Proc. of the Int. Conf. on Artificial Neural Networks, Amsterdam 1993, pp. 63-66, Springer 1993

A Simple Selforganizing Neural Network Architecture for Selective Visual Attention *

D. Heinke, H.-M. Gross

Technical University of Ilmenau, Division of Neuroinformatics
O-6300 Ilmenau, POB 327, Germany
e-mail: dietmar@informatik.tu-ilmenau.de

Abstract

We present a simple neural network architecture which autonomously learns how to control a data driven selective attention process. In order to control the selective attention process a biologically plausible position coding is used which leads to fuzzy representations of position. An associative memory learns the connections between subsequent positions und local features. The result of presenting simple Real-World color images to the neural network architecture is shown.

1 Introduction

Because of the problem of combinatorical explosion a massively parallel data stream, e.g. a complex Real-World scene, can't be processed completely in parallel. Thus, it is necessary to transform a complex spatial data structure into a spatio-temporal data sequence [1] [2]. In order to achieve such type of transformation a data driven selective visual attention was introduced in [1].
The present paper is focused on an architecture for knowledge based control of a data driven selective attention process through a simple self-organizing neural network. After presenting Real-World scenes to the Neural Network architecture the selective attention process should scan the objects of a scene successivly. Hence, this behavior could be interpretated as autonomous knowledge acquisition about objects in a visual scene – without supervised training of single objects in a special learning phase [2].

2 The Neural Network Architecture

2.1 An Overview

Figure 1 shows the architecture of the whole system. In [1] a dynamical network for a data driven selective visual attention process was introduced. In the present work we use a modified version of that *selective attention network* (SAN). SAN generates an activity distribution which corresponds to the position of the actual "focus of attention". On the one hand this activity distribution is used to extract the area of attention out of the input scene and on the other hand it is used for a absolute position coding. Hence, we use activity distribution as a general position coding principle.
In order to be positional invariant the position coding of SAN is transformed into a relative position coding through the *position transformation* (PT). That means we get a coding of the distance and the direction between two successingly following foci of attention. In addition the PT maps relative

*supported by BMFT, Grant No. 413-5839-01 IN 101D - NAMOS-Project

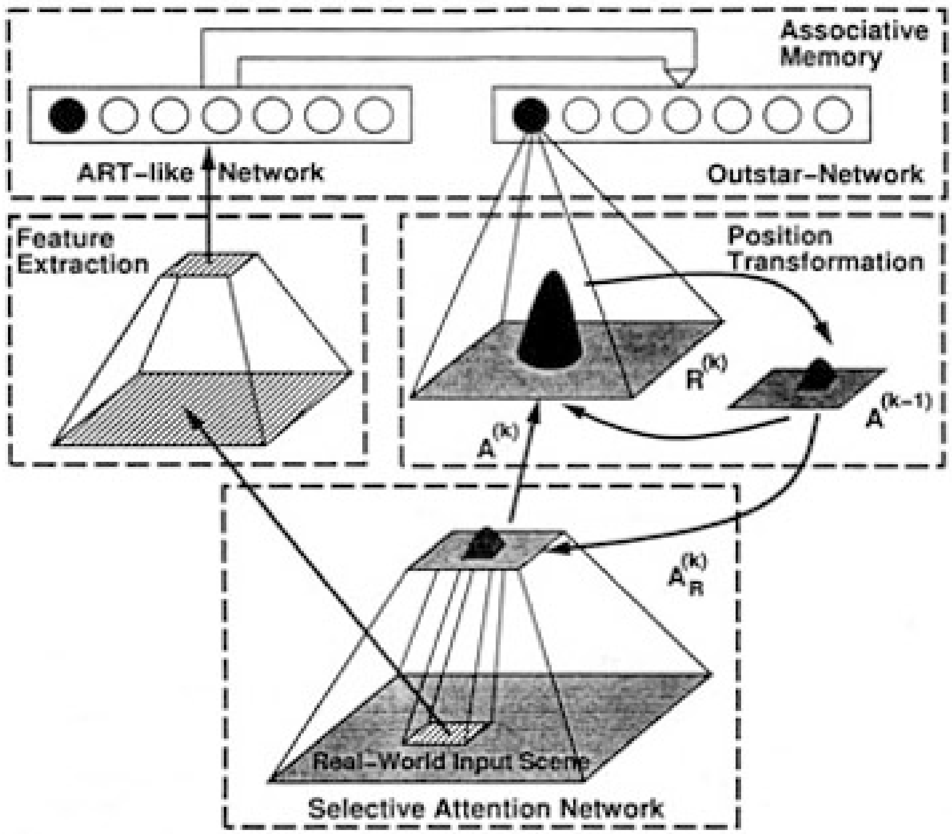

Figure1: The neural network architecture.

position codings into absolute codings. Through this mapping the SAN can be controlled in a top-down manner by the associative memory.
The *associative memory* (AM) is the central part of the whole system. It is able to learn the connection between the relative position coding and the output of the *feature extraction* of the focus area. It is continously learning using the statistics of the SAN behavior which are based on the scenes. Because in different scenes the objects are not always at the same position the frequency of inter-object moves is negligible. Thus, the AM can only learn and stabilize the moves within the objects of a scene. In addition the AM learns by successful control of the SAN. The strength of control increases with its success.

