ICANN '94

Proceedings of the International Conference
on Artificial Neural Networks
Sorrento, Italy
26–29 May 1994

Volume 1, Parts 1 and 2

Edited by
Maria Marinaro and Pietro G. Morasso

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

Self-Organizing a Behaviour-Oriented Interpretation of Objects in Active-Vision

H.-M. Gross, H.-J. Boehme, D. Heinke, T. Pomierski, R. Moeller
Department of Neuroinformatics, Technical University of Ilmenau
D-98684 Ilmenau, Germany

1 Introduction

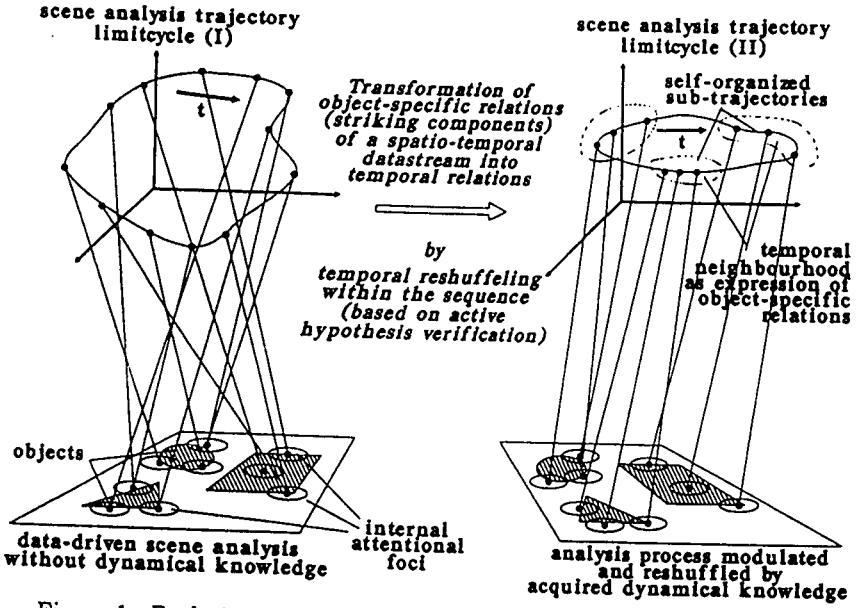

Figure 1: *Evolution of stable temporal relations within sub-trajectories*

This paper deals with a new approach towards self-organization of an intrinsic (behaviour oriented) spatio-temporal object-understanding in the context of an active-vision process in the widest sense. The focus is on the functional architecture and the dynamical principles suited for self-organization of knowledge about complex visual structures, and for a *behaviour-oriented interpretation* of objects in real-world scenes. In our mind, such a self-organization of intrinsic dynamical knowledge is a prerequisite for active, autonomous learning under real-world conditions. In this context an approach is used in our model, whose sub-sequential states of learning are related to the development of the systems behaviour in visual scene analysis. It is well known, that the visual input to an analyzing system dealing with real-world problems is not a set of preselected, figure-ground segregated objects which are to be properly arranged, but the system itself has to select reliably detectable features or input components out

