Proc. ICANN'97, Lausanne October 1997, pp. 919-924, Springer-Verlag 1997

Object Selection with Dynamic Neural Maps*

Fred Henrik Hamker, Horst-Michael Gross
Technische Universität Ilmenau, Neuroinformatik, D-98684 Ilmenau, Germany
http://cortex.informatik.tu-ilmenau.de/~fred
e-mail: fred@informatik.tu-ilmenau.de

Abstract. This contribution presents an approach for object selection. It is based on the functional role of attention in its control of action, that is the selection in space and the selection at the object level to serve the extraction of further object related features. The selection is performed on topographic maps which describe by their local activation the existence of an object hypothesis and is achieved by a dynamical two step process of i) competition and ii) region aggregation. The neural dynamics is based on analog neurons, mathematically described with a nonlinear activation dynamic. This method is appropriate for all objects which are detected by their local features, like color or texture. The performance is demonstrated on two real world selection problems.

1 Introduction

Object selection is a fundamental task in computer vision and robotics. Many algorithms like relaxation labeling [16] and neural algorithms based on competition and cooperation [17] are known for their segmentation capabilities. Nevertheless, active vision, does not require a segmentation of the whole scene. Instead, the selection of only one region or object is decisive for further processing. Especially in the theory of animate vision where vision is regarded as a pointer into the world, a robust algorithm that binds objects in the world to cognitive programs is needed [3]. Thus, selection is often discussed in theories of selective attention. From the functional point of view selection is the extraction of information relevant for the current task [1] [11], or more general, a mechanism for object related parameter specification [12]. Parameters needed in visuo-motor systems might be the location, size and orientation of objects or their features like color or texture. Attentional networks based on the spotlight metaphor [13] [14] have the disadvantage that the attended region of the spotlight is not identical with the winning object. In order to aid further image analysis or learning, selection must be at the object level [15] emerging from a competitive process in parallel across the whole visual field [5]. The presented approach follows this functional rule. A neural competition/ cooperation process selects a potential object on basis of an initial object hypothesis derived from a pre-attentive grouping [6] or from a neural voting of local features like color or texture. After the competition is finished, a neural region aggregation forms an object focus and segments the object from other hypotheses beginning from the seed region emerged from the competition/cooperation process.

---

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

2 System Architecture

The input of the architecture are initial hypotheses about potential objects in the image (Fig. 1). Similar to the segmentation algorithm of Worth et al [17], but on the object level, this network demands at each location a kind of probability for each hypothesis as a result of a pre-attentive process and a following local vote, e.g. the continuous output of a neural classifier. By modifying the activation of the HM according to the subgoals of a system, a goal directed selection can be achieved [8]. In contrast to known segmentation algorithms, the selection of an object is carried out in a two stage process. In the first stage, it is decided where and to what attention should be directed, which is achieved by a competition/cooperation process. This should guarantee that only one object in the image is selected according to the pointing nature of attention. Because of their limited WTA capabilities, one stage models, e.g. [7], do not ensure this. Explaining the result of the selection stage by the pointer analogy, the emerged delimited field of active nodes represent a pointer coupled with the parameter hypothesis directed to a defined location in the image. The second stage generates the object focus, which can be used to extract further object related parameters. This is again achieved by a competition/ cooperation process, but here the input of the map is induced from the seed region emerged in the ASM by a presynaptic excitation of the HM from the FCM. The small focus grows in the FAM until no further input is strong enough to activate a node in the field. By inhibition of the detected region, the winner breaks down and the WTA process detects another object in the image until all objects are found.

