To be presented at ICPR'96,
Vienna, Austria

Task Relevant Relaxation Network for visuo-motory Systems*

Fred Henrik Hamker, Horst-Michael Gross

Technische Universität Ilmenau
Fachgebiet Neuroinformatik
D-98684 Ilmenau, Germany
Phone: +49/3677/69-2858
Fax: +49/3677/69-1665
e-mail: fred@informatik.tu-ilmenau.de

ABSTRACT

The basic idea of our approach is to avoid the separation of perception and the generation of action. Therefore, information extracted from the environment is selected according to the relevance of the systems intended action. Task-specific attention emerges in an attention selection map from a competition among several hypotheses in consideration of task relevant subgoals. This decision initiates the emergence of a task-specific focus of attention in a saliency map which refers to a textured or coloured region in a scene.
The networks underlying the attention selection map and the saliency map consist of interacting columns with local excitatory and global inhibitory coupled feedback. In the attention selection map lateral cooperation is used as a way to integrate task pertinent subgoals. Implemented subgoals are the size of regions, the security of a classification hypothesis and the valuation of the hypothesis for the task.
The performance is demonstrated on a real-world selection of textured objects for a robot grasping task.

1. Introduction and fundamental ideas

Conventional artificial visual systems are often separated between recognition and generation of behaviour. They are usually based on sensory representations without pertinence to the intended behaviour of an agent. These approaches have to interpret the visual data by a special control structure (a "homuncolus") to act appropriately.
In the actual discussion about perception and action, some hints can be found "that perceptual systems have evolved in all species of animals solely as a means of guiding and controlling action" [1].
Therefore, we try to surmount these shortcomings and introduce a neural interactive process, which avoids the separation of perception and the generation of action by implicitly selecting only visual regions significant for the systems behaviour and action at this moment. The selected region serves as a focus of attention for further analysis and finally for the access to the object in this region by a manipulator.
The core of our approach is an attention system which can be divided into an attention selection map and a saliency map. Competition based on lateral cooperation and global inhibition leads to the formation of an assembly which suppresses the inhibition of the input of a saliency map. As a result, the activity stream leads to a formation of a salient region by means of competition and cooperation. The aspect of cooperation shows similarities to models known as relaxation labelling [10] [7], which are known for their segmentation properties [8] [9] [11] [4].

2. Subgoals

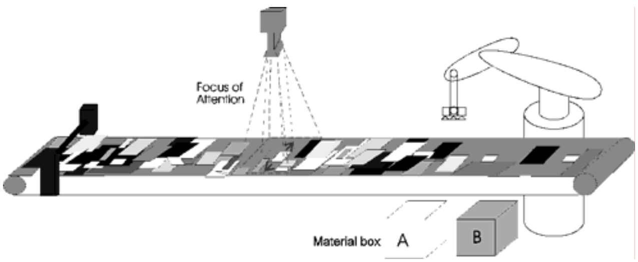

Fig. 1: A vision-controlled robot conveyer system for the separation of recycling material.

A task can be expressed by a variety of subgoals that should be fulfilled in the task. As an example, our contribution

* supported by the BMBF, Grant No. 413-5839-01 M 3014C - MIRIS-Project.

describes the task of a vision-controlled robot conveyor system for the separation of recycling material, e.g., wastepaper (fig. 1).
The object in the visual area of the “highest interest” at this moment should be detected and labelled. The “highest interest” is based on the definition of the task. In contrast to conventional approaches with explicit knowledge based task planning the “highest interest” emerges as a result of interactive information processing modulated by the subgoals. Therefore, the subgoals of the task (fig. 2) have to be transformed into top-down modulated weights between interactive nodes to realize early visual task relevant attention. This concept is based on the assumption that perception depends on the actual situation and that identical sensory situations can cause completely different actions, according to internal state and actual task.
At present implemented subgoals are the "size" of regions, the "valuation", important for preferring certain hypotheses and the "security" to find the proposed material in a receptive field. Clearly, the number of subgoals is not restricted to those discussed in this article, others are also possible, e.g., the minimization of movement access costs or the ability to grasp objects.

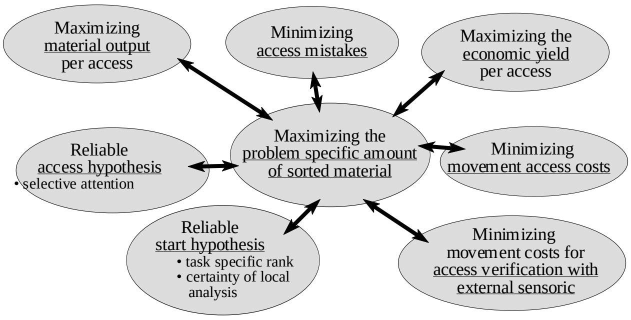

Fig. 2: Subgoals describing the task of the vision-controlled robot conveyer system.

