ICANN '93

Proceedings of the International Conference
on Artificial Neural Networks
Amsterdam, The Netherlands
13–16 September 1993

Edited by
Stan Gielen and Bert Kappen

Springer-Verlag
London Berlin Heidelberg New York
Paris Tokyo Hong Kong
Barcelona Budapest

A Distributed Multicolumnar System for Primary Cortical Analysis of Real-World Scenes*

T. Pomierski, H. M. Gross, D. Wendt

Technical University of Ilmenau
Div. of Neuroinformatics
D-98684 Ilmenau, P.O.Box 327, Germany
e-mail: pomi@informatik.tu-ilmenau.de

Abstract

In this paper we present a distributed multicolumnar system using an intracolumnar principal component analysis (PCA) for a topology preserving mapping of real-world grey level distributions within a two-dimensional intercolumnar Kohonen Feature Map. A two-stage principal component analysis within each processing column allows a similarity preserving description with only a few highly effective fitting parameters suited for a local translation invariant processing.

1. Introduction

Cognitive abilities required for analysis or interpretation of real-world scenes can not be explained by pattern matching completely in parallel. Instead of this a controlled decomposition of a highly parallel and complex visual scene into a sequence of lower dimensional components (meaningful pieces) is more probable. Hereby both preattentive or data-driven and attentive vision mechanisms based on internal system knowledge are of decisive importance for controlling this decomposition process [1]. The regions of a visual scene that are of high interest for an active vision system because of their syntactical meaning (preattentive scene analysis) or because of data-driven activated internal hypothesis about the scene or single components (attentive or knowledge-based vision) will be referred to this paper as internal *regions of attention*. Central point of this paper is to propose a neural network architecture for distributed parallel analysis of internal *regions of attention* selected within a visual scene by the special attention mechanisms mentioned above. In the presented model concept the distributed analysis is realized by an array of cortical processing columns having a structural and functional similarity with the minicolumns localized at the primary visual cortex [7]. This columnar array analyzes parallel the *regions of attention* of the real-world input scene selected by preattentive or attentive mechanisms which are not subject of this paper. Each processing column is able to detect essential input features within the corresponding *analyzing field* inside the *region of attention* on the basis of complex receptive fields. These receptive fields can be structured unsupervised by an adaptive self-organizing process based on a neural motivated principal component analysis (PCA). Oja [3] demonstrated that a particular version of the Hebb rule leads to a synaptic weight vector that has a strong similarity with the principal component of the set of input vectors. As extension according to Sanger [5] the self-organization of principal components corresponding to the largest eigenvalues for input data sets selected randomly from various real-world scenes leads to sets of principal component arrays or receptive fields nearly identical in quality (see Fig. 1). These results lead us to the supposition, that the same sets of complex receptive fields for local input analysis are available to each cortical column independent of its localization within the visual cortex as well as within our simple model architecture. The focus of this paper is to present a new extended neural network approach that enables a local topology preserving principal component analysis within each processing column that guarantees a locally translation invariant description of the local *field of analysis* with the lowest rate of describing parameters. Based on all fitting parameters in all processing columns a mapping into a two-dimensional intercolumnar organized Kohonen Feature Map becomes possible.

*Supported by Ministry of Research and Technology (BMFT), Grant No. 413-5839-01 IN 101D - NAMOS-Project

