Appeared in: *Proceedings of the 3rd ESANN*, D facto, Brussels, pp. 175–180, 1995

An Episodic Knowledge Base for Object Understanding *

U.–D. Braumann, H.–J. Boehme, H.–M. Gross

Technical University of Ilmenau
Department of Neuroinformatics
D–98684 Ilmenau, Germany

Abstract. Standing in close relation with a neural functional model for the self–organized development of a behavior–oriented object understanding (see [3]) this contribution deals with a model for the formation of object hypotheses. Therefore, this paper emphasizes the characteristic of the data to be processed, the internal representation of an emerging object relationship and the self–organization of the internal structure. An illustrating simulation example is given.

1. Introduction

Any environment in which a biological information processing system operates is characterized by information streams. Single events commonly occur as particular cases.

Due to this characteristic of the information entering intelligent biological systems interacting with a natural environment these systems must try to find the best solution of a decision situation by means of a permanent interaction between the incoming data stream and the internal decision making process. We call this principle *Sensory Controlled Internal Simulation (SCIS)* (see [2]), which doesn't allow a separation of training and recall considered to be essential for many neural network models. Instead, SCIS requires online–learning mechanisms.

Against this background a neural overall architecture was outlined and implemented to develop an emerging behavior–oriented object understanding on the basis of self–organization by interaction with its surroundings (analysis of complex visual scenes). The information stream entering the architecture is internally represented according to its characteristic in order to make the currently best possible decision based on this representation depending on the concrete sensory situation.

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

2. Embedding architecture

Fig. 1 depicts the embedding overall architecture of the model of the self–organization of an object understanding emphasizing the partial architecture to be elucidated below.

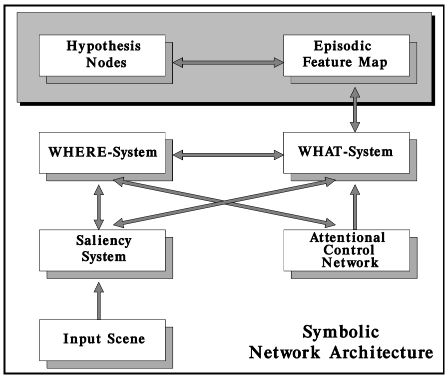

Figure 1: *Symbolic neural network architecture of the overall model for visual scene analysis.*

The overall architecture is explained only as far as it is necessary for the reader's understanding of the architecture described in the presented paper. A more detailed description of the other parts of the overall model is given in [4].

Figure 2: *Two examples from the gallery of scenes presented to the overall model (appears monochromous in print).*

Fig. 2 shows two examples from the gallery of the used stationary color scenes (images) containing objects against structured backgrounds which vary in both their arrangement and their appearance. Neither rotation nor scaling of the size of the objects has been done.

The saliency system computes a saliency map of an image which is the base of the purely data–driven move of a "window of attention" over that image. At each stopping point of that window a complex feature is extracted for the taken detail.

By means of interactions of Where- and What–system stable transitions between features appearing within the input scenes are detected, represented and directly used for the knowledge–based influencing of the scanning process. Here, the Attentional Control Network takes a control function.

Concerning the process of self–organization it is important to assume that objects appear as stable visual structures which are emerging by scanning various scenes and then lead to an increase of the effectiveness of the scanning behavior.

However, the local hypotheses which can be generated by this model one should rather consider as the preliminary stage for an implicit object understanding since more global relationships between features as candidates for objects or parts of objects cannot be represented yet. This is the crucial motivation for the development of the *Episodic Feature Memory* by which a global object relationship shall be represented in order to get hypotheses on complex visual structures in the sense of objects.

3. The Episodic Feature Memory – architecture and behavior

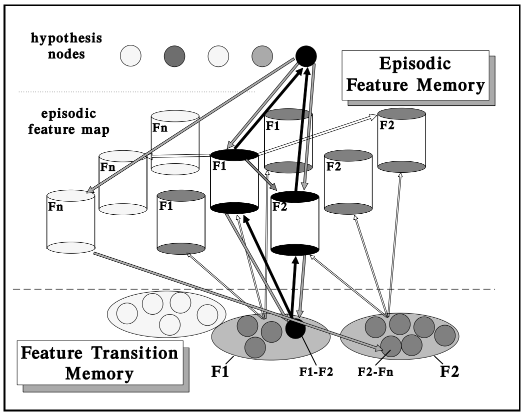

Figure 3: *Sketch of the internal structure of the Episodic Feature Memory (consisting of Episodic Feature Map and Hypothesis Layer) including the interactions with the Feature Transition Memory of the overall architecture. The cylinders symbolize the columns of the Episodic Feature Map.*

