Proc. of ICNN'96, IEEE-Intern. Conference on Neural Networks 1996, Washington, DC, USA, pp. 1540-1545, IEEE Press 1996

Region Selection: Segmentation, Classification and Task Relevance in a single grouping mechanism*

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

We introduce an approach for the fusion of segmentation, classification and examination of task relevance into a grouping mechanism performed by a competitive neural relaxation network. This means, information extracted from the environment is selected according to the relevance of the systems intended action. Due to the fact of a task-specific focus of attention, we avoid the separation of perception and generation of behavior.
Our network for the selection of action relevant visual regions consists of interacting columns with local excitatory and global inhibitory coupled feedback. The lateral cooperation is used as a way to integrate task pertinent subgoals. Possible subgoals are the size of regions, the security of a classification hypothesis and the valuation of the hypothesis for the task. The input activity, received from different hypothesis-layers, evokes several activation areas, which compete in a few iteration cycles. When equilibrium is attained cooperating neurons in one layer remain active, others are suppressed with regard to the relevant subgoals. The performance is demonstrated on a real-world selection of textured objects for a robot grasping task.

1. Introduction

An important issue in image processing is the selection of regions, no matter if in recognizing textures (e.g. [12] [10]) or in the segmentation of color (e.g. [3]). Often morphological operations on segmented and classified images, or a region representation based on the Quadtree approach [13] had to be performed. But the conflict between image segmentation and classification is still unsolved, it is often regarded as a circular problem: Image segmentation delimits different regions without labelling these and classification assigns areas a defined label.
Furthermore, conventional artificial visual systems are often separated between recognition and generation of behavior. They are usually based on sensory representations without pertinence to the intended behavior of an agent. These approaches have to interpret the visual data by a special control structure (a "homunculus") 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 merges the circular problem into a non-circular and which avoids the separation of perception and the generation of action by implicitly selecting only visual regions significant for the systems behavior and action at this moment (fig. 1). 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.

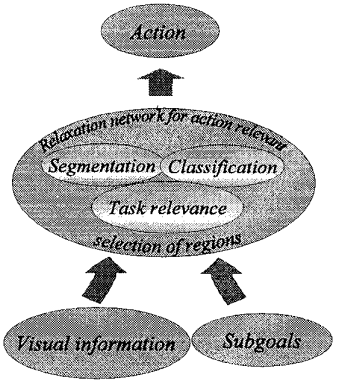

Fig. 1: Region selection merges three tasks into a single process.

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

Proc. of ICNN'96, IEEE-Intern. Conference on Neural Networks 1996, Washington, DC, USA, pp. 1540-1545, IEEE Press 1996

2. System Description

Short range excitatory connections and long range inhibitory connections constitute the typical structure of competitive networks [2] [5] [9]. This concept induced models described as relaxation labelling [7], and which are known for their segmentation properties [8] [11] [14] [6]. In extension to these approaches, we use the lateral cooperation as a way to integrate subgoals. In so far, this approach bridges the gap between sensory representation and motor representation by means of task relevant region selection.

2.1. System Overview

This paper demonstrates, how our relaxation approach can be used to enhance relevant areas in a cluttered scene composed of different textured objects, important for a robot grasping task (fig. 2).

The architecture can be subdivided into a feature extraction and a relaxation part. For feature extraction, the color image taken from a scene is converted into three color-opponent maps (black-white, blue-yellow and red-green) to achieve an increase of information for colored objects. On basis of these color-opponent maps, a topological feature space is spanned, which forms the input space of columnar organized hypothesis-networks which determine the activity of local hypothesis-nodes (fig. 2).

The kind of feature extraction is not subject of this issue, but for completeness it should be remarked that we determine the fractal dimension within a multiresolution pyramid. Thus, we get a multi-dimensional input for every local operating hypothesis-network. To fulfill our requirements of fast learning and optimal separation, we established a supervised incremental network [4] for the generation of local hypotheses.

Strictly speaking, the relaxation part starts after the generation of local hypotheses. The hypothesis-nodes in the relaxation plane are excitatory coupled with their corresponding neighbors and with inhibition nodes which initiate a global inhibitory feedback. Cooperation within the hypothesis-layer strengthens groups of the same hypothesis and inhibition between the hypothesis-layers guarantees that only one hypothesis, i.e., usually one region, remains active. In dependence on the subgoals, another region and as a consequence, another behavior emerges.