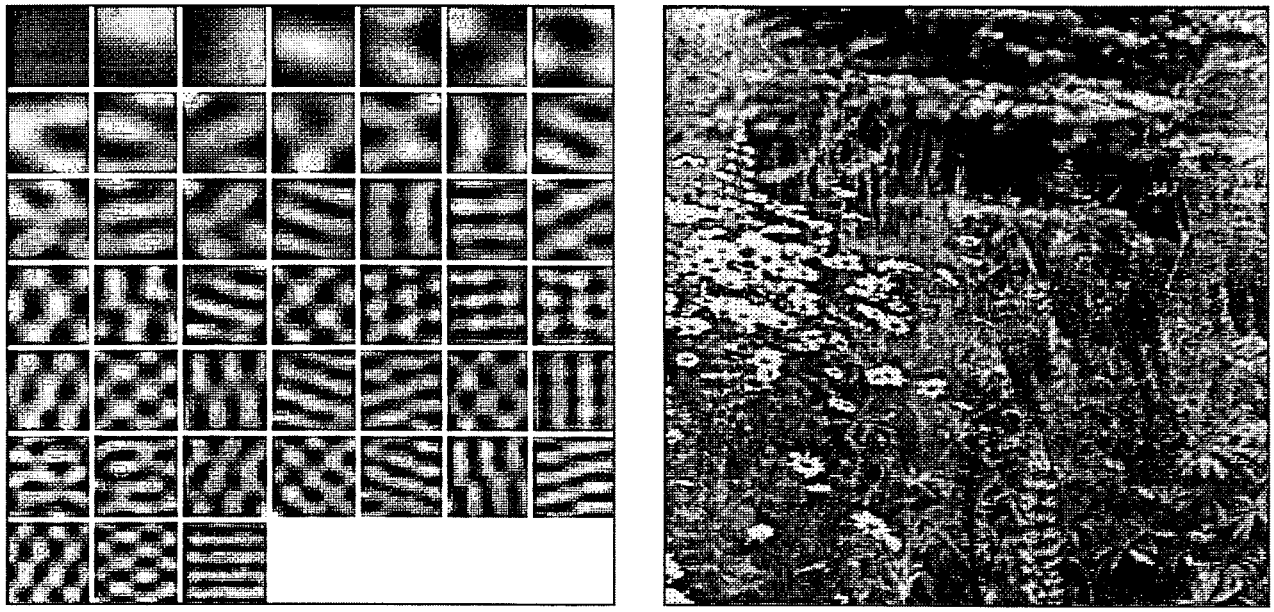

Figure 1: *(left) Self-organized receptive fields (principal component arrays)* $\mathbf{w}^{(1)}, ..., \mathbf{w}^{(i)}, ..., \mathbf{w}^{(n)}$ *structured in an unsupervised learning process by randomly selected input samples out of the shown real-world scene (right). This set of receptive fields is implemented within each processing column of the columnar array (for further explanation see text).*

2. Principal Component Analysis for Self-Organization of Receptive Fields

In the training period numerous images were obtained by recording real-world scenes and discretizing them to 512 pixel square images with 256 grey levels. No attempt was made to correct any optical irregularities. Local training patterns of 16 x 16 pixels were obtained by choosing an area within the image at random. Each training vector was normalized only to interval <0, 1>. This was possible because of the conformity of mean point vector $\mathbf{w}^{(0)}$ and first eigenvector $\mathbf{w}^{(1)}$ of real-world scene data distributions (see Fig. 3). So it is justifiable to indicate real-world scenes as ones without mean value. This is the prerequisite for adaptive unsupervised learning without a-priori knowledge about the data distribution of real-world scenes. For our simulations we used a single-unit rule like that proposed by Oja [3] in 1982.

$$w_j^{(1)}(t+1) = w_j^{(1)}(t) + \gamma(t)\, y(t)\, [x_j(t) - y(t)\, w_j^{(1)}(t)] \tag{1}$$

Here $\mathbf{x} = (x_1, ..., x_j, ..., x_m)^T$ denotes the real valued input normalized to the interval <0, 1>, y is the output signal, $w_j^{(1)}$ is the weight from input unit x_j to the single output unit y, and γ is the learning rate. This rule can be shown to produce a weight vector $\mathbf{w}^{(1)}$ corresponding to that eigenvector of the correlation matrix of all the inputs $\mathbf{x}$ which has a maximal eigenvalue. It extracts the principal component of the input data. The weight vector also tends to unit length. This rule was extended to i output units extracting the principal components in ascending sequence [5]. The learning process consists of two steps for each neuron i. First all parts of the input vector $\mathbf{x}^{(i-1)}$ already representable by weight vector $\mathbf{w}^{(i)}$ are subtracted.

$$\mathbf{x}^{(i)}(t) = \mathbf{x}^{(i-1)}(t) - y_i(t)\mathbf{w}^{(i)}(t),\ \mathbf{x}^{(0)} = \mathbf{x},\ i = 1, 2, ..., n \tag{2}$$