Fig. 3 schematically shows the structure of the Episodic Feature Memory (EFMem). It consists of the Episodic Feature Map (EFMap) and the Hypothesis Layer (HL). The interactions with the hierarchically subordinated part of the overall model is made via the Feature Transition Memory (FTM) of the What–system. This FTM has clusters of units (nodes) which encode relative jumps (moves of the "window of attention") from one starting feature to various target features. Each single node represents just one single feature transition. While in the FTM hypothseses on the next most likely jump can be generated and verified the EFMem tries to take sequences of successfully verified single jumps and keep them as candidates for objects or parts of objects.

The EFMap consists of columns arrangend in clusters, too. Similarly, every cluster also codes a certain feature. The uniformly structured columns of the EFMap are functional autonomous processing units consisting of a number of different subnodes. By this, the different influences of one map element can be treated separately. A detailed description of the internal column–structure cannot be given for space reasons.

Each cluster of the FTM unspecifically activates the corresponding column cluster of the EFMap. If a FTM–node becomes active this signals a successfully predicted feature transition. If one feature transition occurs for the first time a column of the relevant cluster of the EFMap is randomly chosen whereas this bidirectional assignment will not be modified anymore. The mutual weights between the two EFMap–columns are strengthened. Assumed, more active FTM-nodes follow, additional columns might be randomly chosen. Besides, the following feature transitions can already be represented by existing links between the EFMap–columns. This process is continued as long as no further jump can be predicted in the FTM. As a result of this process assemblies of mutual connected EFMap–columns emerge which can activate each other. Each assembly represents a sequence of successful feature transitions.

The nodes of the HL (hypothesis nodes) represent the different established column assemblies in the EFMap. If an assembly arises in the EFMap that not sufficiently matches an assembly already coded by a node of the HL a new node is added to the HL encoding that assembly. The mapping between the column assemblies and the hypothesis nodes changes during the organization process. The activity of the hypothesis nodes shows to which extent the various assemblies have established in the EFMap and thus determines the strength of influence of the EFMem upon the FTM. This influence is affecting the nodes of the FTM indirectly via the corresponding columns of the EFMap.

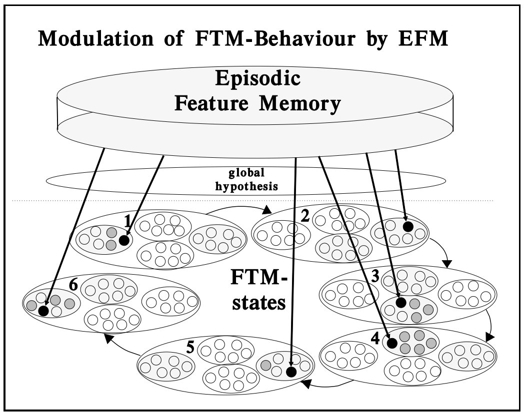

Figure 4: *As an example six stages of the process of hypothesis making are shown. The dark circles symbolize the successfully verified feature transitions. The arrows from the EFMem down to these nodes show the effect of the global hypotheses which could be verified in these cases.*

The cooperation of the FTM and EFMem is shown in Fig. 4. Therefore, several stages of the FTM following each other are depicted symbolically. By means of the inclusion of local feature hypotheses in global hypotheses and the backward–directed projection of them into the FTM the overall model is able to include knowledge about even more complex stable relations into the control of the scan path. The nodes that would have been selected just on the basis of