of the massively parallel input (v.d. Malsburg 1992). Selective attention processes (both active-vision and internal scanning not related to eye movement) and dynamical processing at different organizational levels are widely accepted mechanisms explaining the decomposition of a complex visual scene into components and the subsequent reassembling the recalled internal representations towards an unitary decision (Crick et al., 1990). Therefore, an important aspect of our approach is the data and/or hypotheses driven dissolution of the highly parallel visual input into meaningful components, which can be reassembled freely to new complex visual structures. Only in this case, the analyzing system is able to handle the present input on the base of the knowledge already acquired at any time. In addition, only by such a continuous interaction a self-organization is possible, the aim of which is to bring the actual perception into maximal consistency with the acquired knowledge. The goal of our approach is to find useful ways of exploiting the wealth of dynamic behaviour for such aspects like active, autonomous learning of internal representations, which are assumed to be fundamentally for a behaviour-oriented understanding of visual objects during active-vision. Of our particular interest are such concepts like generation and active verification of dynamical hypotheses about the input in a feedback coupled process of *Sensory Controlled Internal Simulation*. In the context of the systems behaviour during active vision that means the generation and testing of hypothesis about *what* components are to be expected *when* and *where* in the visual field – this is an internal anticipation of a real spatio-temporal selective attention or active-vision process. Therefore our model concept proposed later should be able to map the temporal and spatial characteristics of the data driven serial processing within the intervals between the eye movements shown strongly simplified in Figure 1 into characteristic and dynamically stable internal representations coding stable striking feature relations within the the objects. Without any internal knowledge our functional architecture is not able to establish suitable hypotheses about objects to anticipate an internal scan process. Self-organization of an object understanding means, that typical, reliably detectable striking visual components and their object-specific relations detected in preceding active vision processes more frequently, gradually can be coupled or linked in the temporal domain as *temporal adjacent*. The evolution of stable temporal relations (temporal neighborhood) within subtrajectories is an expression of stable object-specific relations within input data stream and is to understand as a behaviour-oriented internal understanding which components of a visual structure (object) belong together. In our opinion, *temporal neighborhood* in selective attention processes could be a good, possibly the only criterion for an unsupervised segmentation and learning of objects arranged in highly structured visual scenes. Figure 1 (on the right) shows the final state of such a behaviour-oriented transformation. By knowledge-based reshuffling the input sequence the development of such a sequence (limit cycle) of object-specific sub-trajectories is forced which organizes best all relevant input components into a globally consistent decision.

2 Functional architecture

For dealing with autonomous learning and self-organization of such a behaviour-oriented object-understanding under real world conditions we developed a neurobiologically inspired functional architecture. Our concept presented here is

an improved but still very simple computational model of a modular processing hierarchy compared to an earlier approach of us (Koerner et al., 1991). The architecture to be developed in the NAMOS-project is to decompose a complex visual input into a reverberating sequence of reliably detectable fragments (components) ranked by its visual conspicuousness (complexity) and controlled by the self-organized knowledge. The organizational levels of our model schematized in Figure 2 define the following basic abilities and information processing tasks: The **Saliency System** proposed in (Gross et al., 1992) has

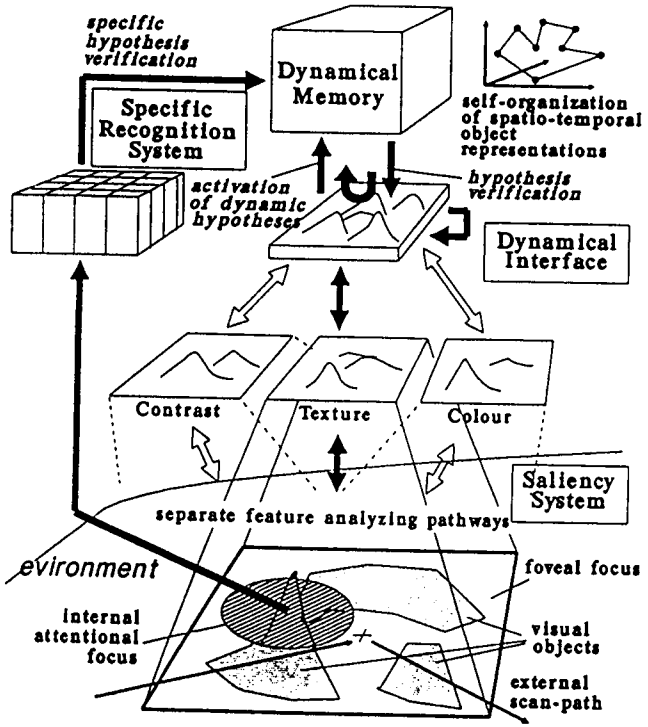

Figure 2: *Sensory Controlled Internal Simulation*

been influenced essentially by the neurophysiological concepts of primary visual processing (DeYoe et al., 1988). The parallel representations at different feature maps yield a measure of the conspicuity of a location in the scene. This is prerequisite for a data-driven decomposition of the visual input into striking and reliably detectable components that can classified as known or unknown by the following levels and can be reshuffled and reassembled freely.

