# Object Recognition in Image Sequences and Robust Associative Image Memory using the Multilevel Hypermap Architecture

Henning Hofmeister
Leibniz Institute for Neurobiology
Magdeburg, Germany
hofmeister@ifn-magdeburg.de

Bernd Brückner
Leibniz Institute for Neurobiology
Magdeburg, Germany
brueckner@ifn-magdeburg.de

## Abstract

*The introduced system for object recognition and tracking uses an associative memory for storing prototypes of objects. The Multilevel Hypermap Architecture (MHA), a self-organizing neural network approach, is used, to construct a robust system. To process form variant objects the MHA is extended to work with masked input data.*

*Because of using scaled input objects, the system is invariant to translation. The invariance to rotation is realized by the associative memory, which is able to learn different instances of the same input object.*

*In our tests we obtained a robust system behavior, because the associative memory is able to minimize disturbances in feature extraction with the learned and recalled features of an object prototype.*

## Keywords

Multilevel Hypermap Architecture, MHA, figure-ground-separation, image sequence analysis, object recognition, object tracking

## 1. Introduction

From the point of machine vision the key to interpret the real world is to extract object information from it. One of the problems is to separate the objects from the background and to find their real boundaries. This is a condition for automated object recognition and object tracking on long image sequences in image analysis systems. Such systems are from great interest, not only in the field of traffic and security.

Classical methods of signal and image analysis are successfully used to solve especially low dimensional problems. For high dimensional problems artificial neural networks become more important nowadays. These adaptations to the biological signal processing system (brain) try to control the complexity of recognition tasks. There is a wide range of usable neural network learning algorithms and of applications in the field of object recognition [1, 8, 13, 14, 16].

Our research in the field of image analysis pursues the goal first to adapt principles of biological vision to computational algorithms for improving machine vision and second of a large use of artificial neural networks. Especially we are dealing with Learning Vector Quantization (LVQ) [11] and Adaptive Resonance Theory (ART) [6].

Our development of the Multilevel Hypermap Architecture (MHA) led to applications in speech recognition, analysis of fMRI data sets and generation of hypotheses. ART we were using for classification of high dimensional data sets and color segmentation in images [9].

Both neural network types use self-learning algorithms. So they are qualified for automated systems.

In this paper an algorithm for figure-ground-separation in long image sequences is described and the ability of the MHA to store objects robustly as an associative memory is shown**.**

## 2. Image Sequence Analysis system

In this paper we present a system for analyzing monocular image sequences, which autonomously describes and stores complex objects in an associative memory structure.

The structure of the technical system is motivated by a simplified biological model of vision.

The processing of visual stimuli takes place in two functional paths with clear anatomical differentiation. The magnocellular "where"-path analyses simple form and motion parameters, has high contrast sensitivity, transient responses and lack of overt wavelength selectivity. The parvocellular "what"-path extracts features like color, colorbased forms or texture by having lower contrast sensitivity, sustained responses and pronounced wavelength selectivity [7]. The striate cortex V1 has a modular structure and each module is capable of analyzing the pattern, wavelength, luminance, movement and depth of stimuli appearing in different portions of visual space [15]. In the higher areas of visual cortex like medio- and inferotemporal cortex (MT, IT) and V4 topographical context will be lost and perception and recognition of a stimulus happens. MT analyses motion and depth, IT is responsible for the recognition of complex forms, like faces, and V4 for the experience of color. Object features described by the visual stimuli are distributed spatially over the matrix of feedback connections of the neurons in the brain [10].

