will be published in
Artificial Neural Networks II - Proceedings of the International Conference on Artificial Neural Networks - ICANN92, Brighton, United Kingdom, September 4-7, 1992, Edited by I. Aleksander and J. Taylor, Elsevier Science Publishers 1992

A Neural Network Hierarchy for Data and Knowledge Controlled Selective Visual Attention*

H.-M. Gross, E. Koerner, H.-J. Boehme, T. Pomierski

Ilmenau Institute of Technology, Department of Neuroinformatics & Cognitive Systems
O-6300 Ilmenau, P.O.B. 327, Germany, email: gross@informatik.th-ilmenau.de

Abstract

In this paper we present a neural implementation of a dynamical network hierarchy for data and knowledge controlled selective visual attention. The model architecture is composed of several interacting subsystems for different processing tasks. All neural subsystems are strongly interrelated bottom-up and top-down both by information and control streams.

1. Introduction

The phenomenon of selective attention in human visual perception points the way out of the dilemma of combinatorical explosion in analysis of real-world visual scenes: there are sequential processing modes intermingled with the parallel one. Selective attention means breaking down the flow of information too high to be managed by the analyzing system in parallel into meaningful pieces of lower dimension. The benefit of selective visual attention is, that the analyzing or identifying system has not to deal with all visual inputs in parallel but only with a limited sequence of presorted groups of input elements [1]. So the analyzing system can focus attention on the most desired visual stimulus among several simultaneously active stimuli, both driven by the input data and controlled by its internal processing state and the already acquired knowledge. This control of the attentional search during the recognition process is a fundamental mechanism for self-organization of sequential and episodical representations - a new type of non-trivial dynamical knowledge [2]. In this paper we present a neural implementation of a dynamical control hierarchy for data and knowledge controlled selective visual attention which is based on some facts from psychophysics and neurophysiology. In order to simulate an attentional search, we had to implement special mechanisms in our model

- to yield a measure of the conspicuity of a location in the visual scene
- to select the most active area (many different feature detectors active) in the highest mapping plane where the different feature maps are superimposed
- to shift the focus of attention from the current to the next striking location in the scene
- to control and manipulate the flow of information in the course of the attentional process taking into account internal system's knowledge about parts of the visual scene.

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

These mechanisms and the corresponding algorithms have been influenced essentially by the neurophysiological concepts of primary visual processing and selective attention of Koch [3] and Orban [4]. They showed that in the first sensory visual area instead of a hierarchical description of the input features a very rich parallel representation of the input by parameter filters is done. Any single input is split into a multiple parametrical description such as orientation, spatial frequency, length, disparity, color, motion etc. The more such parameter filters are triggered by a certain visual input, the stronger the superimposed activation of the receiving area will be.

2. The Functional Architecture of the Network Hierarchy

Our model architecture for data and knowledge controlled selective attention is composed of various interacting dynamical subsystems for different processing tasks. All these multi-layered neural subsystems constituting a dynamical control hierarchy are strongly interrelated by information and control streams. The main subsystems

- Interface Subsystem - **IS**
- Multi-parametrical Scale-Space - **SS**
- Selection and Control Network - **SCN**

are interconnected both bottom-up and top-down by connections of different types (non-adaptive, adaptive, time-varying).