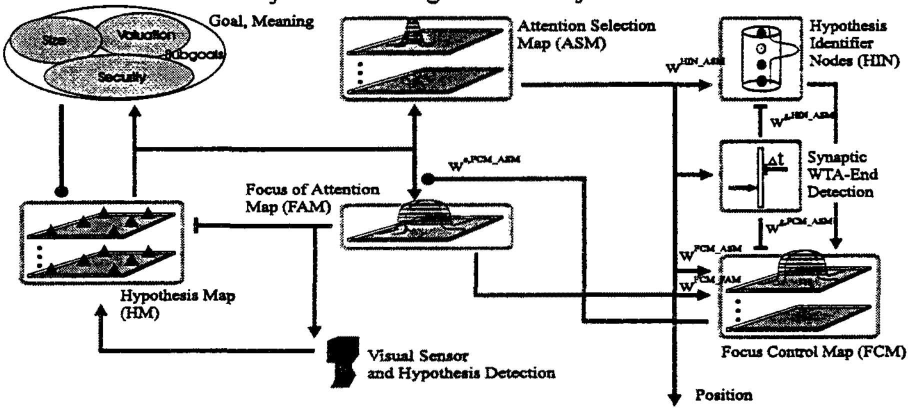

Fig. 1. The architecture consists of five maps. The neurons in the Hypothesis Maps (HM) code by their activation the location and the probability of an object in the image. In the Attention Selection Map (ASM) a WTA-process selects the potential object and its location. The information of what is in the image is recoded in the Hypothesis Identifier Nodes (HIN) after the end of the WTA-process. Finally, the region of the object emerges from the interaction of the Focus Control Map (FCM) with the Focus of Attention Map (FAM), according to a region aggregation process starting with the seed region detected in the ASM.

2.1 Neural Field Dynamics

The natural selection in attention can be viewed as a cooperative/competitive process [2] [10] and constitutes a basic and proven framework for selection. The dynamics of pattern formation in lateral-inhibition type neural fields are described by Amari [2]:

$$\tau \cdot \dot{z}(x,t) = -z(x,t) + \int w(x-x')f(z(x',t))dz' + \xi(x,t) + I(x,t) - \Theta \quad (1)$$

The transfer function f leads to a nonlinear system. w describes the kernel of recursive lateral feedback, ξ a stochastic term, I the external input and Θ a threshold.
Equation (1) is often demonstrated to show a good competitive behavior. Nevertheless, the neural interpretation of the term $-z(x,t)$ as a linear loss term of a neuron can be replaced by a non-linear loss term, which is biologically more plausible [10] and shows a very good WTA behavior [9].

$$s(z) = c \cdot \ln\left(\frac{z_{\max}+z}{z_{\max}-z}\right) \quad (2)$$

Using equation (2) as a nonlinear loss term, we get a neural map, in which each neuron is described by its average spike activity, which is limited to a maximum z_{max}. Because the average spike activity is only allowed for nonnegative values the differential equation is thresholded. The basic equation of each neuron in the dynamic maps reads:

$$\tau \cdot \dot{z}_{ki} = \Psi_{\min}\left(z_{ki},\ -s(z_{ki}) + \sum_{j\in\Omega,\, l\in\Upsilon} y_{lj}(t)\cdot w_{kilj} + I_{ki} - \Theta\right) \quad (3)$$

where Ψ_{min} is defined with a maximum operator and the Heaviside function H as:

$$\Psi_{\min}(x_1,x_2) = \max\{\mathrm{H}(x_1)\cdot x_2,\ x_2\} \quad (4)$$

The average spike activity z is equal to the output y of the node, no special transfer-function is needed.

2.2 Attention Selection (First Stage)

Eq. (5) shows the basic equation of each node in the ASM.

$$\tau \cdot \dot{z}_{ki}^{ASM}(t) = \Psi_{\min}\left(z_{ki}^{ASM}(t),\ -s(z_{ki}^{ASM}(t)) + I_{ki}^{ASM}(t) - \Theta^{ASM}\right) \quad (5)$$

The first input of a node in (6) denotes the local cooperation, determined by e.g. a Gaussian function of the distance. The local activity representing the probability of the hypothesis can be found in the second term. The third term denotes the feedforward global inhibition, modulated by the average hypotheses activity.

