OBJECT DETECTION USING NEURAL SELF-ORGANIZATION

Arpad Barsi

Department of Photogrammetry and Geoinformatics
Budapest University of Technology and Economics
H-1111 Budapest, Muegyetem rkp. 3, Hungary
barsi@eik.bme.hu

Commission III, Working Group III/4

KEY WORDS: Neural networks, Object detection, Modeling, Data structure

ABSTRACT:

The paper presents a novel artificial neural network type, which is based on the learning rule of the Kohonen-type SOM model. The developed Self-Organizing Neuron Graph (SONG) has a flexible graph structure compared to the fixed SOM neuron grid and an appropriate training algorithm. The number and structure of the neurons express the preliminary human knowledge about the object to be detected, which can be checked during the computations. The inputs of the neuron graph are the coordinates of the image pixels derived by different image processing operators from segmentation to classification. The newly developed tool has been involved in several types of image analyzing tasks: from detecting building structure in high-resolution satellite image via template matching to the extraction of road network segments in aerial imagery. The presented results have proved that the developed neural network algorithm is highly capable for analyzing photogrammetric and remotely sensed data.

1. INTRODUCTION

Artificial neural networks have quite long history. The story has started with the work of W. McCulloch and W. Pitts in 1943 (Rojas 1993). Their paper presented the first artificial computing model after the discovery of the biological neuron cell in the early years of the twentieth century. The McCulloch-Pitts paper was followed by the publication from F. Rosenblatt in 1953, in which he focused on the mathematics of the new discipline (Rosenblatt 1953). His perceptron model was extended by two famous scientists in 1969: M. Minsky and S. Papert.

The year 1961 brought the description of competitive learning and learning matrix by K. Steinbruch (Carpenter 1989). He published the "winner-takes-all" rule, which is widely used also in modern systems. C. von der Malsburg wrote a paper about the biological self-organization with strong mathematical connections (Malsburg 1973). The most known scientist is T. Kohonen, who published several books on the *instar* and *outstar* learning methods, associative and correlation matrix memories, and – of course – self-organizing (feature) maps (SOFM or SOM) (Kohonen 1972; Kohonen 1984; Kohonen 2001). This neuron model has great impact on the whole spectrum of informatics: from the linguistic applications to the data mining.

The Kohonen's neuron model is commonly used in different classification applications, such as the unsupervised clustering of remotely sensed images. The paper of H.C. Sim and R.I. Damper gives a demonstration, how the SOM model suits for object matching purposes with images of tools under translation, rotation and scale invariant circumstances (Sim 1997).

The goal of automatic road detection is very clear in the paper of R. Ruskoné et al., who apply a two-level processing technique in combination of road segment extraction and a production net (Ruskoné 1997). A. Baumgartner et al. describe a context based automatic technique for road extraction (Baumgartner 1997), while S. Hinz and his colleagues developed a road extractor in urban areas (Hinz 2001). The research of P. Doucette et al. focuses on the simulated linear features and the detection of paved roads in classified hyperspectral images HYDICE with the use of Kohonen's SOM method (Doucette 1999; Doucette 2001).

2. SELF ORGANIZING NEURAL NETWORKS

The self-organizing feature map (SOFM) or self-organizing map (SOM) model is based on the unsupervised learning of the neurons organized in a regular lattice structure. The topology of the lattice is triangular, rectangular or hexagonal. The competitive neurons have a position (weight) vector of dimension *n*:

$$\mathbf{m} = \left[\mu_1, \mu_2, \ldots, \mu_n\right]^T \in \mathbb{R}^n \tag{1}$$

Furthermore the input data points have similar coordinate vector:

$$\mathbf{x} = \left[\xi_1, \xi_2, \ldots, \xi_n\right]^T \in \mathbb{R}^n \tag{2}$$

The learning algorithm consists of two blocks: the first is the rough weight modification, called *ordering*, the second one is the fine settings, called *tuning*. The iterative algorithm starts with the neuron competition: a winner neuron must be found by the evaluation of the following formula:

$$c = \arg\min_i \left\{ \left\| \mathbf{x} - \mathbf{m}_i \right\| \right\} \tag{3}$$

where $i = 1 \ldots q \in \mathbb{N}$, having *q* neurons, and $c \in \mathbb{N}$.