3. System Architecture

This paper demonstrates, how our approach can be used to select relevant areas in a cluttered scene composed of different textured objects, important for a robot grasping task (fig. 3). The whole architecture can be subdivided into a feature extraction and an attention part. The first part is not subject of this issue. But for better understanding attention resulting from the relaxation networks, some facts of the early feature extraction are remarked.

3.1. Overview

For feature extraction, the colour image taken from a scene is converted into three colour-opponent maps (black-white, blue-yellow and red-green) to achieve an increase of information for coloured objects. On basis of these colour-opponent maps, a topological feature space is spanned, which

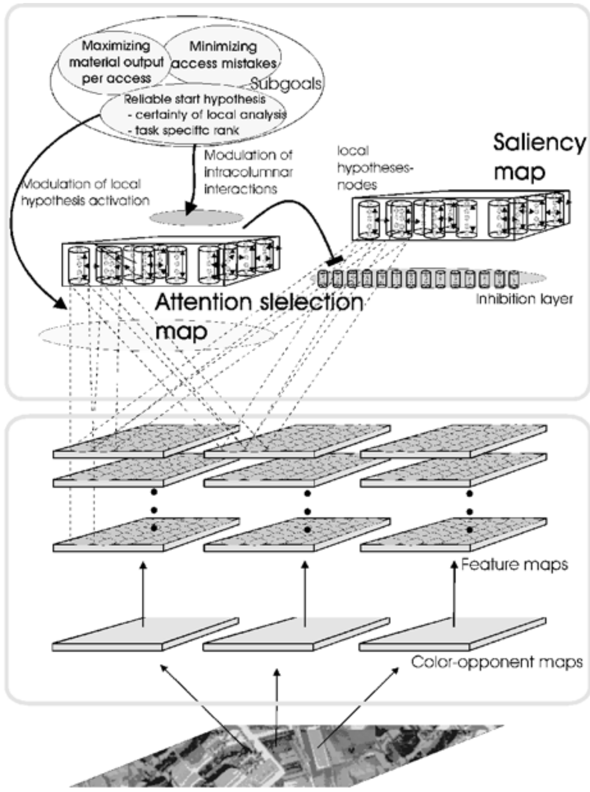

Fig. 3: The selection of textured objects. The proposed architecture comprises image acquisition, a conversion into 3 colour-opponent maps, a feature extraction on these maps, a generation of a hypothesis map, the relaxation process in an attention selection map modulated by different subgoals and the formation of a region of attention in a saliency map.

forms the input space of columnar organized hypothesis networks that determine the activity of local hypothesis nodes (fig. 3). To fulfill our requirements of fast learning and optimal separation, we established a supervised incremental network [2] for the generation of local hypotheses.
The hypothesis nodes in the attention selection map are excitatory coupled with their corresponding neighbours and with inhibition nodes, which initiate a global inhibitory feedback. In dependence on the subgoals lateral cooperation and global inhibition result into the formation of a winner assembly [5]. Changing subgoals lead to another assembly, and as a consequence another behaviour emerges. Though competition and the selection of a region in one relaxation network has the advantage of a compact architecture [6] the emergence of only one region can not guaranteed. A reliable formation of only one region of attention is performed in the saliency map after the competition dynamics in the attention selection map. The saliency network is similar to the attention selection network, but the interactions are not modulated by subgoals.

3.2. Subgoal controlled competition

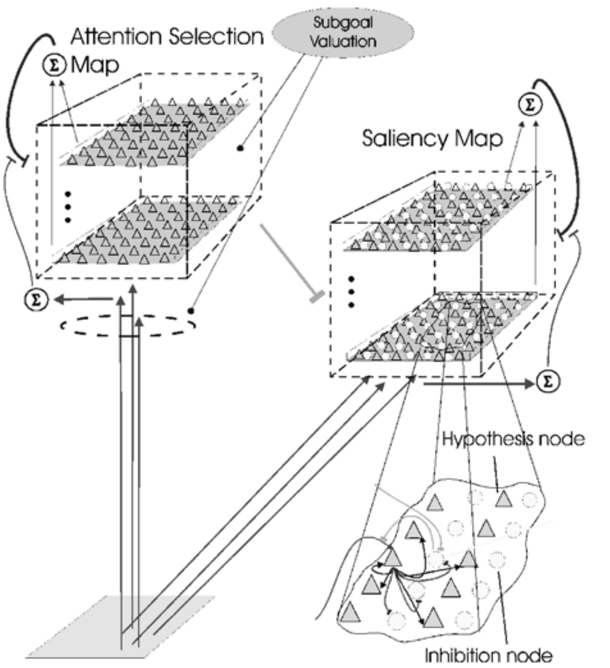

Fig. 4: The architecture of the relaxation networks consists of local excitatory interactions, global inhibition and highly specific modulations from subgoals.