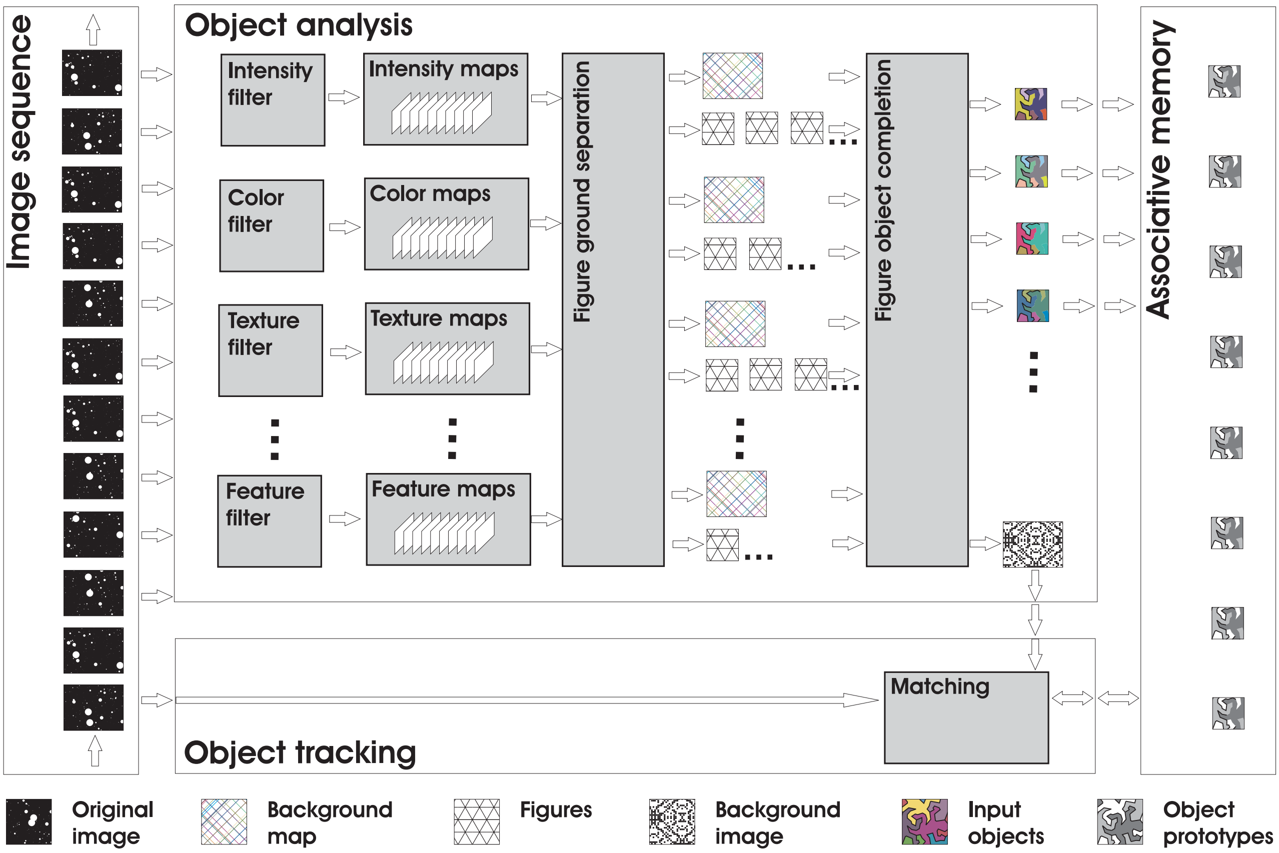

**Figure 1: System model**

Inspired by the biological model we developed the following technical analysis system shown in Fig. 1 to recognize and track unknown and form variant objects. As input we use a monocular sequence of trichromatic images (RGB) from a stationary camera. Over a time of *n* images a stack is created representing the history of the object movement. The history is necessary for the suggested figure-ground-separation. The first step is a feature extraction. The filter modules extract in parallel different features from each image of the stack. So a stack of feature maps is generated. After that figure-ground-separation for each feature take place resulting in a set of featured figures from the object and a map from the background. For further processing the figures must be scaled to the same size. They are superimposed by a map, so that pixels not belonging to the object are marked. For improving the robustness of the object extraction it is meaningful to combine the separation results of several filter modules. Therefore next step is an object completion and we get as result scaled and masked images from the objects, which occurred in the sequence. The contour of the input objects for the associative memory follows the contours of the moving objects in the image sequence. Background pixels are marked as such. For robust learning of an input object we suggest a multi-level scaling like in cortical neurons coding differently large ranges of visual space. So it is possible to get an input object with different spatial scales by a determined size and to get a more or less detailed view on the objects. Therefore the input objects are translation invariant.

The associated memory is able to learn the input objects unsupervised and creates prototypes. Each prototype represents a moving object from the image sequence by all the features analyzed by the filter modules.

For an easy object tracking we suggest to locate the objects by a matching algorithm between the prototypes and the figures from the current image of the sequence separated from the ground. Conditions for the object tracking are prototypes and a background image. But after initial analyzing of the first few images of the sequence the system fulfils these conditions. Then the object recognition and tracking tasks are running in parallel.

In the proposed application we use long sequences of 576*720 images. For a three-level spatial scaling input objects of sizes 4x4, 16x16 and 64x64 are created.