Afterwards the adaptation of all components $w_j^{(i)}$ of weight vector $\mathbf{w}^{(i)}$ takes place.

$$w_j^{(i)}(t+1) = w_j^{(i)}(t) + \gamma_i(t)\, y_i(t)\, x_j^{(i)}(t),\ i = 1, 2, ..., n,\ j = 1, 2, ..., m \tag{3}$$

The number of output units was determined by experiments showing that in context of our desired application 45 principal components are sufficient for extracting the essential information out of

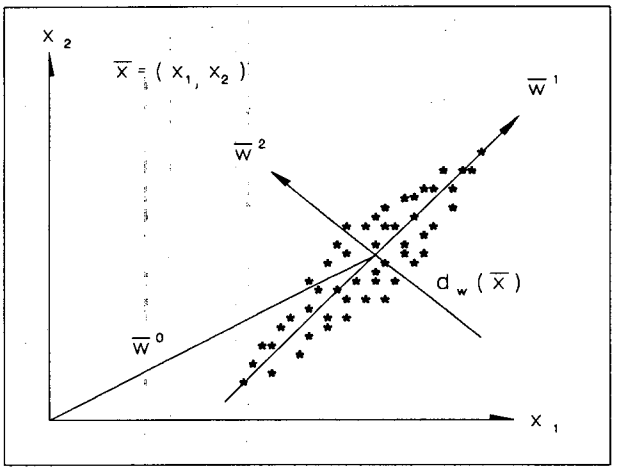

Figure 2: *Principal depiction of the possible correlation of two variables x_1 and x_2. The mean of the data distribution is not equal to zero. Accordingly, first it is necessary to extract the mean before determining a more favorable system of perpendicular coordinates.*

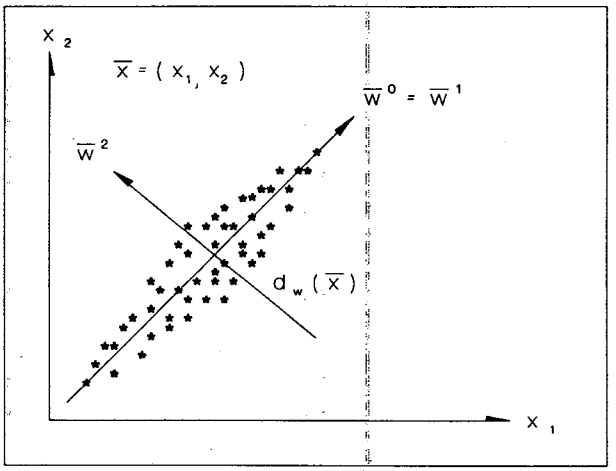

Figure 3: *Principal depiction of randomly selected pairs of grey-valued pixels x_1 and x_2 situated side by side within the visual scene in Fig. 1. The mean point vector $\mathbf{w}^{(0)}$ of the data distribution and the eigenvector corresponding to the largest eigenvalue $\mathbf{w}^{(1)}$ are identical here (for more explanations see text).*

16 × 16 square samples taken from real-world scenes (17% of all possible eigenvectors). The net was initialized by setting all the weights $w_j^{(i)}$ to small random values so that the square sum of the components of each weight vector $\mathbf{w}^{(i)}$ was approximately unity. Training consisted of applying randomly selected inputs and updating the weights. We used a total of 50.000 presentations. Because of the output feed-back to the input neurons:

$$\mathbf{x}^{(i)} = \mathbf{x}^{(i-1)} - y_i \mathbf{w}^{(i)}, \; i = 1, 2, ..., n \tag{4}$$

any weight modification of neuron i has consequences for neuron $i+1$. So it is necessary to give the weight vector $\mathbf{w}^{(i)}$ of neuron i the chance to stabilize before weight vector $\mathbf{w}^{(i+1)}$ of neuron $i+1$ gets able to learn effectively (see Fig. 4). That is the reason why convergence of weights is assisted by gradually reducing the learning rate in following way:

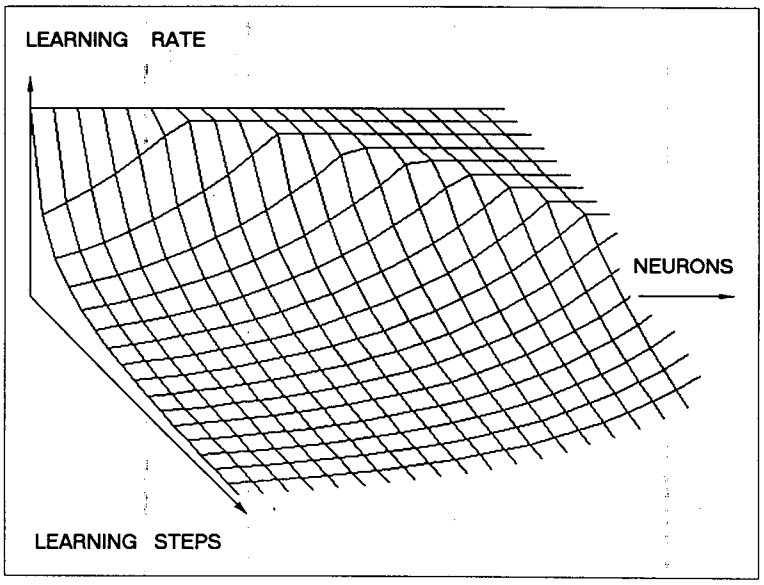

Figure 4: *Gradually reducing learning rate for neurons i. The mathematical description of the time varying and neuron dependent learning rate is presented in text.*

$$\gamma_0(t) = \frac{a}{t+1} \tag{5}$$

$$\gamma_i(t) = \gamma_{i-1}(t) \; v, \; i = 1, 2, ..., n \tag{6}$$

Here $\gamma = (\gamma_1, ..., \gamma_i, ..., \gamma_n)^T$ is the learning rate for the neurons $\mathbf{n} = (n_1, ..., n_i, ..., n_n)^T$, v the threshold parameter, t the learning step and a the starting value. This leads to a weight stabilization of all 45 neurons after 50.000 learning steps (see Fig. 1). This principal component analysis evoked by visual stimulation of real world and resulting in forming complex receptive fields is comparable in the broadest sense with the development of specialized receptive fields in the kitten visual cortex [6].

The ability of the self-organized receptive field set to extract nearly complete the information of any real-world scene is shown in Fig. 5. To code the images, each scene was segmented in $m \times m$ parts of $n \times n$ pixels without overlap. Each segment in the scene was analyzed by the complete receptive field set learned on the basis of a completely different visual scene (Fig. 1, right). In this way for each segment 45 fitting values for coding are determined. Fig. 5 shows the two scenes after reconstruction from the $m \times m \times 45$ fitting values.