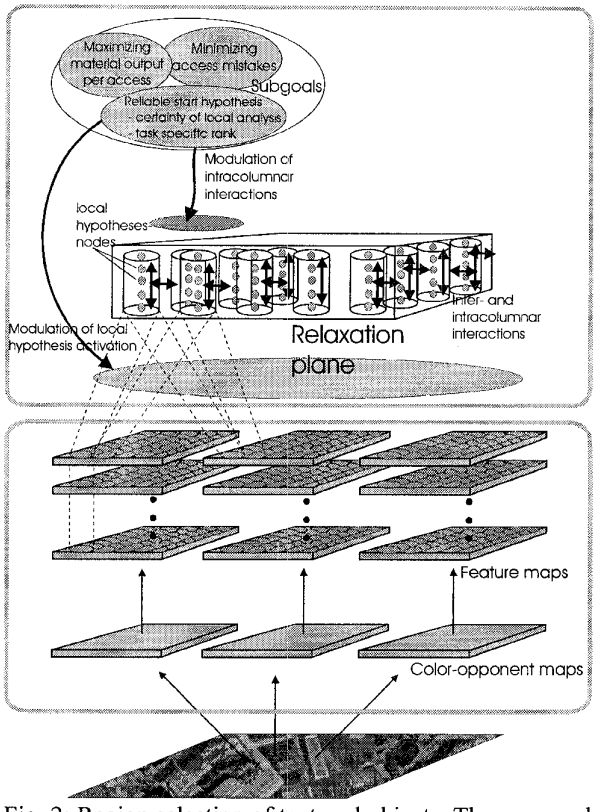

Fig. 2: Region selection of textured objects. The proposed architecture comprises image acquisition, a conversion into 3 color-opponent maps, a feature extraction on these maps, a generation of a hypothesis-map and the relaxation on this map modulated by different subgoals.

2.2. Relaxation Architecture

Local hypotheses modulated by a subgoal "valuation", important for preferring certain hypotheses, feed the relaxation process (fig. 3). The number of hypothesis-layers is identical with the number of local hypotheses. Each hypothesis-node is connected excitatory with its nearest neighbors within the same layer. A mutual support of different, similar hypotheses is optional and can be realized with excitatory weights between layers. Competition is induced by inhibitory feedback. For a flexible determination of competitors, the activity of each layer is summed up and inhibitory distributed into the layers by a distribution function. In order to prevent the occurrence of a region

Proc. of ICNN'96, IEEE-Intern. Conference on Neural Networks 1996, Washington, DC, USA, pp. 1540-1545, IEEE Press 1996

in case of a too weak input, the feedback is modulated by a presynaptic inhibition.
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.

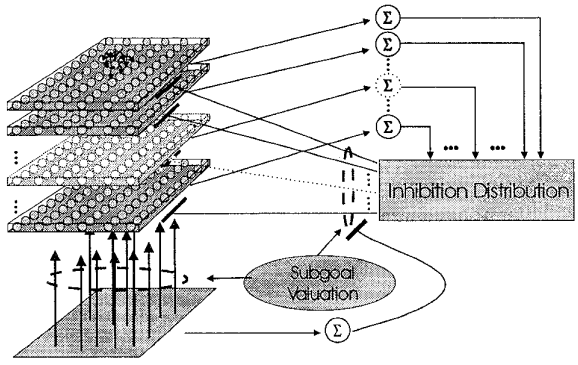

Fig. 3: The architecture of the relaxation network consists of local excitatory interactions, global inhibition and highly specific modulations from subgoals.

2.3. Network Dynamics

The output y_{ki} of the hypothesis-node i in column k is calculated from the internal activity of the node z_{ki} with a nonlinear saturation function.

$$y_{ki} = \varphi(z_{ki}); \qquad \varphi(z) = \begin{cases} z & for\ 0 < z \leq z_{\max} \\ z_{\max} & for\ z > z_{\max} \\ 0 & for\ z \leq 0 \end{cases} \tag{1}$$

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

$$\tau \cdot \dot{z}_{ki} = -z_{ki} + (I^a_{ki} + I^b_{ki} + I^s_{ki}) \tag{2}$$

The local intra- and intercolumnar interactions are considered in the following term:

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

For cooperation only within a hypothesis-layer the set Ω consists solely of $\Omega = \{i\}$. Ψ denotes the set of excitatory coupled neighbors 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:

$$I^a_{ki} = w_{kiki} \cdot B_i \cdot y_{ki} + h_{ki} \cdot B_i \tag{4}$$

In this equation, B_i stands for the "valuation" of the different hypotheses concerning their importance for action. The input activity h_{ki} of the hypothesis-nodes terms the security of the system to find the hypothesis in the analyzed area. The global inhibitory feedback, in which β allows to determine the inhibition between the hypothesis-layers, is calculated in consideration of the input activity h_{lj} of all nodes in the hypothesis-layers:

$$I^s_{ki} = -\frac{1}{\sum_{l,j} h_{lj}} \sum_{l,j} \beta_{ij} \cdot y_{lj} \tag{5}$$