2.2 Feature Extraction

The feature extraction of the SAN consists of two multiresolution pyramids operating on a special color space, the WMM-space. The first pyramid is based on Laplace-filtering in the W-domain, the intensity image. The second pyramid uses the Euclidean distance to the average hue of the image in the $M_{rg}M_{yb}$ domain [1]. In order to have a strict local feature extraction in the focus area we simply use the WMM-space in an additional multiresolution pyramid.
For improving the feature extraction we currently discuss Eigenvectors to get a texture based enhancement [3].

2.3 Position Transformation

The position transformation uses two mapping rules (see Fig. 1). The first rule maps two successive SAN position codings $\mathbf{A}^{(k-1)}, \mathbf{A}^{(k)} \in I\!R^m \times I\!R^m$ into a relative position coding $\mathbf{R}^{(k)} \in I\!R^{2m} \times I\!R^{2m}$

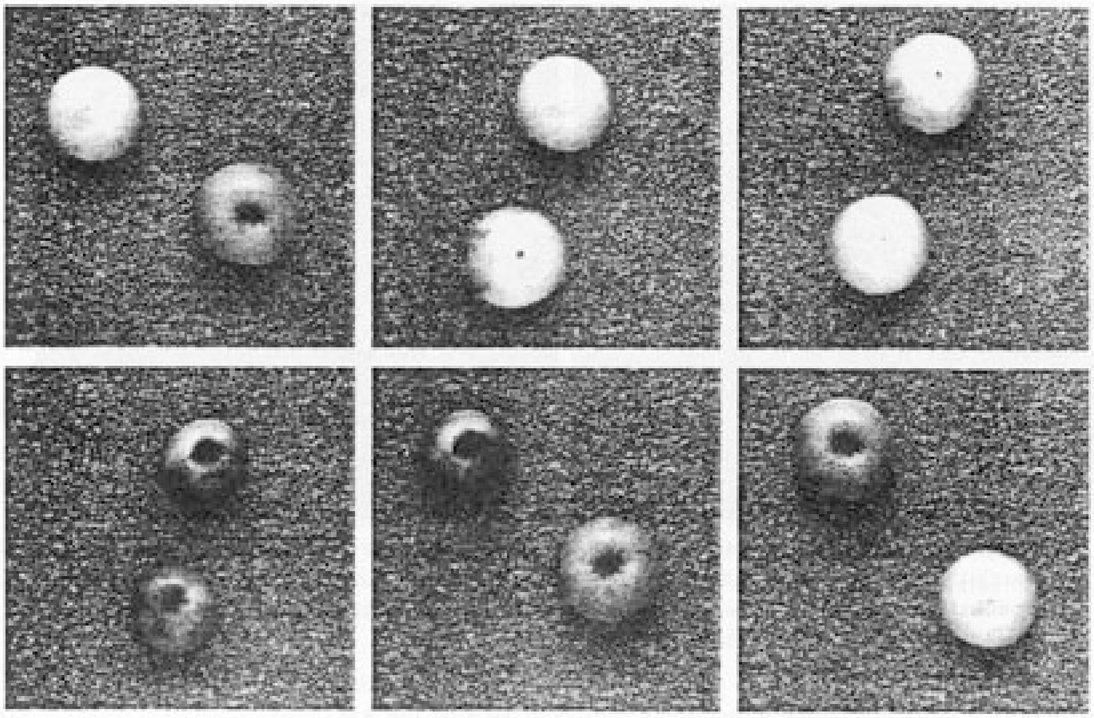

Figure2: The Figure shows examples of typical input images. The different objects are placed at different positions in a scene. This is the essential prerequisit for a selforganizing learning within the associative memory (see text).

through correlation:

$$r_{io}^{(k)} = \sum_{j=0}^{n-1} \sum_{l=0}^{n-1} a_{jl}^{(k-1)} a_{i+j,o+l}^{(k)}$$

The second mapping rule used in the opposite direction transforms a relative position coding $\mathbf{R}^{(\mathbf{k})} \in IR^{2n} \times IR^{2n}$ into a absolute position coding $\mathbf{A}_{\mathbf{R}}^{(\mathbf{k})} \in IR^{n} \times IR^{n}$ through convolution:

$$a_{io}^{(k)} = \sum_{j=0}^{n-1} \sum_{l=0}^{n-1} a_{jl}^{(k-1)} r_{i-j,o-l}^{(k)}$$

Hence, we use a biological plausible way of a implicit representation of positions. In addition this coding principle is a fuzzy representation, which makes the whole system more robust and can be simply learned by the AM.