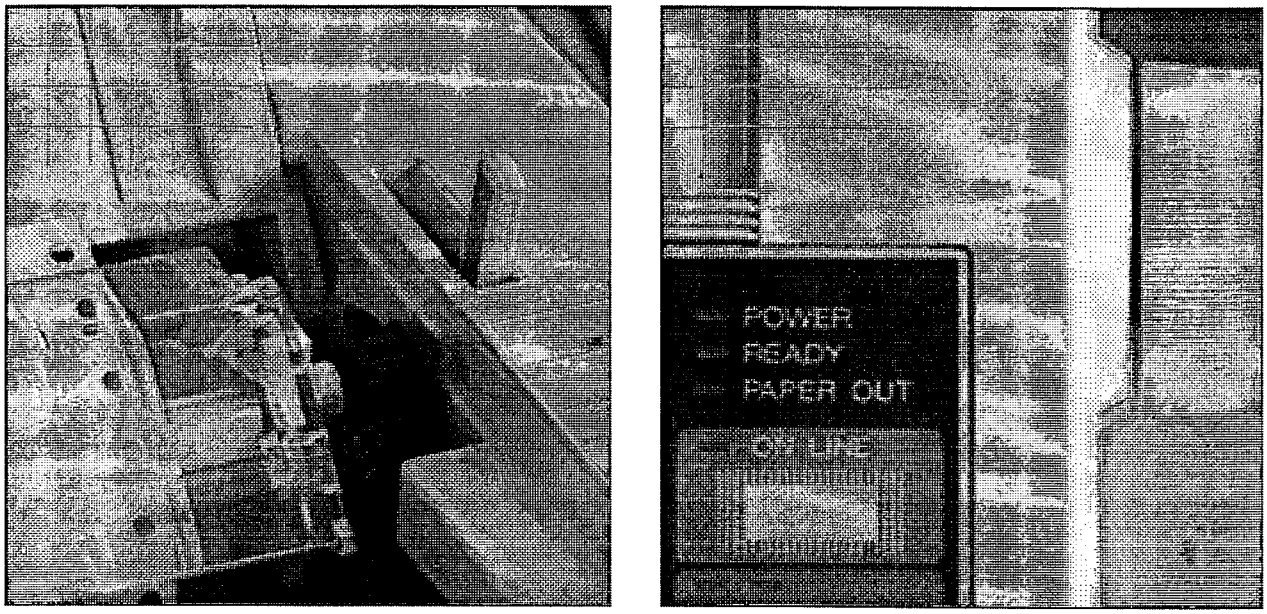

Figure 5: *Two views of a printer are shown after reconstruction from fitting parameters derived from the set of receptive fields shown in Fig. 1. These principal components have been self-organized on the basis of the real-world scene shown at the right hand side of Fig. 1. Both reconstruction test scenes shown here have been first segmented and analyzed by the receptive fields of Fig. 1. and then reconstructed from 45 fitting values for each segment. It is not possible to find essential differences between the original and the reconstructed scenes, so it is justified to present only the reconstruction results.*

3. Intracolumnar Processing of Local Receptive Fields

In our model architecture (see Fig. 6) each processing column analyzes a definite local domain of the focus region, the so called *region of analysis*, by complex receptive fields self-organized by the proposed neural learning mechanism. These *regions of analysis* enclose in our model 32×32 square pixels for each column. Because of this field dimension 16×16 processing columns analyze parallel a *region of attention* of 152×152 pixels, as the overlapping between regions of analysis has been chosen to 75%. Presupposing grey-level distributions of the real-world scenes as additive superposition of periodic process parts, a convolution of the self-organized receptive fields (16×16 pixels) with the *regions of analysis* (32×32 pixels) is in accordance with an analysis of the period determined by the receptive field of first order $\mathbf{w}^{(1)}$. The motivation for this convolution is the following: There were no internal relations detectable between fitting vectors for *analyzing regions* shifted inside the scene by only one pixel. The convolution presents a way to escape from this dilemma. By equalizing local translations similar fitting vectors for similar grey-level distributions are generated at the same time. Naturally, this is connected with a loss of precision, and in addition, this represents an irreversible process of analysis. Each of the 16×16 columns determines the average of activation for all neurons with the same receptive fields situated within its *region of analysis.*

$$f_i = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} || \mathbf{w}^{(i)} \mathbf{x}_{kl} ||, \; i = 1, 2, ..., 45 \tag{7}$$

In this way for each column an averaged 45-dimensional vector $\mathbf{f}$ of fitting values can be determined, which describes the mean conformity of the complex receptive fields with their *regions of analysis*. Here $\mathbf{f} = (f_1, ..., f_i, ..., f_{45})^T$ denotes the averaged fitting vector, $\mathbf{w}^{(1)}, ..., \mathbf{w}^{(i)}, ..., \mathbf{w}^{(45)}$ denote the two-dimensional arrays of principal components (complex receptive fields), and $\mathbf{x}$ is the real-valued input of the *analyzing region*. The 1024 grey values of one *region of analysis* are reduced to 45 real fitting values, this is a compression rate of 95%. The first component of the fitting vector f_1 describing the steady component was not used further for analyzing the features independent of the mean grey-level. A simple clustering of the 16 × 16 fitting vectors determined in this way for different *regions of attention* was showing a conformity with the human visual feeling. Several tests with different *regions of analysis* showed a generally exponential decay within the fitting vectors. This was a good prerequisite for an additional principal component transformation of the (45 – 1) fitting values to 2 (x and y) per processing column describing the *region of analysis* sufficiently. A simple cluster analysis of this two-component fitting vectors showed a still sufficient description of the *analyzing regions*. Based on these results each column is transmitting its fitting values x and y into a two-dimensional intercolumnar Kohonen Feature Map [2]. All columnar activations within a *focus of attention* (as mentioned above we implemented 16 × 16 processing columns per focus) contribute to activation distributions within the two-dimensional feature map. This map could be understood as an alphabet of all possible local grey-level distributions within real-world scenes. In result of internal competition and selection processes within the Kohonen Map only one cluster can be activated in a particular region of the map for a certain time whereas the others are suppressed until this cluster is deactivated. Using a topographical correct reciprocal projection from the map to the multicolumnar system that columns contributing to the just winning cluster are activated sequentially (see Fig. 7). The neural implementation of the analyzing and the Kohonen-based selection subsystems is the subject of our present work.