3. Results

To demonstrate the performance, we used the architecture in fig. 2 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. 4). We performed three simulations on the same input scene

Proc. of ICNN'96, IEEE-Intern. Conference on Neural Networks 1996, Washington, DC, USA, pp. 1540-1545, IEEE Press 1996

(fig. 5) with changing subgoals "size", "valuation" and "security". With regard to the chosen subgoals another region emerges.
As explained, we transform our RGB-image into three color-opponent maps (fig. 7). The feature extraction on these maps forms the input space of the columnar organized hypothesis-networks. The local continuous output of these 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 mountain scenery with hills and valleys (fig. 8). The region selection algorithm optimizes the representation with regard to the intended task by means of local interactions and global inhibition.The following results demonstrate the robustness of forming and selecting regions according to different subgoals in an image containing journals and cardboard (fig. 5).

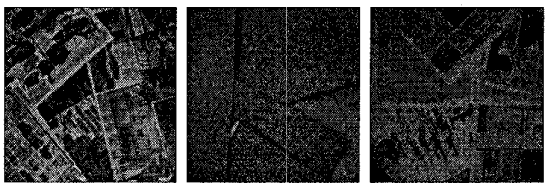

Fig. 4: An example of teach images concerning to classes of journal, cardboard and newspaper.

• First experiment

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 ensures the journal region to win the competition (fig. 9 and 6).

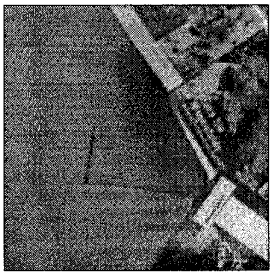

Fig. 5: Color scene composed of cardboard and journal.

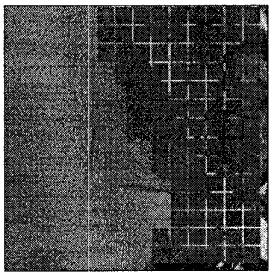

Fig. 6: Color scene with selected region.

Fig. 7: Color-opponent maps (red-green, blue-yellow and black-white).

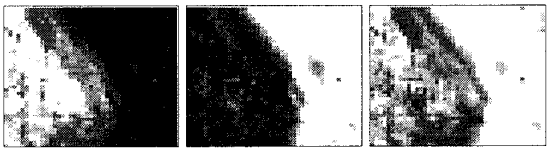

Fig. 8: Hypothesis-maps of journal, cardboard and newspaper. Strong hypotheses are indicated black.

Fig. 9: Emergence of regions in hypothesis-maps.

• Second experiment

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.

Proc. of ICNN'96, IEEE-Intern. Conference on Neural Networks 1996, Washington, DC, USA, pp. 1540-1545, IEEE Press 1996

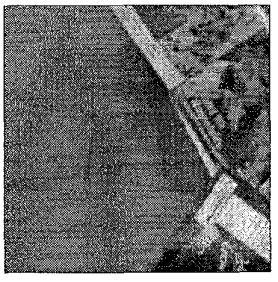

Fig. 10: Scene composed of cardboard and journal.

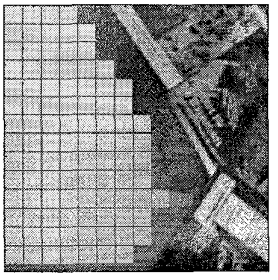

Fig. 11: Color scene with selected region.

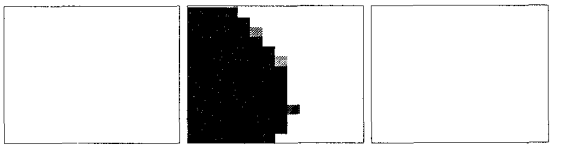

Fig. 12: Emergence of regions in hypothesis-maps.

• **Third experiment**

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 ¾ 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.

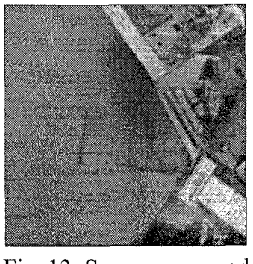

Fig. 13: Scene composed of cardboard and journal.