2.4 Associative Memory

The *associative memory* (AM) consists of two neural networks. The first network is similar to ART2 [4] and represents the features extracted in the focus area in a sparse coded way. This representation forms a suitable input for the second neural network, an outstar network [5].
The outstar-network learns relative position codings which are frequently met with a defined output of the ART2-like network. Its weight matrix $\mathbf{W} \in IR^{p} \times IR^{2n} \times IR^{2n}$ is accessed through: $r_{ij} = w_{Kij}$, where K is the actually active output node of ART2 and p the number of the output nodes. The result of this operation is mapped through the position transformation and controls the SAN through a simple additive superposition. The weights of the outstar network are adapted in the following way:

$$\tau \frac{dw_{Kij}}{dt} = -w_{Kij} + a_{ij}^{(k)},$$

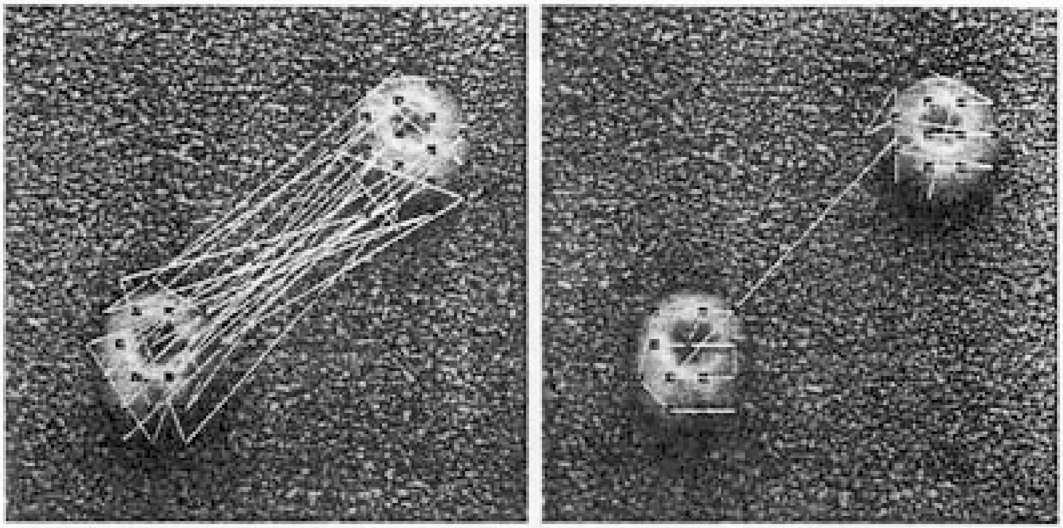

Figure3: The left part of the Figure shows the behavior of the neural network architecture with only data driven selective attention. The right part shows the behavior controlled by the associative memory after presenting 50 different images to the neural network architecture.

where $\mathbf{D^{(k)}} \in IR^{2n} \times IR^{2n}$ is the normalized relative position coding: $\mathbf{D^{(k)}} = \mathbf{R^{(k)}}/||\mathbf{R^{(k)}}||$. This learning rule leads to gradually increasing control of SAN by the AM.

3 Results

The whole neural network architecture was confronted with 50 different color images. Fig.2 depicts examples of them. All scences consist of different kind of fruits in different positions. Fig.3 compares the behavior of the neural network architecture before being confronted with the 50 images and afterwards. The "known" objects are scanned successivly because of the dominating control of the data driven selective process by the AM.

4 Conclusion

Currently we extent the present approach through the concept of efficiency [2]. This concept will be applied to two areas: At first the dynamic of the SAN can be used to achieve a more efficient use of processing time. At second the neural network architecture learns how to perform an efficient scan process within a given set of images. In this frame other top-down control mechanism than simple additive superposition are discussed.

References

[1] H.-M. Gross, E. Koerner, H.J. Boehme, T. Pomierski, "A Neural Network Hierachy for Data and Knowledge Controlled Selective Visual Attention", Proc. of ICANN92, 1, 825-828, 1992.

[2] H.-M. Gross, H.-J. Böhme, D. Heinke, R. Möller, T. Pomierski, "Steuerung parallel-sequentieller Verabeitungsprozesse und Strukturierung dynamischer Repräsentationen", will be published by Springer-Verlag.

[3] T. Pomierski, H.-M. Gross, D. Wendt, "A Distributed Multicolumnar System for Primary Cortical Analysis of Real-World Scenes", this volume.

[4] G.A. Carpenter, S. Grossberg, "ART2: self-organization of stable category recognition codes for analog input patterns", Applied Optics, 26, 23, 1987.

[5] S. Grossberg, "Some Networks That Can Learn, Remember, and Reproduce Any Number of Complicated Space-Time Pattern I", J. of Mathematics and Mechanisics, 19, 1, 53-91, 1969.