The second step of the learning algorithm is a weight update for epoch *t*+1 in a simple way:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(t)\left[\mathbf{x} - \mathbf{m}_i(t)\right] \tag{4}$$

where *t* means the epoch, and the coefficient function is defined as follows:

$$h_{ci}(t) = \begin{cases} \alpha(t) & \text{if } i \in N_c(t) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

This formula contains the speciality of the Kohonen model, namely the time dependent consideration of the neurons' neighborhood in form of $N_c(t)$. The neighborhood can be interpreted as concentric squares around the winner in case of a rectangular neuron lattice. If the neighboring neurons are in the acceptable region, their weights will be updated by a factor of $\alpha(t)$. For this function the limits are known: it must be between 0 and 1, mostly it decreases monotonically.

3. THE SONG MODEL

As the author of this paper applied the Kohonen type neuron model for road junctions, great anomalies were recognized. This was caused by the regularity of the grid structure, which was quite far from the graph-like road crossings.
The basic idea of the extension of the SOM model was to integrate the graph structure with its flexibility in form. The newly named *Self Organizing Neuron Graph* (shortly SONG) model has an undirected acyclic graph, where the computational elements are located in the graph nodes. The edges of the graph ensures the connections between the participating neurons. The learning algorithm of the SONG technique has kept the Kohonen rule for winner selection and weight update, but the neighborhood had to be formulated in a different way.
The connections (edges) of the graph are mostly given by the adjacency, commonly in form of a matrix. The elements of the matrix are defined after Equation 6:

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{if node i and j are connected} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

This matrix is theoretically a symmetric square matrix of $q \times q$. The nonzero matrix elements represent the direct relations between two neurons. The term adjacency can be extended in order to express more information than a binary "direct neighbor" or not. Therefore the graph algorithms are implemented, which derive the generalized adjacency matrix $\mathbf{A}^k$, where $k \le n$. This generalized adjacency matrix has only zero elements in the main diagonal (there are no loops), all the other values give the number of the intermediate edges between two nodes. The most simple generalization algorithms are the *matrix power technique* (Henley 1973; Swamy 1981) and the *Floyd-Warshall method* (Warshall 1962; Reinhard 2002).
The generalized adjacency matrix makes the modification of the learning coefficient formula for graphs possible:

$$h_{ci}(t) = \begin{cases} \alpha(t) & \text{if } \mathbf{A}_{ci}^k < d(t) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

In the above expression the condition is evaluated for the row of the winner neuron of the matrix.
As it was mentioned regarding the Kohonen model, there is an ordering and a tuning phase. These two computational blocks are inherited in the SONG technique, too. Because the graph adjacency is invariant during the run (the connections of the neurons are fixed), the adjacency matrix can be created prior the iterations (Barsi 2003a).
The calculations have been accelerated applying buffering around the neurons. Such buffers are to be interpreted as boundaries for the potential data points, so the distance computations and therefore the whole processor load can be limited. Figure 1 illustrates the generalized adjacency in gray shades and the buffers in case of a letter figured graph.

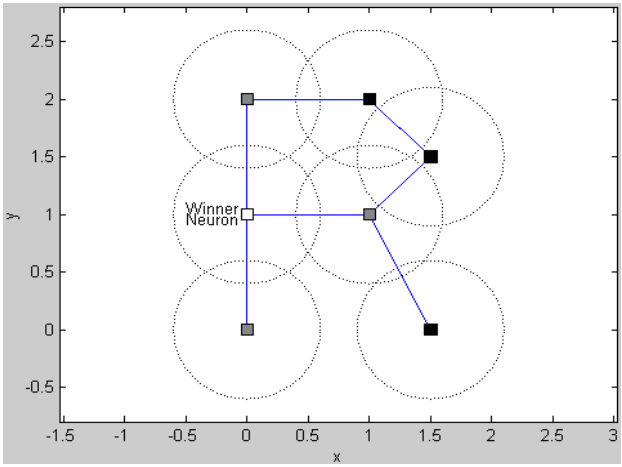


Figure 1. Neuron graph with buffers, where the neurons are colored by the generalized adjacency values