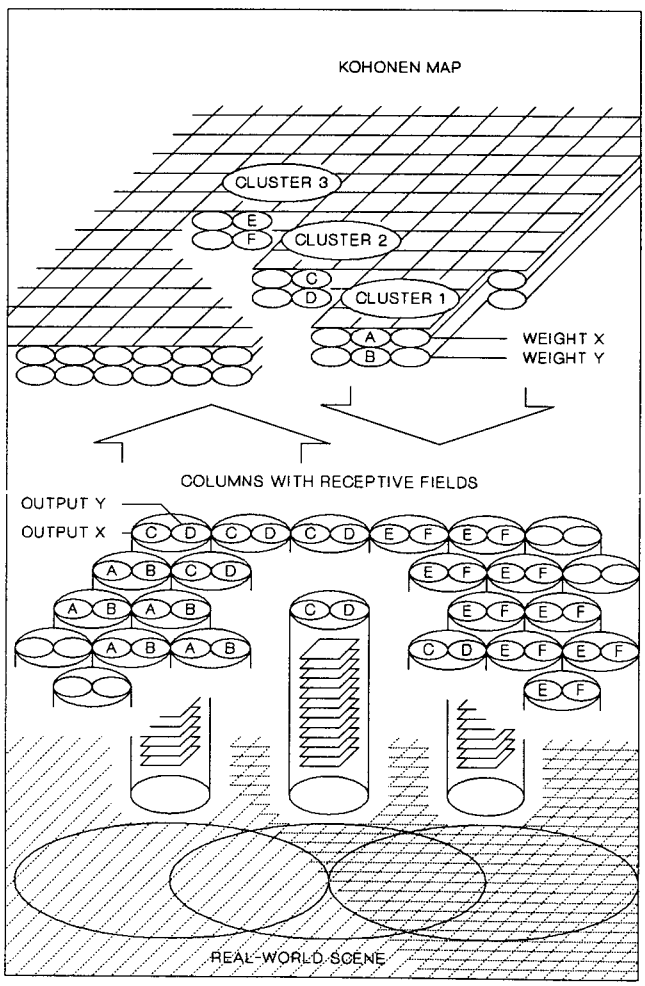

Figure 6: *The distributed multicolumnar system for primary cortical analysis, consisting of a subsystem for the two-phase principal component analysis (see text) and a subsystem for sequential selection of topological adjacent columnar analyzing results. AB and EF are columnar activations (analyzing results) describing completely different grey-level distributions (for instance textures) in a complex real-world scene. The activation CD is indicating a border-region of overlapping visual structures AB and EF. This cluster is localized within the Kohonen Feature Map between neuron populations sensitive for the features AB and EF.*

4. Results and Conclusions

By using such a two-stage principal component analysis (PCA) it is possible to describe local grey level segments of real-world scenes by only two fitting parameters (x, y) sufficiently. It became possible to realize a topology preserving mapping within an intercolumnar feature map. An analyzing region overlap of 75% per column leads to very compact cluster-formed activations within the feature