Local hypotheses modulated by a subgoal "valuation", important for preferring certain hypotheses, feed the relaxation process in the attention selection map (fig. 4). The number of hypothesis layers is identical with the number of local hypotheses. The output y_{ki} of the hypothesis node i in column k is calculated from the internal activity of the neuron z_{ki} with a nonlinear saturation function.

$$y_{ki} = \varphi(z_{ki}); \qquad \varphi(z) = \begin{cases} z & for\ 0 \prec z \le z_{\max} \\ z_{\max} & for\ z \succ z_{\max} \\ 0 & for\ z \le 0 \end{cases} \tag{1}$$

The differential equation of the internal activity dynamics consists of different terms determined by the interaction in the network, the subgoals and the input:

$$\tau \cdot \dot{z}_{ki} = -\mathrm{f}(z_{ki}) + (I_{ki}^{a} + I_{ki}^{b} + I_{ki}^{s}) \tag{2}$$

Each hypothesis node is connected excitatory with its nearest neighbours within the same layer. A mutual support of different, similar hypotheses is optional and can be realized with excitatory weights between layers. These local intra- and intercolumnar interactions are considered in I^b:

$$I_{ki}^{\mathrm{b}} = \sum_{j \in \Omega, l \in \Psi} y_{lj} \cdot w_{ki\,lj} \qquad for\ l \neq k \tag{3}$$

For cooperation only within a hypothesis layer the set Ω consists solely of $\Omega = \{i\}$. Ψ denotes the set of excitatory coupled neighbours in the field. Recurrent excitatory connections from each node to itself strengthen considerable its activity and the local hypothesis performs a permanent input during the relaxation process:

$$I_{ki}^{\mathrm{a}} = w_{ki\,ki} \cdot B_i \cdot y_{ki} + h_{ki} \cdot B_i \tag{4}$$

In this equation, B_i denotes the “valuation” of the different hypotheses concerning their importance for the task. The input activity h_{ki} of the hypothesis nodes terms the security of the system to find the hypothesis in its receptive field.
Competition is induced by inhibitory feedback. In order to prevent the occurrence of a region in case of a too weak input, the feedback is modulated by a presynaptic inhibition, which is expressed by the division of the modulated input activity h_{ki} :

$$I_{ki}^{\mathrm{s}} = -\frac{1}{\sum\limits_{l,j} B_j \cdot h_{lj}} \sum_{l,j} y_{lj} \tag{5}$$

In extension to merely one competition, this algorithm is able to transform the information of a scene into a sequential order, a decision sequence, relevant for the actual task by inhibition of the preceding winners in additional cycles.

3.3. Emergence of a focus of attention

In order to grasp an object from the conveyor belt the emergence of a hypothesis, equivalent with a label, and the point of access is sufficient. For further analysis of the object the saliency map forms a definite region of attention on basis of the hypothesis map.
The inhibition nodes which prevent the information flow from the hypothesis map to the saliency map is suppressed from the winner in the attention selection map with preservation of its place (fig. 4). The activation from the hypothesis map spreads out in the saliency map, but is restricted to the activation boundaries. Thus, an assembly of active nodes in the saliency map is linked to the actual most interesting visual object in the scene.
The output y_{ki} of the hypothesis node i in column k of the saliency map is also calculated from the internal activity of the neuron z_{ki} (1) but with a lower saturation z_{max}.
For the internal activity dynamics a leaky integrator is sufficient. The interactions are only driven by the input:

$$\tau \cdot \dot{z}_{ki} = -z_{ki} + (I_{ki}^{b} + I_{ki}^{s}) \tag{6}$$

Each hypothesis node is also connected excitatory with its nearest neighbours within the same layer and with itself.

$$I_{ki}^{\mathrm{b}} = \sum_{j \in \Omega, l \in \Psi} y_{lj} \cdot w_{ki\,lj} \tag{7}$$

To prevent a global spread out of the activity, competition is induced by inhibitory feedback:

$$I_{ki}^{\mathrm{s}} = -\frac{1}{\sum\limits_{l,j} h_{lj}} \sum_{l,j} y_{lj} \tag{8}$$

Each hypothesis node suppresses neighbouring inhibition nodes to gate the activation form the input into the saliency map (fig. 4). To demonstrate the performance, we used the architecture in fig. 3 for region selection in a cluttered scene of three groups: newspapers, journals and cardboard. Our hypothesis network is teached with only two presentations of features gained form three images of each class (fig. 5). We performed three simulations on the same input scene (fig. 6) with changing subgoals "size", "valuation" and "security". With regard to the chosen subgoals another region emerges.