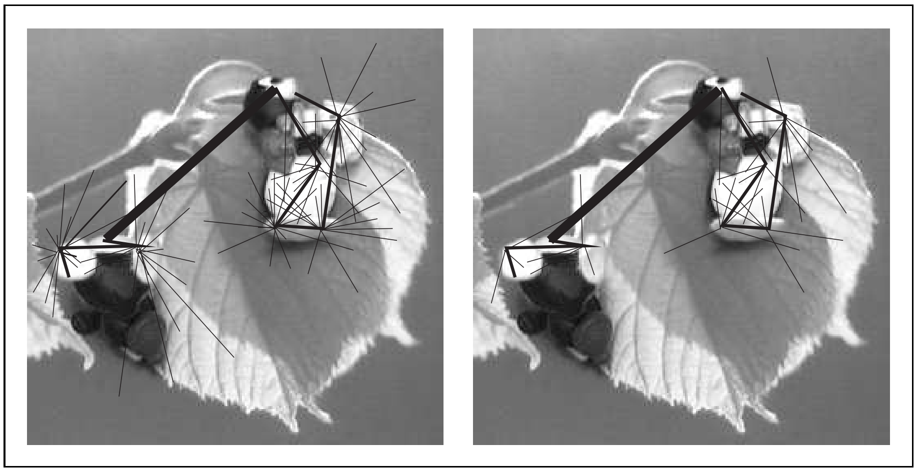

Figure 5: *Comparison of two simulations: the left was made without the effect of a global hypothesis whereas the right could benefit from that influence. The thick line in both images indicates a purely data–driven jump. See text for more details.*

local hypotheses alone are marked gray. The black labelled nodes depict those which finally were selected by the fusion of local and global hypotheses in order to control the scanning of the image.

The two simulation results depicted in Fig. 5 clarify how the backward-directed projection from EFMem onto the FTM has effect on the scanning of real scenes. The scan path in the left image was exclusively modulated through the available local hypotheses on features following each other immediately. In both images a different number of thin lines are marking just not executed jumps which were suggested by hypotheses but could not be verified, while the successfully executed jumps are marked with medium thick lines. In the right simulation example the scan path for the same image was modulated by hypotheses affected by the EFMem. It shows that compared to the previous example far less non–verifiable hypotheses arose. Faulty hypotheses cannot be completely suppressed with the current implementation stage for several reasons: Firstly, the complete gallery of scenes (100 images) has been presented to the system only once. Secondly, plenty of feature transitions appear in several objects, thus a heavy overlapping within the EFMap–column–assemblies results. Thirdly, the presented architecture is just an approach towards an episodic representation of feature sequences and subject to further refinement.

4. Summary

As described in the above paper the most important goal of our mentioned overall model is to develop a behavior–oriented object understanding mainly based on self–organization. The information to be processed are sequences of complex features extracted while scanning visual scenes, whereas the way of scanning itself expresses the level of the acquired object understanding. Objects are assumed to appear as stable visual structures within scenes of the used gallery. As mentioned above, it doesn't seem to be sufficient to use only local hypotheses to manage the scanning of scenes. That is why we developed a model monitoring the scanning based on local hypotheses in order to detect global relationships of features as candidates for objects. By this, a mechanism of forming global hypotheses on objects or parts of objects can emerge. The *Episodic Feature Memory* is that model where such relationships are represented both as mutual activating assemblies of columnar units and as single hypothesis nodes. As a result of the influence of global hypotheses the scanning of the scenes became much more effective.

References

[1] Boehme, H.–J., Braumann, U.–D., Gross, H.–M.: *A Neural Network Architecture for Sensory Controlled Internal Simulation.* In: Proceedings of the ICANN–94, Vol. 2, Springer–Verlag, Berlin – Heidelberg – New York, pp. 1189–1192, 1994

[2] Gross, H.–M., Boehme, H.–J., Heinke, D., Moeller, R. and Pomierski, T.: *Steuerung parallel–sequentieller Verarbeitungsprozesse und Strukturierung dynamischer Repraesentationen. (in German)* In: Informatik–Fachberichte zum Statusseminar Neuroinformatik, Schloß Maurach am Bodensee, 1992

[3] Gross, H.–M., Boehme, H.–J., Heinke, D., Pomierski, T. and Moeller, R.: *Self–Organizing a Behaviour–Oriented Interpretation of Objects in Active–Vision.* In: Proceedings of the ICANN–94, Vol. 1, Springer–Verlag, Berlin – Heidelberg – New York, pp. 58–61, 1994

[4] Heinke, D., Gross, H.–M.: *A neural Network Architecture for Self–Organizing of Object Understanding.* In: 39. Internationales Wissenschaftliches Kolloquium, Vol. 2, Technische Universitaet Ilmenau, pp. 124–131, 1994