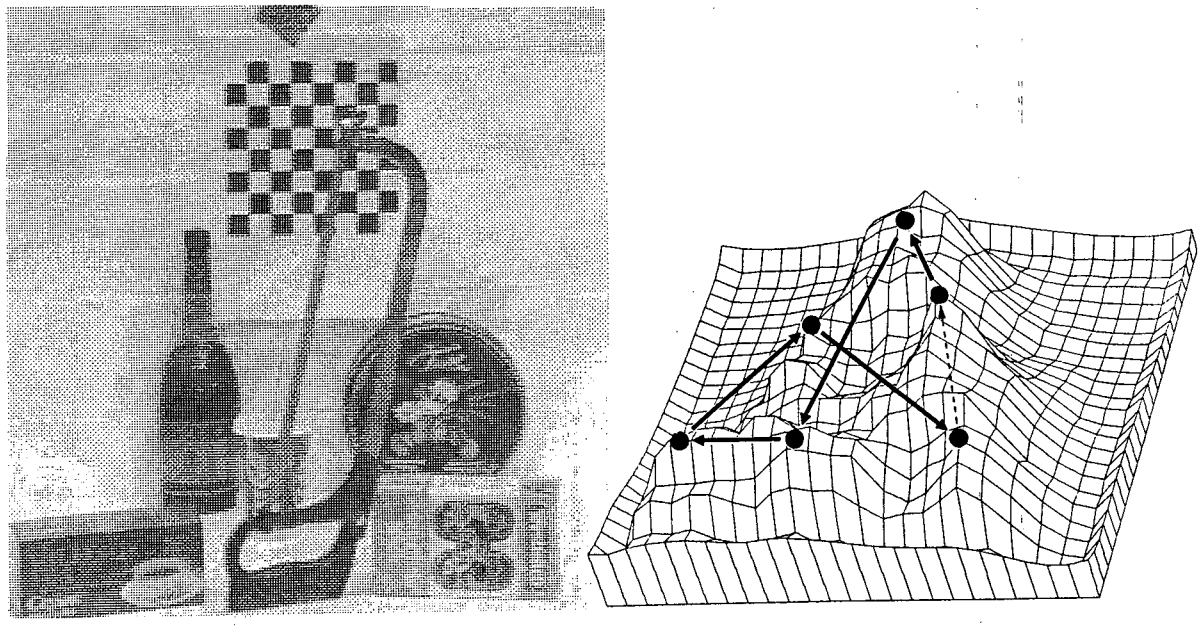

Figure 1: (Left side) Natural visual input scene for testing the sequential attentional search. (Right side) Sequential selection of the most active bubbles in the activity superposition plane, shifting of the focus of attention between the most striking locations in the scene.

To our mind selective visual attention requires the freedom to select only those parts out of the input which are needed at that time in the analyzing or recognition process. Therefore we have implemented the controllable **Interface Subsystem - IS**, which is based on a detailed model architecture of the thalamo- reticular complex proposed in [1]. Operating on this Interface

the succeeding higher subsystems can actively modulate and manipulate the bottom-up flow of input data streams giving them the freedom of formulating and testing hypothesis on special parts of the parallel input.

The **Multi-parametrical Scale Space - SS** operating on the Interface Subsystem detects in distinct analyzing pathways differences in local conspicuous features of the IS activity pattern (fig. 1 - left side). Additionally, the input is not only analyzed for the different features in the initial level of resolution but also in different lower resolutions. By reducing the resolution and size we get a processing pyramid that realizes both an enormous reduction of the amount of input data and some form- and position invariance at the top. This local invariances are essential for succeeding parallel-sequential pattern recognition mechanisms [5] not presented in this paper. Consequently the **SS**-subsystem can be considered as a set of several pyramidal organized topographical maps of the visual scene. Each of these analyzing maps includes at each resolution level parameter filters for different elementary features (texture density, color and intensity contrasts, spatial frequency). In this way any single input location is split into a multiple parametrical description. Up to now analyzing pathways for differences in local texture density and for local contrast in color and gray-value intensity have implemented as pyramids in the model, other scale space pathways are at work. By weighted superposition of the neural activity between the several feature maps an encoding of high syntactic complexity (many different feature detectors activated at the same time) into an blurred activity distribution at the top of the pyramid is realized (fig. 1 - right side) . The more such different parameter filters are triggered by a certain visual location, the stronger the total activation of this specific area in the highest scale space level will be.