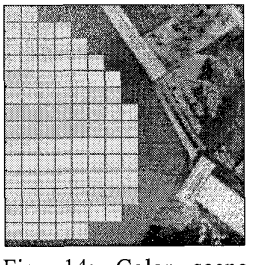

Fig. 14: Color scene with selected region

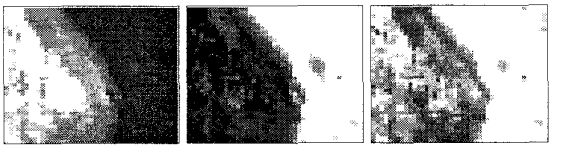

Fig. 15: Hypothesis-maps with reduced activity of the hypothesis journal.

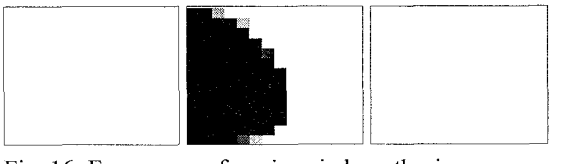

Fig. 16: Emergence of regions in hypothesis-maps.

Conclusion

The introduced model of region selection is part of a Robot-Vision system and serves as an attentional mechanism for grasping relevant regions. The emergence of a region and the decision is made without explicit rules and in dependence on defined subgoals of a task.
Thus, our approach goes beyond a mere segmentation, it combines perception with action.
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 (fig 17).
The result of the region selection mechanism depends on the quality of the hypothesis-maps. An improvement of feature extraction is expected to lead to better hypothesis-maps and as a consequence to more exact regions (compare the hypothesis-maps in fig. 8 with the image in fig. 5).
For further work our aim is to use this model for forming a

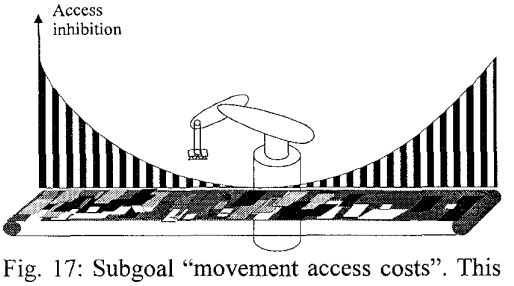

Fig. 17: Subgoal "movement access costs". This subgoal can be established by an inhibition in dependence on the distance to the manipulator.

sequence of interesting regions in a large image. Therefore, an extension of this algorithm to a dynamic linked grouping may be helpful to realize a stronger competition among assemblies of the same hypotheses.

Appendix

Parameters of the above mentioned equations are given here. In all simulations the local excitatory connections remained the same and are shown in figure 18. Global inhibitory weights are chosen in all simulations to:

0	0	0	0	0	0	0
0	0	0.2	0.3	0.2	0	0
0	0.2	0.8	0.9	0.8	0.2	0
0	0.3	0.9	0.3	0.9	0.3	0
0	0.2	0.8	0.9	0.8	0.2	0
0	0	0.2	0.3	0.2	0	0
0	0	0	0	0	0	0

Fig. 18: Weights ($w_{ki\,lj}$) from the central node ki to itself and to its neighbors lj.

Due to a limited memory capacity, the resolution of the relaxation plane is reduced compared to the resolution of the hypothesis-maps - the block effect is not part of the algorithm.

$$\beta_{ki} = 360 \quad \textit{for } i \neq k$$
$$\beta_{kk} = 30$$

Parameters in the first and in the third experiment:

$$B_1 = 1 \quad B_3 = 1$$
$$B_2 = 1$$

Parameters second experiment:

$$B_1 = 0.8 \quad B_3 = 1$$
$$B_2 = 1$$

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] Amari, S.; Arbib, M. A.: *Competition and Cooperation in Neural Networks.* Berlin: Springer Verlag, 1982.

[3] Celenk, H.: *A color clustering technique for image segmantation.* Comp. Vision, Graphics and Image Processing, vol. 52 (1990), pp. 145-170.

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

[5] Grossberg, S.: *Studies of mind and brain.* Dordrecht: D. Reidel, 1982.

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

[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] Kohonen,T.: *Assoziative Memory - A system theoretic approach.* Berlin: Springer Verlag, 1977.

[10] Mosquera, A.; Cabello, D.; Carreira, M. J.; Penedo, M. G.: *Texture image segmentation using a modified hopfield network.* New Trends in Neural Computation, IWANN' 93, 1993, pp. 657-663.

[11] 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.

[12] Pachowicz, W.: *Semi-autonomous evolution of object models for adaptive object recognition.* IEEE Transactions on Systems, Man, and Cybernatics, vol. 24 no. 8 (1994), pp. 1191-1206.

[13] Samet, H.: *A tutorial on quadtree research.* In: Multiresolution image processing and analysis. Springer Series in Information Sciences. Vol. 12. Berlin: Springer: A. Rosenfled, pp. 212-223, 1984.

[14] 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.