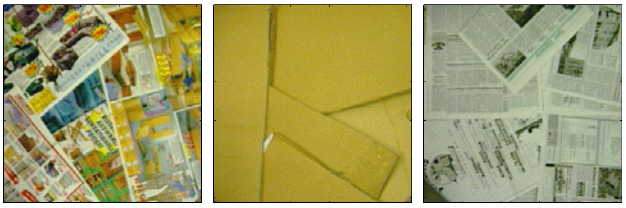

Fig. 5: An example of teach images concerning to classes of journal, cardboard and newspaper.

The local continuous output of the columnar organized hypothesis networks, denoted as the local hypothesis, terms the security to find the proposed material in its receptive field. On a global view, the local hypotheses form a kind of hypothesis scenery with hills and valleys (fig. 8).
The following results demonstrate the robustness of forming and selecting regions according to different subgoals in an image containing journals and cardboard (fig. 6).

- **First experiment: Influence of subgoals "security" and "size"**

In our first experiment the "valuation" of the different hypotheses is equal. Thus, the size of a coherent region and the security of the local hypotheses are decisive for the competition.
Despite the larger size of cardboard, the higher security of journal hypotheses (fig. 8) ensures the journal region to win the competition (fig. 9 and 7).

Fig. 6: Scene composed of cardboard and journal.

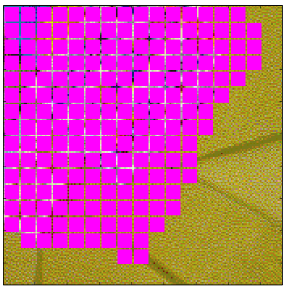

Fig. 7: Result: Scene with selected region.

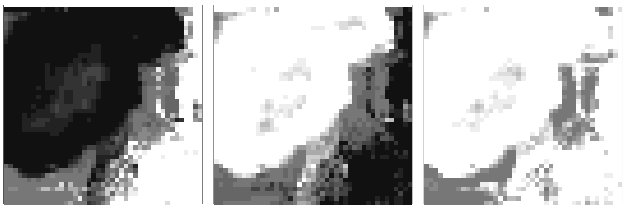

Fig. 8: Hypothesis maps of journal, cardboard and newspaper. Strong hypotheses are indicated black.

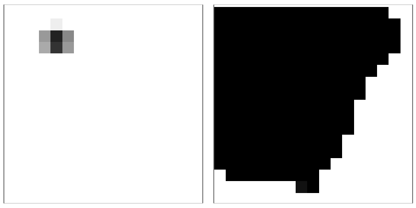

Fig. 9: All other activities despite of those shown in the above maps were suppressed. Left: Result of the selection process in the attention selection map for journal. Right: Emergence of a region for jounal in the saliency map.

- **Second experiment: Decreasing subgoal "valuation"**

In reducing the "valuation" of journals in this experiment, the cardboard hypothesis is successful (fig. 12 and 11). This means, the hypothesis journal is now, despite its high security, not worth enough to look at in detail and therefore not a candidate for grasping.

Fig. 10: Scene composed of cardboard and journal.

Fig. 11: Result: Scene with selected region.

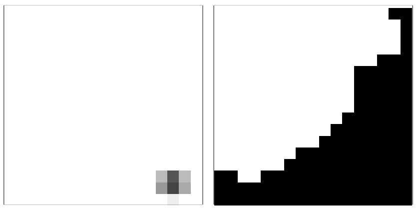

Fig. 12: All other activities despite of those shown in the above maps were suppressed. Left: Result of the selection process in the attention selection map for cardboard. Right: Emergence of a region for cardboard in the saliency map.

- **Third experiment: Decreasing subgoal "security"**

If the performance of the classification task for journals was lower, in contrast to the first simulation, the hypothesis of journals might not win the competition once more. To demonstrate this, we reduced the output of the classification network for journals to ⅔ of its original activity (fig. 15). Now in case of the same valuation, cardboard is the winner, because the main advantage of the journal hypotheses, the high security, is reduced.