The **Selection and Control Network - SCN** realizes a kind of cortically controlled selection of the most conspicuous location in the visual field which has been encoded as activity peak or bubble in the Scale Space. In result of an internal relaxation process in this feedback network only a single, the most conspicuous location of the visual scene will remain actively. When the input to **SCN** has multiple activity peaks because of several striking locations in the visual scene (fig. 1), the network has to select the maximum one or that peak with the highest competition energy (extension and altitude of the activity bubble). Since this relaxation process toward a stationary solution needs some time, a higher knowledge based processing level has enough time to activate its acquired knowledge about the presented input. By a feedback from that higher processing level to **SCN** the activated knowledge can modulate directly the dynamics of the sharpening process in **SCN** [5]. Therefore this subsystem has been implemented as controllable single-layer feedback network. The activity dynamics of each neuron in the subsystem can be described mathematically by the following differential equation:

$$T\frac{dz_{ij}(t)}{dt} + z_{ij}(t) = e_{ij}(t) + c_{ij}(t) + \Theta(e_{ij}(t)) * \mu * \sum_{\substack{k=i-a \\ l=j-a}}^{\substack{i+a \\ j+a}} y_{kl}(t) - \nu * \left(\sum_{\substack{k=1 \\ l=1}}^{N} y_{kl}(t) - \sum_{\substack{u=i-b \\ v=j-b}}^{\substack{i+b \\ j+b}} y_{uv}(t) \right)$$

with the nonlinearity

$$y_{ij}(t) = \Phi(z_{ij}(t))$$

and the input modulated cooperation power

$$\Theta(e_{ij}) = \begin{cases} 0 & : \quad e_{ij} < \alpha * \frac{1}{N^2} * \sum_{\substack{i=1 \\ j=1}}^{N} e_{ij} \\ e_{ij} & : \quad \text{else} \end{cases}$$

$e_{ij}(t)$ denotes the excitatory input to neuron (i,j), c_{ij} the excitatory or inhibitory top-down control signal from the knowledge base to neuron (i,j) and μ und ν denote the coupling parameters for cooperative and competitive connections.

3. Dynamics of the model and simulation results

After presenting an input to the hierarchy several decisions on the input are developing in SS and SCN because of the different syntactic complexity at the several locations in the input (fig. 1). The local cooperation and ensemble competition between the neurons in the **SCN** suppress activity bubbles with weaker competitive power and sharpens the remaining one so that only that bubble with the highest energy survives. A moving of the selective attention can be forced in our framework by either adaptation of the activity at the once selected location in the several levels of the Scale Space or by switching off the selected representation in the interface for a certain time via feedback from **SCS**. These channel-specific mechanisms of adaptation or controlled inhibition take the selected decision for a certain time out of discussion. In this way the next decision can develop only on the remaining parts of the input and the next-grade complex known input configuration will start this search process anew. In a time sharing manner other input parts can be selected, thus creating a reverberating sequence of limited numbers of such decisions. In this way our system decomposes a complex visual input into a time sharing sequential representation of local input parts. Without a knowledge based top-down manipulation of the systems dynamics, the established sequence depends only on the local syntactic complexity of the selected input locations. The *attentional selection* based on internal systems knowledge about spatial relations and form features of several input parts can be realized by the knowledge controlled modulation of the preattentive dynamics mentioned above. To this end an adaptive parallel-sequential classifying system like **GNOM** proposed in [5] has been implemented in the hierarchy.

Our paper will demonstrate with the example of visual scene analysis at hand, that the proposed network hierarchy is well suited for realizing an input data driven *selective visual preattention.* Because of the network topology and dynamics and the functional architecture of our model a coupling of preattentive and knowledge based attentive vision mechanisms is possible. This coupling is an object of research at present.

References

[1] **Koerner,E., Tsuda,I., Shimizu,H.** Parallel in Sequence–Towards the Architecture of an Cortical Processor. In Parallel Algorithms and Architectures, Akad.-Verlag Berlin 1987, p.37-47

[2] **Koerner, E., Boehme, H.-J.** Organization of an Episodic Knowledge Data Base. In Proceedings of ICANN91, vol. 1, pp.873-878, North-Holland 1991

[3] **Koch,C., Ullmann, S.** Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4(1985) p. 219-227

[4] **Orban,G.A.** Neural operations in the visual cortex. Springer Verlag Bln., Hdbg., NY, Tokyo 1984

[5] **Gross,H.-M., Koerner, E., Pomierski,T.** GNOM—A Modular Network Architecture for Adaptive Parallel–Sequential Pattern Recognition. In Proceedings of ICANN91, vol. 1, pp. 747-752, North-Holland 1991