Based on its inherent dynamics the **Dynamical Interface** carries out a sequential search for the most striking components within the visual field by shifting its internal attentional focus. The Dynamical Interface is modulated both bottom-up by the sensory input from *Saliency System* and top-down by spatio-temporal recall from *Dynamical Memory* in context of an unspecific hypotheses verification. In this way the Interface and with that the sensory data stream can be controlled by activated hypotheses about the input so that the interesting components can be reshuffled in time according to the state of internal hypthesis activation (Gross et al. 1992, Heinke et al., 1993). The **Specific Recognition System** operates on the internal attentional focus controlled by the *Dynamical Interface* within the *Saliency System*. In cooperation with the *Dynamical Memory* it determines in a *specific hypothesis verification* the similarity of the actual internal focus feature set to the feature sets extracted and learned autonomously in previous cycles (see Pomierski et al., 1993). The **Dynamical Memory - (DM)** is the highest organizational level of our architecture. DM is activated and driven by the established spatio-temporal sequence of striking input components and can act as a guide in attentional control and input decomposition based on the knowledge already accumulated within the system. Therefore DM is interacting reciprocally with the *Dynamical Interface* in so-called 'hypothesize-verification-cycles'. All activated hypotheses interfere back to the *Dynamical Interface* and try to control the course of data-driven search within this system. Via this feed-back DM can search for that input components which would support one of the activated hypotheses. In this sense, DM uses its internal self-organized knowledge for flexible activation and continuous verification of hypotheses about the input (see (Boehme et al., 1994)). This forces the development of such a sequence of decisions which organizes best all selected input components into a globally consistent decision. If it is impossible to activate internal hypotheses by a data-driven input sequence this input has to be accepted by the DM as a new sensory situation. On this background a model was developed and presented in (Boehme et al., 1992) as a first very simple attempt to extract the inherent structure of a spatio-temporal data stream in active-vision or selective attention processes. Based on this first model the DM tries to map each sensory input sequence into a characteristically memory trace. So, the sequence of decisions on certain striking input components is transferred into a spatio-temporal representation within DM. For the case, that the input is partially unknown, DM can generate sub-hypotheses on the base of the already accumulated knowledge. If no interpretation of the input pattern sequence is possible, this input sequence will be learned. In first simulations (see Heinke et al., 1993) our system was able to change its behaviour in scanning unknown complex scenes, that is a result of the transformation of detectable object-specific relations uncoupled with respect to time at the beginning into more and more stable temporal relations by autonomous learning. In next time we will couple all different organizational levels of our model in a comprehensive simulation.

References

Boehme, H.-J. et al. (1992); Proc. of ICANN'92, 1381-84
Boehme, H.-J. et al. (1994); submitted paper to ICANN'94
Crick, F. & Koch, Ch. (1990); Sem. in Neurosc. 2 (1990) 263-275
DeYoe, E.A., Van Essen, D.C. (1988) TINS, 11 (1988) 5, 219-227
Gross, H.-M., Koerner, E. (1991); Proc. ICANN'92, 825-828
Heinke, D., Gross, H.-M. (1993); Proc. ICANN'93, 63-66
Koerner, E., Boehme, H.-J. (1991) Proc. of ICANN'91, 873-78
v.d.Malsburg, Ch., Buhmann, J. (1992) Biol. Cybern., 67 (1992) 233-242
Pomierski, T., Gross, H.M. (1993) Proc. of ICANN'93, 142-47