Fig. 13: Scene composed of cardboard and journal.

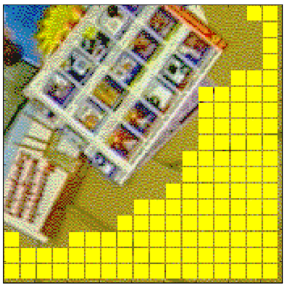

Fig. 14: Result: Scene with selected region

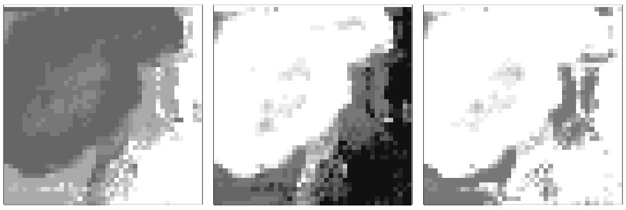

Fig. 15: Hypothesis maps with reduced activity of the hypothesis journal.

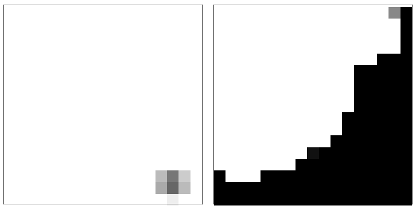

Fig. 16: All other activities despite of those shown in the above maps were suppressed. Left: Result of the selection process in the attention selection map for cardboard. Right: Emergence of a region for cardboard in the saliency map.

Conclusion

Our introduced model is part of a visuo-motory system and serves as mechanism of attention control for grasping relevant objects. The emergence of a region and the decision is made in dependence on defined subgoals of a task. Thus, our approach goes beyond a mere segmentation, it combines perception with action selection.

In a technical system the subgoals can be determined by an operator. Future work will focus on biological relevant acquisition and representation of subgoals. Although the influence of internal state, motivation and experience in controlling selective attention is not clear, there are some hints that the frontal cortex plays a major role in suppressing the irrelevant and enhancing the relevant sensory information [3]. The result in the saliency map depends on the quality of the hypothesis map. An improvement of feature extraction is expected to lead to a better hypothesis map and as a consequence to more exact regions (compare the too large hypothesis map of journal in fig. 8 with the image in fig. 6).

References

[1] Allport, A.: *Selection for action: Some behavioral and neurophysiological considerations of attention and action*. In: Perspectives on Perception and Action. Lawrence Erlbaum Associates: H. Heuer, A. F. Sanders, pp. 395-419, 1987.

[2] Fritzke, B.: *A growing neural gas network learns topologies*. Advances in Neural Information Processing Systems, vol. 7 (1995).

[3] Fuster, J. M.: *The Prefrontal Cortex. Anatomy, Physiology, and Neuropsychology of the Frontal Lobe.* Second Edition - New York: Raven Press, 1989.

[4] Grossberg, S.; Mingolla, E.: *Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations.* Percep Psychophys, vol. 38 (1985), pp. 141-171.

[5] Hamker, F.; Gross, H.-M.: *Region finding for attention control in consideration of subgoals*. Neural Network World, 6 (1996) no.3, pp.305-313.

[6] Hamker, F.; Gross, H.-M.: *Region selection: segmentation, classification and task relevance in a single grouping mechanism.* Proceedings of the ICNN’96, Washington, 1996. To appear.

[7] Hummel, R. A.; Zucker, S. W.: *On the foundations of relaxation labelling processes.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-5 (1983), pp. 267-287.

[8] Kienker, P. K.; Sejnowski, T. J. Hinton, G. E.; Schumacher, L. E.: *Separating figure from ground with a parallel network.* Perception, vol. 15 (1986), pp. 197-216.

[9] Mozer, M. C.; Zemel, R. S. Behrmann, M.; Williams, C. K. I..: *Learning to segment images using dynamic feature binding.* Neural Computation, vol. 4 (1992), pp. 650-665.

[10] Rosenfeld, A; Hummel, R. A.; Zucker, S. W.: *Scene labeling by relaxation algorithms.* IEEE Transactions on Systems, Man and Cybernatics, vol. 6 (1978), pp. 267-287.

[11] Worth, A. J.; Lehar, S.; Kennedy, D. N.: *A recurrent cooperative/competitive field for segmentation of magnetic resonance brain images.* IEEE Transactions on Knowledge and Data Engineering, vol. 4 no. 2 (1992), pp. 156-161.