The described SONG model is based on the adjacency information. During the developments an other type has been created, which is based on the distance between the neurons. In the distance SONG model the generalized distance matrix $\mathbf{D}^n$ is used instead of $\mathbf{A}^k$, but this second algorithm is significantly slower because of the lack of the invariance of the distance matrix: the neuron weights (positions) are permanently changing in every iteration steps, so the distance must be modified (Barsi 2004). This results that the distance SONG method can only be applied in refinement of the positions of the adjacency SONG algorithm.

4. RESULTS

The application of the SONG technique for image analyzing tasks is discussed in three types of examples. The first example is a template matching use, where the previously given fiducial structure has to be matched. The complexity of the graphs is shown in the second group of experiments, where a given building structure has to be found and the structure of a complex building has to describe in form of a neuron graph. The last tests implement a running window-type application of the SONG algorithm: segments of a road network are detected by a kernel graph.

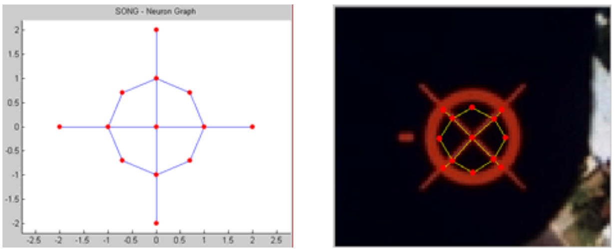
4.1. Fiducial detection

The interior orientation of an aerial image can be automated if the fiducial marks can be detected without human interaction. The camera manufacturers apply specific figures as fiducials, which have given geometry; they can be described even in graphs. The rough skeleton of the fiducial mark was drawn as a graph (Figure 2a).

In the first application a color Wild RC20 aerial camera image was dropped into RGB components, then the red channel was segmented by histogram threshold. The pixel coordinates of the binary image were the data points for the test run.

Because the fiducials of this camera type are in the corners of the images, only the small image corners were cut out and preprocessed. The result of the algorithm can be seen in Figure 2b.

The ordering algorithm had 100 epochs, the starting and end learning rates were 0.9 and 0.0. The adjacency distance has been decreased from 4 to 1 (direct neighborhood). The tuning had 1000 epochs, 0.1 starting learning rate and zero at the end, while the neighborhood was set back from 2 to 1.



a. initial graph structure b. final position

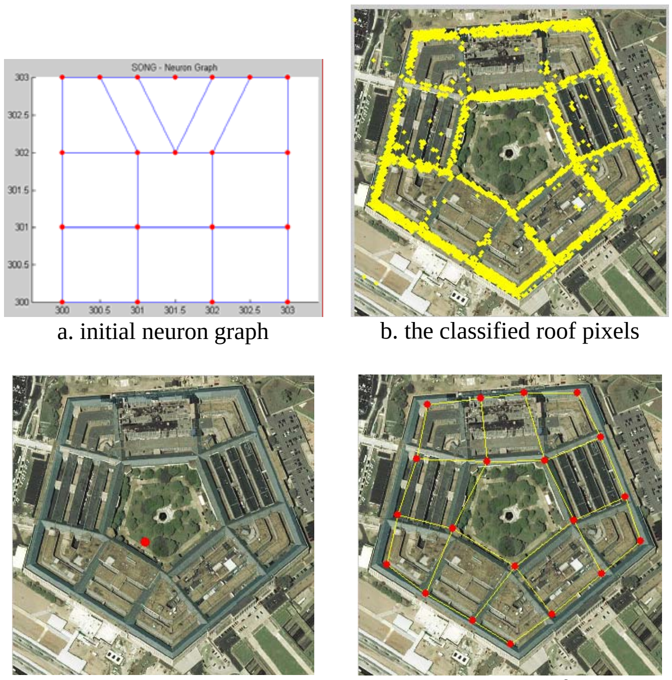
Figure 2. Search of fiducials by SONG.