$$I_{ki}^{ASM}(t) = \sum_{j\in\Omega,\, l\in\Upsilon} y_{lj}^{ASM}(t)\cdot w_{kilj} + y_{ki}^{HM} - \frac{w_{inh}}{\sum_{l,j} w_{kilj}^{*}\cdot y_{lj}^{HM}}\, y^{GI}(t);\quad w_{kilj}^{*} = \frac{1}{L\cdot J}\quad \forall\, l,j \quad (6)$$

This global inhibition is described by an interneuron in (7).

$$\tau^{GI}\cdot \dot{z}^{GI}(t) = \Psi_{\min}\left(z^{GI}(t),\ -s(z^{GI}(t)) + \sum_{l,j} y_{lj}^{ASM}(t)\right) \quad (7)$$

A more detailed look at the third term in (6) reveals, that the higher the average hypotheses activity, the lower the global inhibition. This means, a low hypotheses activity leads to a strong inhibition and as a result no node is able to succeed, i.e. no attention emerges. Parameters influencing this behavior are w_{inh} and Θ.

2.3 Object Focus (Second Stage)

The information about what hypothesis is selected is represented in the HIN.

$$\tau \cdot \dot{z}_{i}^{HIN}(t) = \Psi_{\min}\left(z_{i}^{HIN}(t),\ -s(z_{i}^{HIN}(t)) + I_{i}^{HIN}(t)\right) \quad (8)$$

In (9), each node i represents a hypothesis. Because the superior hypothesis can only be extracted after the WTA process has been decided, the nodes i are inhibited by a term, which detects a change in the activity of the ASM nodes by comparing the original and the with T_d delayed output (WTA-end detection). When the significant inhibition from the WTA-end detection stops, the self excitation and mutual inhibition leads (first term in Eq. (9)) the node with the highest input, which represents the successful hypothesis in the ASM (second term in Eq. (9)) to achieve its maximum activation. All other nodes are suppressed. The threshold θ^{HIN} is important, because it prevents low activity in the ASM to spread all over the network. The operator Λ sets a negative argument to zero.

$$\begin{aligned} I_i^{HIN}(t) &= \sum_l w_{il} \cdot y_l^{HIN}(t) + \sum_k w^{HIN\_ASM} \cdot \Lambda\left(y_{ki}^{ASM}(t) - \theta^{HIN}\right) \\ &- \sum_k w^{g,HIN\_ASM} \cdot \left|y_{ki}^{ASM}(t) - y_{ki}^{ASM}(t - T_d)\right| \end{aligned} \tag{9}$$

An activation of nodes in a FCM layer is only possible if the corresponding HIN is active, otherwise the remaining input does not surpass the threshold θ^{FCM}.

$$\tau \cdot \dot{z}_{ki}^{FCM}(t) = \Psi_{\min}\left(z_{ki}^{FCM}(t),\ -s(z_{ki}^{FCM}(t)) + I_{ki}^{FCM}(t) - \theta^{FCM}\right) \tag{10}$$

The input of the FCM (11) comprises the activity of the ASM, the activity of the HIN, a diverge feedback from the FAM and the inhibition from the WTA-end detection, which again suppresses the activity until the WTA in the ASM has been finished.

$$\begin{aligned} I_{ki}^{FCM}(t) &= w^{FCM\_ASM} \cdot \Lambda\left(y_{ki}^{ASM}(t) - \theta^{FCM}\right) + y_i^{HIN}(t) + \sum_{l \in \Phi} w_l^{FCM\_FAM} \cdot y_l^{FAM}(t) \\ &- w^{g,FCM\_ASM} \cdot \sum_k \left|y_{ki}^{ASM}(t) - y_{ki}^{ASM}(t - T_d)\right| \end{aligned} \tag{11}$$

The activation loop comprised of FCM (10) and FAM (12) nodes controls the segmentation. The emerged seed region in the ASM activates FCM nodes which force the way for the HM activation into the FAM. A recurrent pathway from the FAM to the FCM enlarges the object focus.

$$\tau \cdot \dot{z}_k(t) = \Psi_{\min}\left(z_k(t),\ -s(z_k(t)) + I_k^{FAM}(t)\right) \tag{12}$$