## 3. Figure ground separation

By image acquisition with a stationary camera we can assume stable background pixels and varying pixels over a sequence if changes happen. So, for detection of moving objects the pixel difference between successive images is usable. In the analysis system the history (a number of previous images) is considered and for each pixel a vector is created only from the stable feature values of the considered sequence, which describe the background. Changing feature values are faded out. In Fig. 2 an intensity map of a traffic scene and the background map belonging to it are shown. The background is calculated from a sequence of twenty images. The three moving objects on the road (two cars and a cyclist) are suppressed.

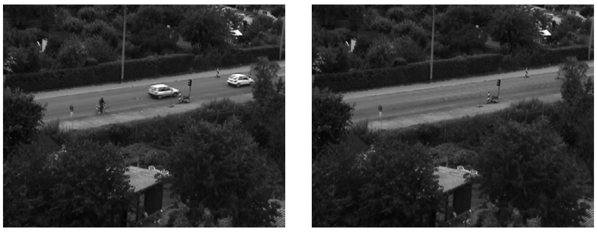

**Figure 2: Intensity map and background map**

The figures of the moving objects are selected by subtraction the calculated background map from the current feature map. Figures are separated by labeling connected pixels and each of them is scaled to the size of the input object images (4x4, 16x16, and 64x64). In these images all pixels not belonging to the object are marked with a special value. So the objects are masked by their real contour boundaries. As result of the figure-ground-separation we get a background map for each filter module (see Fig.1, 2) and a set of three images for each figure of a moving object (Fig. 3).

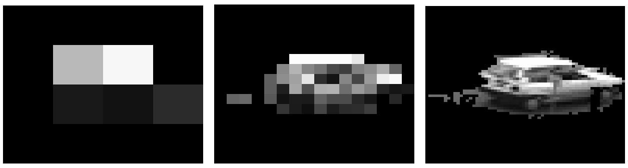

**Figure 3: Masked figures of an object in 3 scales (4x4 pixel, 16x16 pixel, 64x64 pixel)**

## 4. Associative Memory

### 4.1 The Multilevel Hypermap Architecture

One type of Learning Vector Quantization (LVQ) is the Hypermap principle introduced by Kohonen [11]. This principle can be applied to both LVQ and SOM algorithms. In the Hypermap the input pattern is recognized in several separate phases: the recognition of the context around the pattern to select a subset of nodes is followed by a restricted recognition in this subset. This architecture speeds up searching in very large maps and may carry out stabilizing effects, especially if different inputs have very different dynamic ranges and time constants [2].

The modification and extension of the Hypermap, the Multilevel Hypermap Architecture (MHA), are described in [2], the system model is shown in Fig. 4.

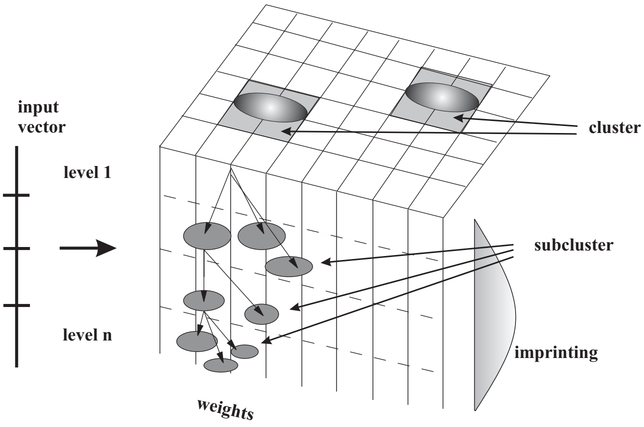

**Figure 4: The Multilevel Hypermap Architecture (MHA)**

Instead of two levels proposed in the Hypermap [12], the data and the context level, the MHA supports several levels of data relationship and a hierarchical unsupervised clustering. Therefore the input vector consists also of an arbitrary number of levels. In the MHA there is the same number of levels in the weight vector of each unit and these levels are related to the corresponding levels of the input vector. A varying number of levels for the units of the map is supported.

The MHA is trained with the different levels of the input vector whose representation is a hierarchy of encapsulated subsets of units, the so called clusters and subclusters, which define different generalized stages of classification.

Classification is achieved by finding the best matching node for each level of the hierarchy and by determining the square mean error of matching. In principle the algorithm handles different numbers of levels in the input vector.

One advantage of the MHA is the storage of hierarchical relationships of data. This will be useful for the generation of hypothetical relationships, i.e. relations that are not trained by input data. The MHA find it by itself by analyzing trained data.

By means of MHA it is possible to analyze structured or hierarchical data, i.e.