4.2. Detecting building structure

The detection of man-made objects focuses very often on buildings. The SONG method was therefore tested in such tasks. There were two experiments executed: (1) the right position of a given structured building had to be found and (2) the structure of a given (positioned) building had to be detected.

The first test applied one of the first images taken about the Pentagon building in Washington DC after the attack on 11. September 2001. The image was captured by the QuickBird sensor with a ground resolution of 0.6 m. The initial neuron graph was given: the structure of the famous building is known (Figure 3a). The input data points were produced by a maximum likelihood classification of the color image pixels, for that training areas of two roof types were marked and used. The result of the image classification gave a binary image, where the true pixels were the elements of the roof (Figure 3b). The coordinates of these pixels were read out and fed into the SONG algorithm. The ordering phase of the algorithm had 300 epochs, the starting learning rate and neighborhood were 0.01 and 6, while the finishing state had values of 0 and 1 (direct neighborhood) respectively. In the starting step (Figure 3c) the 20 neurons of the graph were placed somewhere within the building, after the 10000 step long tuning (with a learning rate interval of 0.003 – 0 and strict direct neighborhood) the graph has found the building (Figure 3d). During the tuning phase the direct neighborhood ensured, that the neurons merely refined their geometrical positions instead of rough changes.

The iterative evaluation of the 3302 roof pixels took only a couple of minutes on an Intel PIII machine (Barsi 2003b).



a. initial neuron graph b. the classified roof pixels

c. the starting neuron graph position d. neuron graph after the finished iterations

Figure 3. The Pentagon test

The other building analyzing test used an another satellite image: a 1 m IKONOS image, taken from Singapore in August 2000. The input data set was established by a rule-based RGB pixel classification, focusing similarly on the roof. In order to get smaller training data set the identified points were resampled. The test applied four given graph structures having 11, 13, 19 and 21 neurons in the nodes (Figure 4).

Figure 4. Detecting building structure (Singapore) in IKONOS image – an intermediate state with 13 neuron graph

The four variations were quite similarly controlled: the ordering phase had about 200-600 epochs, the tuning had 500-3000. Learning rate was between 0.01 (ordering) to 0.00001 (tuning). The starting neighborhood was increased from 4 to 10 with the complexity grow of the neuron graph in the ordering; the counterpart tendency was noticed during the tuning with a decreased neighborhood from 2 to 1 (Barsi 2003c).

4.3. Detecting roads

The presented self-organization technique can be applied also in a running-window environment. In this case the necessary structuring (moving) element was a simple cross having 5 processing elements (neurons) and 4 connections. The test image was a 0.4 m ground resolution black-and-white orthophoto about the Frankfurt am Main region. The image cut-off was captured over a sparse village part near to the Frankfurt international airport.

The image preprocessing was solved by thresholding the intensities, so the further steps used binary data. The adequate parameterized technique identified the junctions as it is shown in Figure 5.

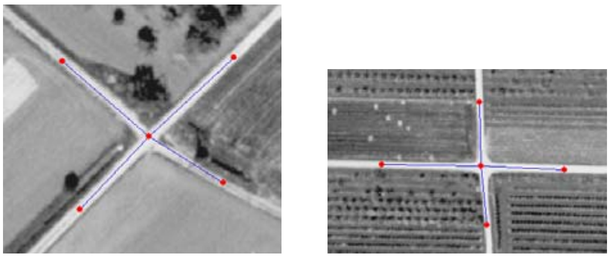


Figure 5. Recognizing a four-arm junctions by the cross kernel

The road detection approach has a new type of SONG application: the input image has been partitioned, then the previously defined and in the whole detection procedure fixed structuring element is evaluated by the SONG technique. The developed SONG algorithm was limited for the ordering phase; the tuning was not so important; the processing speed could be accelerated in this way. This means that the structuring element must be created and described before the run. The partitioning is carried out by creating regular square raster on the image. The applied structuring element was the same as in Figure 5, a 5 neuron cross with 4 connections. The adjacency matrix and the rule to build the initial neuron weights were constructed in the starting step.