The presynaptic excitation of the HM activity from the FCM is shown in the first term of (13), in which $\varphi(x)$ denotes a threshold function between zero and one, which is zero for arguments below zero and one for arguments exceeding one.

$$\begin{aligned} I_k^{FAM}(t) &= y_{ki}^{HM}(t) \cdot \varphi\left(\sum_i w^{e,FCM\_FAM} \cdot y_{ki}^{FCM}(t)\right) \\ &+ \sum_{l \in \Phi} w_{kl} \cdot y_l(t) - w^{FAM\_GI} \cdot y^{GI} \end{aligned} \tag{13}$$

Similar to the ASM, each node receives an input from its neighbors, but here this input is very low, it only bridges activity gaps in the HM.

$$\tau^{GI} \cdot \dot{z}^{GI}(t) = \Psi_{\min}\left(z^{GI}(t),\ -s(z^{GI}(t)) + \sum_k y_k^{FAM}(t)\right) \tag{14}$$

The global inhibition threshold activation defines the necessary input for a FAM node to become a part of the object focus. The stabilization of a pattern in FAM is detected by two nodes with different time constants. After detecting a stable pattern, all areas of the HM are inhibited by the emerged object region and another competition begins.

3 Results

The following simulations show results from two projects running at our institute. In most interacting systems, the detection of only one object in an image is of interest. But for completeness, the sequential selection of all attended regions is shown.

The first example, comes from a multisensorial robotic system for sorting [8], demonstrated in the field of waste paper recycling (Fig. 2). The hypotheses are generated by the extraction of local color information followed by a neural preclassification into journal, cardboard and newspaper.

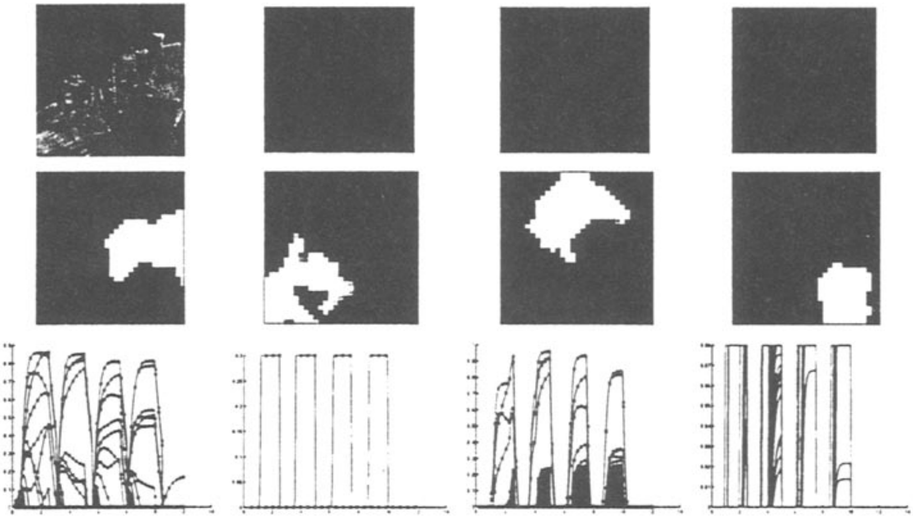

Fig. 2. Selection of objects on a conveyor-belt. *Top, left to right*: original image (256x256), journal, cardboard, and newspaper hypothesis (32x32). *Middle, left to right*: all sequential detected regions. *Bottom, left to right*: Activity over time in the ASM, HIN, FCM and FAM. The hypothesis images show a preclassification error for journal, which was classified as newspaper. The network selects all materials in the image, but not the misclassification, the activity of the ASM in this region increases but it is not strong enough to attract attention and the algorithm stops independently (bottom, left).