- o data with priorities, e.g. projection of hierarchical data structures in data bases
- o data with context (data bases, associative memories)
- o time series, e.g. speech, moving objects
- o data with varying degrees of exactness, e.g. sequences of measured data.

One advantage of the MHA is the support for both, the classification of data and the projection of the structure in one unified map. The resulting hierarchy has some redundancy like in biological systems.

An overview of our last works about MHA and further details on the algorithm gives [2, 4, 5]. Also in the previous years some real world applications using the MHA were reported in the literature [3, 5].

### 4.2 Adaptation for an Associative Image Memory

To use the MHA for the described object tracking system some extensions and definitions are needed. First of all the MHA is now able to learn masked data. With this feature the scaled and masked input objects (see Fig. 5) are learned. The different input objects (images) are assigned to the levels of the MHA, especially the 4x4 image is assigned to the 1. level, the 16x16 image to the 2. level and so on.

Therefore we get a generalization depending on the resolution of the image. The better the resolution, the more disturbances are expected and the higher is the number of the used level of the MHA (obviously with less generalizing effects). With learning the input objects the MHA creates prototypes of objects. These prototypes are used from the tracking system to find and track an object in a scene. The associative memory is unsupervised and self-organized, i.e. learning and recall of learned objects happened at the same time.

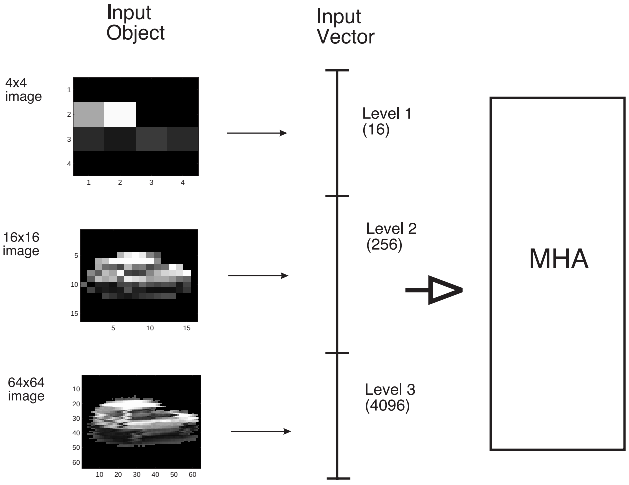

**Figure 5: The scaled input objects are transformed into the input vector for training the MHA. The different image resolutions are assigned to the corresponding levels of the input vector (in summary 4368 elements).**

To be invariant to the rotation of an object the associative memory (MHA) builds instances of the object prototypes, which are all learned views of an input object. Because of the hierarchical structure of the MHA similar instances are clustered, i.e. are assigned to the object prototype.

In the matching process of the object tracking the recalled object prototypes are combined with the background image and labeled as known objects. For these known objects the tracking is processed.

## 5. Results

The ability of the MHA as an associative memory should be demonstrated by the results of a simulation experiment of the suggested object recognition system.

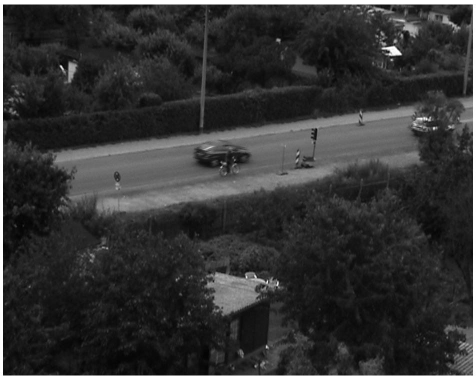

**Figure 6: Image 36 from traffic sequence**

A video sequence from a camcorder with 67 images was presented to the system. The first twenty images were used for creating the initial background image for the figure-ground-separation. From the following 47 images the object figures were separated and the input objects were created. Also the background image was updated continuously. In the sequence there were five moving objects (four cars and a cyclist). A MHA with 25x25 nodes was trained with the input object images. In the traffic sequence (one image from it is shown in Fig. 6) we observed a disturbance. The object of the cyclist (in the center of the image in Fig. 6) covers partially the object of the car. The effect on the prototype is shown below.