The main question in the practical implementation was (1) to find the right size of the image partitioning and (2) to get the right parameter set for this special case. Both questions were answered after executing an experiment series with various size partitions and different SONG control parameters. The best result after visual evaluation was a partition size of 20 pixel by 20 pixel (8×8 m).

The control parameters have not too large variety, because the given neuron graph was not large; the highest neighborhood could be 2. The free parameters were the starting learning rate and the number of epochs. From earlier experience, these two controls have strong interaction, so during the tests both are varied: learning rates were between 1.0 and 0.5 and the number of epochs between 20 and 500.

The SOM and therefore (because of the inheritance) SONG have an observation: after a "critical" number of epochs, the system gets a stabile status, ie. the neuron weights don't change anymore. This critical epoch number was searched, and found at 50 for the 20×20 partitions. Then the effective starting learning rate was found at 1.0.

The essence of the algorithm for this running window style computing model is the following:

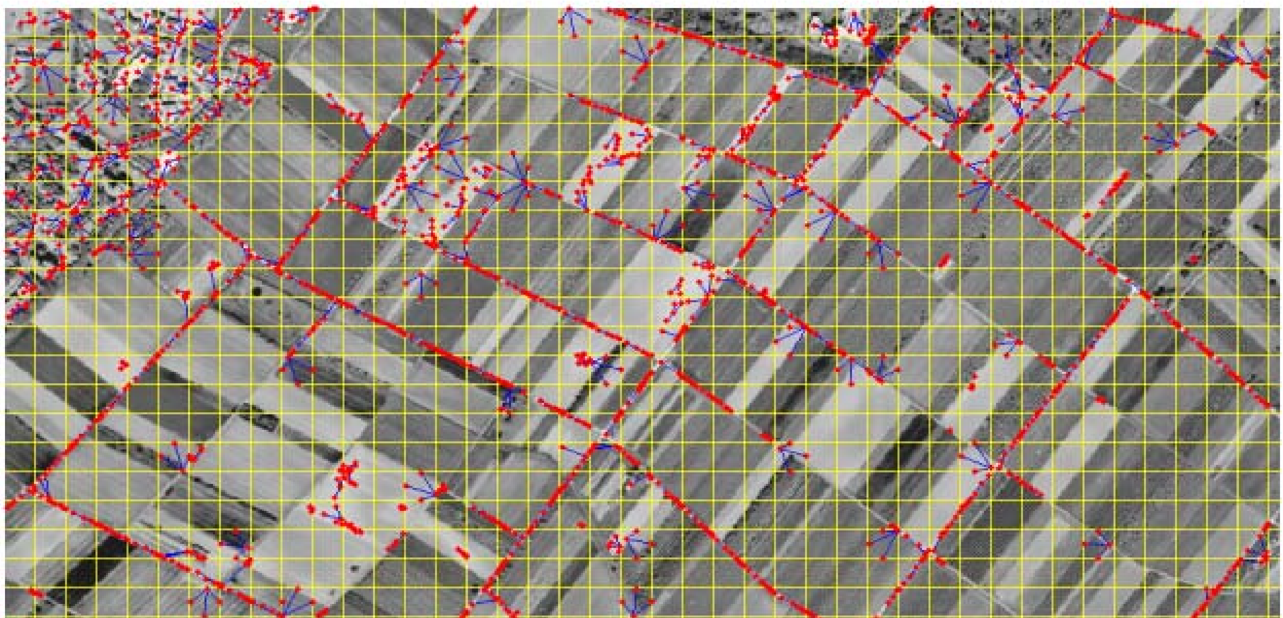


Figure 6. Road detection by running window SONG algorithm

```
for all partitions
  selecting the data points
  if the set is not empty
    reading pixel coordinates
    for all ordering epochs
      calculating learning rate and d(x)
      for all data points
        for all neurons
          calculating distance
        endfor
        for all neurons
          if in calculated neighborhood
            modifying weights
          endif
        endfor
      endfor
    endfor
  endif
endfor
```

The above algorithm has strong nested loop structure. The method is therefore computation intensive and sensitive for the loop control parameters. Fixing the number of epochs, neurons, partition size, the number of data points has also great influence on the performance. It was observed, that the smaller sized partitions resulted faster runs, than bigger partitions, which is not unexpected looking at the above algorithm.