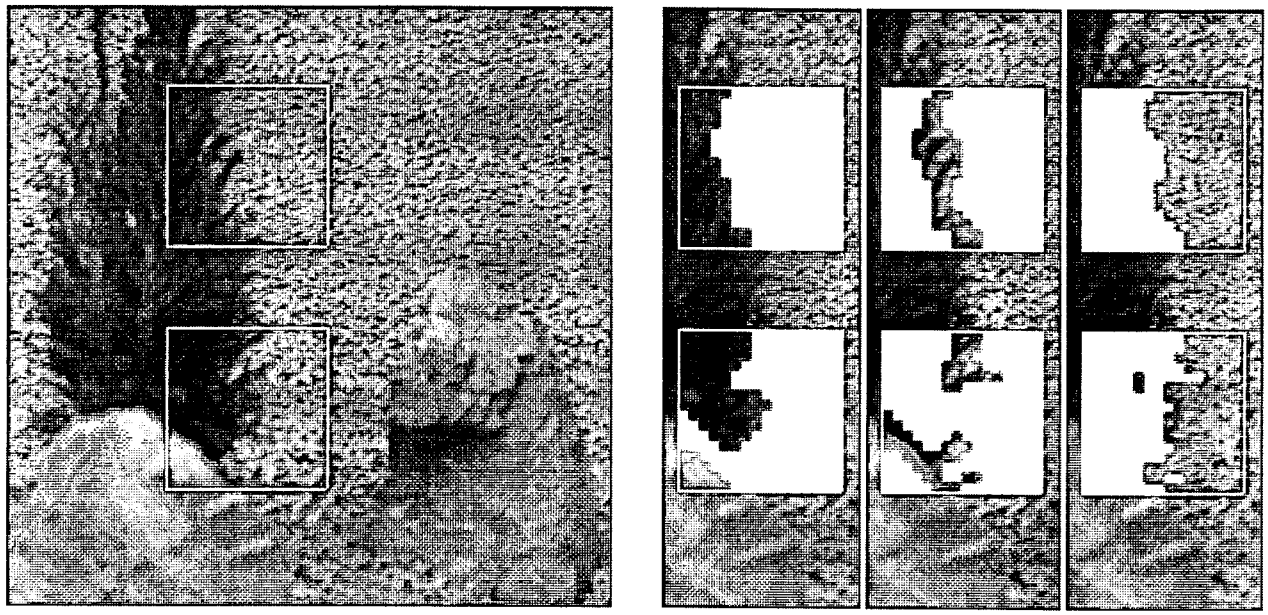

Figure 7: *(left) Real-world input scene. White frames show different regions of attention, that are processed by the distributed multicolumnar system. (right) Temporal developing of segmentation decisions within the regions of attention, based on the two-phase principal component analysis and the sequential selection in the Kohonen Feature Map. Similar regions of the attentional focus are selected and activated coherently time after time.*

map evoked by grey level distributions equal in quality inside the *region of attention*. The columnar architecture proposed in this paper has been tested at different real-world scenes. A typical result is illustrated in Fig. 7. It is interesting to establish that texture crossings independent of their direction are assigned here to one common cluster of mixed textures. In our further work this model architecture will be implemented for the analysis of coloured real-world scenes, organized in special colour difference spaces like that proposed in [1].

References

[1] **Gross, H. M., Koerner, E., Boehme, H. J., Pomierski, T.** A Neural Network Hierarchy for Data and Knowledge Controlled Selective Visual Attention. In Proceedings of ICANN92, vol. 1, pp. 825 - 828, Brighton, 1992.

[2] **Kohonen, T.** Self-Organization and Associative Memory, 3rd Edition, Springer Verlag, Berlin, 1989.

[3] **Oja, E.** A Simplified Neuron Model as Principal Component Analyzer. In J. Math. Biol., vol. 15, pp. 267 - 273, 1982.

[4] **Ritter, H., Schulten, K., Martinetz, T.** Neuronale Netze, 2. erweiterte Auflage, Addison-Wesley, Bonn, 1991.

[5] **Sanger, T. D.** Optimal unsupervised learning in a single-layer linear feedforward neural network. In Neural Networks, vol. 2, pp. 459 - 473, 1981.

[6] **Singer, W., Rauschecker, J. P.** The Effects of Early Visual Experience on the Cat's Visual Cortex and their Possible Explanation by Hebb Synapses. In J. Physiol., vol. 310, pp. 215 - 239, 1981.

[7] **Szenthagothai, J.** Theorien zur Organisation und Funktion des Gehirns. In Naturwissenschaften, pp. 303 - 309, 1985.