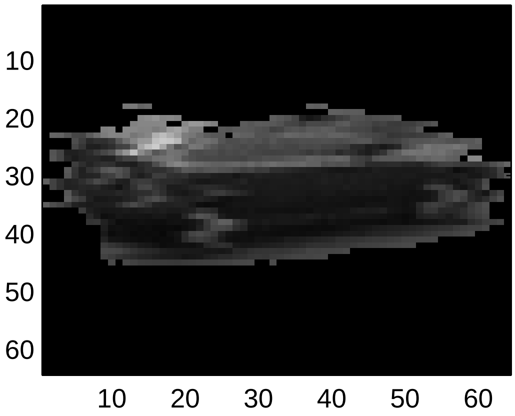

**Figure 7: 64x64 image of input object**

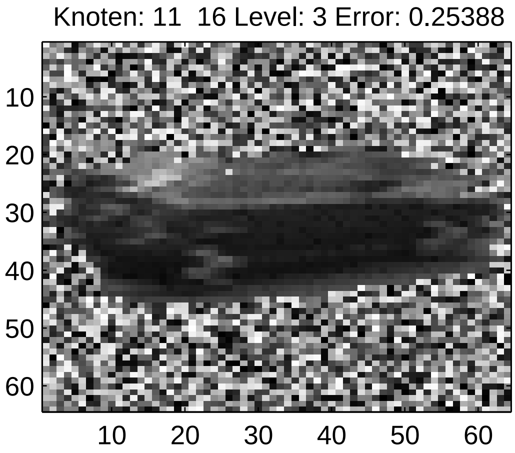

**Figure 8: 64x64 image of object prototype**

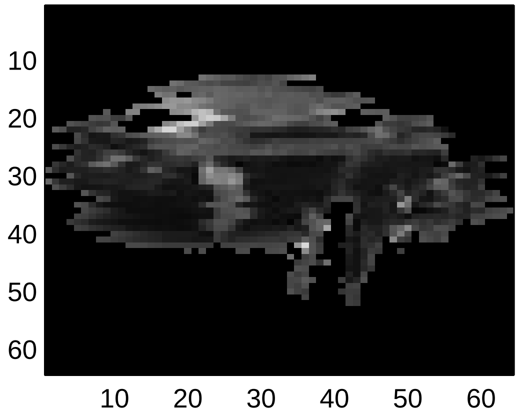

**Figure 9: 64x64 image of disturbed input object**

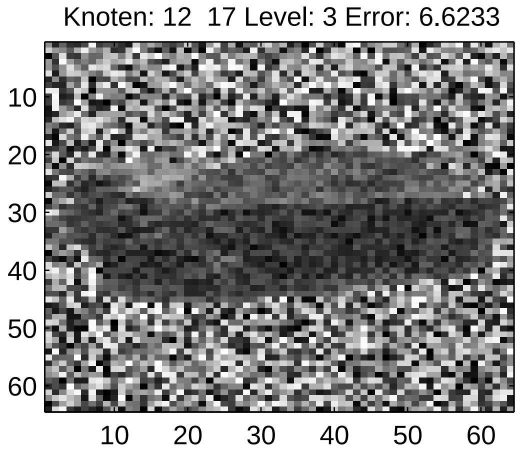

**Figure 10: 64x64 image of disturbed object prototype**

Fig. 7 shows the 64x64 feature image of the input object from an image of the sequence before the disturbance occurs.

We use here the 64x64 images for presentation, because they are better to view then the smaller images. The results of the first two levels of the MHA are comparable to the results of level 3, but with a higher significance (generalization effect of the MHA's hierarchical structure).

Fig. 8 shows the level 3 prototype of the object 'car' generated by the MHA. It's clearly to be seen that only the masked object is learned. Pixels from the surrounding lead to coincident values. The shown prototype is represented by node {11,16} from the 25x25 MHA.

Fig. 9 shows the input object created from the car covered by the cyclist resulting from image 36 (see also Fig. 6).

In Fig. 10 the result of the recall for the disturbed input object (from Fig. 9) is shown. The resulted prototype is represented by node {12,17}. This means that the prototype of the disturbed object belongs to the same cluster like the undisturbed. The MHA has learned an instance of the object. The similarity of both recalls is obviously comparing Fig. 8 and Fig. 10, and the generalization effect of the MHA, too.

In spite of the disturbance by another object the tracking continuous uninterruptedly. So the robustness of the associative memory stabilizes object tracking.

## 6. Summary and Conclusions

In the last years we were dealing with block matching for motion detection. Because of problems by using it for separating objects from a scene we are now using a new figure-ground-separation algorithm for motion detection.

To construct a robust system, the object tracking is combined with an associative memory, where prototypes of objects are stored. The associative memory is an adaptation of the Multilevel Hypermap Architecture (MHA), a self-organizing neural network approach. To process form variant objects, which are constructed as masked images, the MHA is extended to work with masked input data.