The function d(x) is the permanently decreasing neighborhood.

The result of the image analysis can be seen in Figure 6, which shows the regular partitions, too. The structuring elements have detected visually most of the road segments and their junctions, only some small errors were occurred. These errors must be eliminated by implementing constrains for the final neuron graph structure during the processing, which is the topic of the current research.

5. CONCLUSION

The newly developed self organizing neuron graph is a flexible tool in analyzing different types of remotely sensed images. The template matching problem was solved by a given fiducial mark graph. Because the camera manufacturers have own figures for fiducials, the SONG algorithm is capable to fit not only a single neuron graph to the assumed image part containing the fiducial, but a whole series. Measuring the fitness for all matching, the highest fitness identifies furthermore the type of camera. In this way, the SONG technique recognizes the camera itself. The SONG matching is fast, it can provide an alternative solution in the automatic interior orientation.

The building and road detection belong to the object detection approaches, which are the most interesting in the modern digital photogrammetry. The building detection, if we have any preliminary hypothesis about the building structure, is a relational matching task. In this sense, the SONG method can help to realize similar solutions, which are mostly related to the relational matching. The shown example with the Pentagon building can be generalized: if one can describe the structure of a building in form of a graph, it can be found in an image using the developed self-organizing technique; furthermore this building can be "traced" in an aerial image strip or block. Only the same preprocessing operators (e.g. classifiers) must be executed, prior the SONG run.

The other presented building detection experiment proved that this method can be used to detect the structure of an unknown building by creating a hypothesis neuron graph and testing its suitability. The given example has shown that this hypothesis testing can be a way to improve the current version of the algorithm. The test was an alternative solution for getting the skeleton of an object, solved by the application of artificial intelligence instead of the known classical skeleton operator.

The most interesting test was the road detection. In this situation a black-and-white orthoimage ensured the necessary data points by simple thresholding, which is a very fast image processing technique. After the segmentation, the SONG algorithm has found the rough road structure. This can be interpreted in two ways.

Once these road segments are objects to perform further grouping methods to obtain real road network. In this way the creation of a classical GIS-type topology, then a follow-up topology analysis and restructuring can lead to the network.

The other possible application of the obtained results is their interpretation as first approximation of the road network. Using this philosophy, the connecting processing steps can be buffering the graph edges. With the fusion of the independent buffers, we will get a subsample of the image, where the probability of the roads in the image is relatively high. Thereafter the different checking algorithms can be executed, which test these image subsamples on containing roads. If the test is positive, the exact road segment and its geometric position can be detected using more accurate methods. In this checking the SONG technique can be involved in the complex algorithm, i.e. with also the tuning phase.

The result image of the road detection contains some noisy parts, especially in urban areas. The method should be used with a preliminary masked input image, which doesn't have any urban land covers. Other alternative would be to establish an urban parameter set, which can have better performance under built-in circumstances.

The last test points to a new possibility for using highly parallelized algorithms in photogrammetric image analysis, because after partitioning the input image, the same steps of the algorithm must be evaluated for each image parts. If a multi processor computing environment (e.g. dual Pentium PC, or even computer cluster or grid) is available, the method can be implemented and used.

The general assumption of the SONG technique is to own an adequate initial neuron graph. A nice initialization could be the use of an obsolete topographic map, where the novel method would be responsible to check the old map (database) content to the new image information. In this meaning the method suits also to the map update procedures.

The road detection test lasted in Matlab (interpreter type) environment about two minutes on an Intel P4 1.7 GHz machine, all the other are significantly faster thanks to the compiler (MS Visual C++) realization. The image size was 555×827 pixel. This fact also underlines the performance power of the method.

