Proc. ICANN'97, Lausanne October 1997, pp. 913-918, Springer-Verlag 1997

SAIM: A Model of Visual Attention and Neglect

Dietmar Heinke[1] and Glyn W. Humphreys[2]

[1] Technical University of Ilmenau, Div. of Neuroinformatics, D-98684 Ilmenau, Germany, email:dietmar.heinke@informatik.tu-ilmenau.de
[2] University of Birmingham, School of Psychology, Birmingham B15 2TT, UK

Abstract. This paper examines the reason for a particular impairment of cognitive functioning in brain-damaged patients called visual neglect. To achieve this goal a Selective Attention Identification Model (SAIM) was developed which performs translation-invariant object recognition. SAIM uses a constraint satisfaction routine based on a continuous Hopfield network to map an object into a focus of attention. The simulation results show that SAIM is a successful model of visual attention and visual neglect.

1 Introduction

There is a growing impact of neuropsychological findings on the understanding of the cognitive functioning of the brain. (e.g [1]). Neuropsychology is mainly concerned with cognitive deficits of brain-damaged patients. In this article we focus on examine the reason for a particular impairment called "visual neglect". The term "visual neglect" is used to refer to brain-damaged patients who fail to respond appropriately to stimuli presented on the side of space contralateral to their brain lesion. They fail to eat food on one side of an object, to cancel lines on one side of a sheet, to draw one half of an object or to read words on one side of a text. Classically, neglect is related to lesions of the right parietal lobe [4]. In order to examine how visual neglect might emerge following damage to an object recognition system, we developed a model called SAIM (Selective Attention Identification Model, Fig. 1), which aims at a translation-invariant object recognition [2] [5]. It does this by mapping from locations on a retina through to a smaller "attentional" window, the Focus of Attention (FOA), with activation within the FOA providing the input to an object recognition system. This approach is similar to the model of [7], which focused on anatomical issues, whereas our work concentrates on psychological and neuropyschological modelling.

2 The Network

The architecture of SAIM is illustrated in Fig. 1. It shows three different subnetworks: the contents network, the selection network and the knowledge network. These networks will be introduced in this section.

2.1 Contents Network

The contents network contains "sigma-pi" units that determine the activation values assigned to units in the FOA:

$$y_{ij}^{FOA} = \sum_{k=1}^{N} \sum_{l=1}^{N} y_{kl}^{VF} \cdot y_{ikjl} \qquad (1)$$

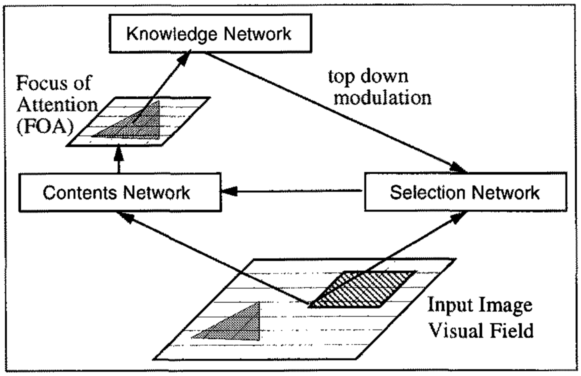

Fig. 1. Overview of SAIM. In order to achieve a translation-invariant object recognition, SAIM maps the visual field through to a smaller FOA. This mapping is performed by two networks: The contents network contains "sigma-pi" units, that determine the activation values assigned to units in the FOA by combining multiplicatively activation in retinal units with that in units in the selection network. There is one unit in the contents network for every unit in the FOA. The selection network determines which retinal units have their activation values mapped through to the FOA (via the contents network). Which retinal units come to be mapped through to the FOA is determined by process of mutual constraint satisfaction between units in the selection network, and mutual constrain satisfaction is in turn achieved by the network embodying certain constraints in its pattern of inter-connectivity. The knowledge network introduces knowledge about objects into SAIM and modulates the behaviour of the selection network in a top down way.

Here, y_{ij}^{FOA} is the activation of units in the FOA, y_{kl}^{VF} the activation of units in the visual field and y_{ikjl} the activation of units in the selection network. N is the size of the visual field. Every unit in the visual field can be mapped through to any FOA unit, if the appropriate unit in the selection network is active. In this way, the contents network enables a translation-invariant representation of the contents of the visual field.