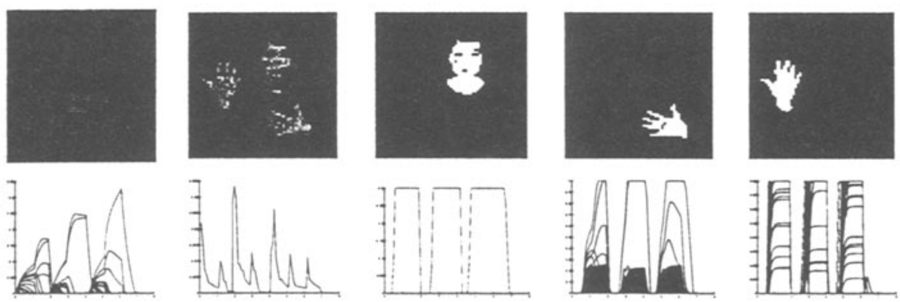

Fig. 3. Selection of human skin areas. *Top, left to right*: original image (256x256), skin hypothesis (64x64), all sequential detected regions. *Bottom, left to right*: Activity over time in the ASM, WTA-END, HIN, FCM and FAM. All skin areas are detected separately.

The second example comes from a project aimed at posture recognition to equip a mobile robot with the capability to react on human commands [4]. Therefore, a skin detector based on color assigns each pixel a skin hypothesis.

Conclusion

The proposed two step process of selection is demonstrated to fulfill the functional necessity of attention. The first step guarantees the selection of only one candidate and the second step realizes a description on the object level. Thus, our approach can serve as a powerful method for selection in active vision or generally for object specific information routing in visuo-motor systems.

References

[1] Allport, A.: *Selection for action: Some behavioral and neurophysiological considerations of attention and action.* Perspectives on Perception and Action., 395-419, 1987.

[2] Amari, S.: *Dynamics of pattern formation in lateral-inhibition type neural fields.* Biological Cybernatics, vol. 27 (1977), 77-87.

[3] Ballard, D. H.: *Deictic codes for the embodiment of cognition.* Behavioral and Brain Sciences, to appear 1997.

[4] Böhme, H.-J.; Groß, H.-M.; Braumann, U.-D.; Heinke, D.; Pomierski, T.; Brakensiek, A.: *Active-Vision zur Selbstorganisation von Verhalten in senso-motorischen Systemen.* In: Proc. Workshop "Aktives Sehen in techn. u. biol. Systemen", Hamburg , 60-66, 1996.

[5] Desimone, R; Duncan, J.: *Neural mechanisms of selective visual attention.* Anual Review of Neuroscience, vol. 18 (1995), 193-222.

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

[7] Hamker, F.; Gross, H.-M.: *Region selection: segmentation, classification and task relevance in a single grouping mechanism.* ICNN'96, Washington, 1540-1545, 1996.

[8] Hamker, F.; Gross, H.-M.: *Task Relevant Relaxation Network for visuo-motory Systems.* ICPR'96, Vienna, 406-410, 1996.

[9] Kaski, S; Kohonen, T.: *Winner-take-all networks for physiological models of competitive learning.* Neural Networks, vol. 7 (1994), 973-984.

[10] Kohonen, T.: *Physiological interpretation of the self-organizing map algorithm.* Neural Networks, vol. 6 (1993), 895-905.

[11] Milner, A. D.; Goodale, M. A.: *Visual pathways to perception and action.* Progress in Brain Research, vol. 95 (1993), 317-337.

[12] Neumann, O.: *Beyond capacity: A functional view of attention.* In: Perspectives on Perception and Action. Lawrence Erlbaum Associates, 361-393, 1987.

[13] Olshausen, B.; Anderson, C.; van Essen, D.: *A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information.* The Journal of Neuroscience, vol. 13, (1993), 4700-4719.

[14] Treisman, A.; Gelade, G.: *A feature-integration theory of attention.* Cognitive Psychology, vol. 12 (1980), 97-136.

[15] Yantis, S.: *Multielement visual tracking: Attention and perceptual organization.* Cognitive Psychology, vol. 24 (1992), 295-340.

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

[17] 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), 156-161.