As the paper presented, the generalization of the original Kohonen-type SOM can be extended by graphs. The newly developed SONG method has been proved its capability in different photogrammetric and remote sensing tasks. The technique has shown how to cope with different types of tasks using the same algorithmic background. The method is important in the point of view of the artificial intelligence and neural networks, because the general suitability and applicability has also been proved.

ACKNOWLEDGEMENTS

The author expresses his thanks to the Alexander von Humboldt Foundation, when the work was started; to the Institute of Photogrammetry and GeoInformation, Hannover; to the Hungarian Higher Education Research Program (FKFP) for partly financing the research.

REFERENCES

Barsi, A., 2003a. Neural Self-Organization Using Graphs, *in: Machine Learning and Data Mining in Pattern Recognition, Lecture Notes in Computer Science,* Vol. 2734, Springer Verlag, Heidelberg, pp. 343-352

Barsi, A., 2003b. Graph Based Neural Self-Organization in Analyzing Remotely Sensed Images. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2003,* Toulouse, Vol. VI, pp. 3937-3939

Barsi, A., 2003c. Neural Self-Organization in Processing High-Resolution Image Data, *ISPRS-Earsel Joint Workshop High Resolution Mapping from Space 2003,* Hannover, p. 6

Barsi, A., 2004. Generalization of topology preserving maps: A graph approach. *International Joint Conference on Neural Networks*, Budapest, *accepted for publication*

Baumgartner, A., Eckstein, W., Mayer, H., Heipke, C., Ebner, H., 1997. Context-supported road extraction. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) Monte Verità*, Birkhäuser Verlag, Basel, pp. 299-308

Carpenter, G.A., 1989. Neural network models for pattern recognition and associative memory. *Neural Network*, No. 2, pp. 243-257

Doucette, P., Agouris, P., Musavi, M., Stefanidis, A., 1999. Automated extraction of linear features from aerial imagery using Kohonen learning and GIS data. in: P. Agouris – A. Stefanidis (eds): *Integrated Spatial Databases – Digital Images and GIS*, Portland, Lecture Notes in Computer Sciences 1737, Springer, Berlin, pp. 20-33

Doucette, P., Agouris, P., Stefanidis, A., Musavi, M., 2001. Self-organised clustering for road extraction in classified imagery. *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 55, No. 5-6, pp. 347-358

Henley, E.J., Williams, R.A., 1973. *Graph theory in modern engineering*, Academic Press, New York

Hinz, S., Baumgartner, A., Mayer, H., Wiedemann, C., Ebner, H., 2001. Road extraction focussing on urban areas. in: Baltsavias et al. (eds): *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*, Swets & Zeitlinger, Lisse, pp. 255-265

Kohonen, T., 1972. Correlation matrix memories. *IEEE Transactions on Computers*, Vol. 21, pp. 353-359

Kohonen, T., 1984. *Self-organization and associative memory*. Springer, Berlin

Kohonen, T., 2001. *Self-organizing maps*. Springer, Berlin

Malsburg, C. von der, 1973. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, No. 14, pp. 85-100

Reinhard, E., 2002. http://www.cs.ucf.edu/~reinhard/classes/cop3503/floyd.pdf

Rosenblatt, F., 1958. The perceptron. A probabilistic model for information storage and organization in the brain. *Psychological Review*, Vol. 65, pp. 386-408

Rojas, R., 1993. *Theorie der neuronalen Netze. Eine systematische Einführung*. Springer, Berlin

Ruskoné, R., Airault, S., 1997. Toward an automatic extraction of the road network by local interpretation of the scene. in: D. Fritsch – D. Hobbie (eds): *Photogrammetric Week*, Wichmann, Heidelberg

Sim, H.C., Damper, R.I., 1997. Two-dimensional object matching using Kohonen maps. *IEEE International Conference on Systems, Man and Cybernetics*, Orlando, Vol. 1, Part 1/5, pp. 620-625

Swamy, M.N.S., Thulasiraman, K., 1981. *Graphs, networks and algorithms*, Wiley, New York

Warshall, S., 1962. A theorem on Boolean matrices, *Journal of the ACM*, Vol. 9, No. 1, pp. 11-12