlyak S., Nikitovich D., "Using a multi-port architecture of neural-net associative memory based on the equivalency paradigm for parallel cluster image analysis and self-learning," Proc. of SPIE Vol. 8662, 86620S (2013).

5. Krasilenko V.G., Lazarev A.A., Nikitovich D.V., "Modeling and possible implementation of self-learning equivalence-convolutional neural structures for auto-encoding-decoding and clusterization of images," Proceedings of SPIE Vol. 10453, 104532N (2017).