The introduced system for object recognition and tracking is invariant to translation because of using scaled input objects. Different image resolutions of object features are used for classifying and storing objects in the associative memory in form of prototypes.

The system is invariant to rotation, because the associative memory is able to learn different instances of the same input object and the hierarchical structure guarantees the consistency of all instances of an object prototype.

In our tests we obtained a robust system behavior, because the associative memory is able to minimize disturbances in feature extraction (e.g. shadow or partly covered objects) with the learned and recalled features of an object prototype.

Further works are going to implement more feature filters in the system, especially for texture features of the input objects. Also improvements in the history mechanism of the scene analysis and optimization of speed are planned.

## Acknowledgements

This work was supported by the LSA grant 3133B/0089B.

## References

[1] Bhandarkar, S.M., Koh, J., and Suk, M., Multi-scale Image Segmentation using a Hierarchical Self-Organizing Feature Map. *Neurocomputing*, **vol. 14**, No. 3, pp. 241-272 (1997).

[2] Brückner, B., Improvements in the Analysis of Structured Data with the Multilevel Hypermap Architecture. In: Kasabov et.al.: Progress in Connectionist-Based Information Systems, vol. 1, Springer-Verlag, Singapore, 342-345 (1997).

[3] Brückner, B., Gaschler-Markefski, B., Hofmeister, H., and Scheich, H., Detection of Non-Stationarities in Functional MRI Data Sets using the Multilevel Hypermap Architecture. In: Proceedings of the IJCNN'99, Washington, pp. 1–4 (on CD), ISBN: 0-7803-5532-6 (1999).

[4] Brückner, B., and Hofmeister, H., Generating Hypotheses Using the Multilevel Hypermap Architecture. In: Proceedings of the ICONIP2001, Fudan University Press, Vol. 1, pp. 58-62, Shanghai (2001).

[5] Brückner, B., and Wesarg, T., Modeling Speech Processing and Recognition in the Auditory System using the Multilevel Hypermap Architecture. In: Seiffert, U. and Jain, L.C. (Eds.), Self-Organizing Neural Networks. Recent Advances and Applications, Springer Series on Studies in Fuzziness and Soft Computing, Vol. 78, Springer-Verlag, Heidelberg, pp. 145-164 (2001).

[6] Carpenter, G.A., Grossberg, S., A Massively Parallel Architecture for a Selforganizing Neural Pattern Recognition Machine. *Computer Vision, Graphics, and Image Processing*, **vol. 37**, pp. 54-115 (1987)

[7] De Yoe, E.A., and van Essen, D.C., Concurrent processing streams in monkey visual cortex. *Trends Neurosci.*, **11**, 219-226 (1988).

[8] Grossberg, S., and Wyse, L., A neural network architecture for figure ground separation of connected scenic figures. *Neural Networks*, **vol. 4**, pp. 723-742 (1991).

[9] Hofmeister, H., Brückner, B., Color Classification to Improve Block Based Motion Estimation in RGB-Image Sequences. Proceedings of 6th ICONIP 1999, vol. 3, pp.1224-1229 (1999)

[10] Kohonen, T., Self-Organization and Associative Memory (Third edition). Springer-Verlag, Berlin (1989).

[11] Kohonen, T., Selforganizing Maps (Second Edition). Springer-Verlag, Berlin (1997).

[12] Kohonen, T., The hypermap architecture. In: Kohonen et. al.: Artificial Neural Networks, Elsevier Science Publishers, Helsinki, 1357-1360 (1991).

[13] Konen, W.K., Maurer, T., and von der Malsburg, C., A Fast Dynamic Link Matching Algorithm for Invarainat Pattern Recognition. *Neural Networks*, **vol. 7**, Nos. 6/7, pp. 1019-1030 (1994).

[14] Mel, B.W., SEEMORE: Combining Color, Shape, and Texture Histogramming in a Neurally Inspired Apporach to Visual Object Recognition. *Neural Computation*, **9**, 777-804 (1997).

[15] Schiller, P.H., Striate cortex. In: Encyclopedia of Neuroscience, ed. Adelman, G., Birkhäuser, Boston Basel Stuttgart, pp. 1148-1149 (1987).

[16] Wöhler, C., and Anlauf, J.K., An Adaptable Time Delay Neural Network Algorithm for Image Sequence Analysis. *IEEE Transactions on Neural Networks*, **vol. 10**, no. 6, pp. 1531-1536 (1999).