2.2 Selection Network

The selection network (Fig. 2) determines which retinal units have their activation values mapped through to the FOA (via the contents network), as depicted in Fig. 1. Within the selection network are separate "control layers" of units, and the units in each control layer are activated by input from specific retinal locations and each control layer controls one unit in the contents network. Which retinal units come to be mapped through to the FOA is determined by process of constraint satisfaction between units in the selection network. The constraints aim at representing correctly the contents of the selected region in the visual field in the FOA. For this constraint satisfaction process the energy function approach by [3] is used, where minima in the energy function are introduced at a network state in which the constraints are satisfied. The constraints are:

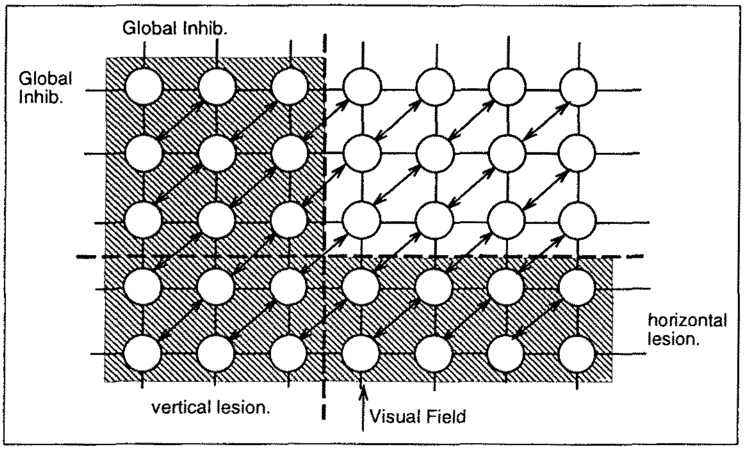

Fig. 2. Contents network. The topology derived from the energy function has three main feature: global inhibition between and within control layers and excitatory connections on the diagonal. Two type of lesioning are considered: The "vertical" lesion only affected the left part of the selection network and restricted the effect of the lesion to the left part of the visual field. The "horizontal" lesion affected inputs into the left side of the FOA for inputs across the visual field.

1. Only one unit in each control layer must be maximally active and all other units must have zero output activity. If this were not the case, there would be an overlap between the contents of retinal units in the FOA. Therefore, the maximally active unit has to suppress all other units through a global inhibition in a winner takes all fashion. The energy function for a WTA behaviour can be defined as:

$$E_{WTA}(y_i) = a \cdot (\sum_i y_i - 1)^2 - \sum_i y_i \cdot I_i \tag{2}$$

Where I_i is the input. This energy function is minimal, when all y_i's are zero except one y_i and corresponding I_i has the maximal value of all I_is (see [6]). This equation would force one unit of the control layer to be one; however there should be the possibility for attention not to be allocated to any location. Thus, a two state winner take all (WTA2) was introduced:

$$E^{(1)}_{WTA2}(y_{ikjl}) = \sum_{ij}(\sum_{k,l} y_{ikjl} - 1)^2 \cdot (\sum_{k,l} y_{ikjl})^2 \tag{3}$$

The multiplicational part of this equation introduces a second minimum into the energy function, where all units can stay zero and so attention is not directed to any location. This equation adds a threshold function to the WTA behaviour, whereby the output activity of the units stay zero as long as the input activity is below a certain level. The input to the WTA2 will be introduced below.

2. The same competition has to occur between corresponding columns of the control layers, because only one unit in each column must be maximally active, otherwise, the contents of one retinal unit is mapped more than one time onto

the FOA. This constraint leads to a WTA2 behaviour between control layers ($E^{(2)}_{WTA2}(y_{ikjl})$).

3. The selection network has to maintain the spatial relationships between retinal units in the FOA. Therefore, an appropriate energy function should have its minima, if units on the diagonal of the contents network are active:

$$E_{neighbour}(y_{ikjl}) = -\sum_{i,j,k,l}\sum_{s=-L}^{L}\sum_{r=-L}^{L} g_{sr} \cdot y_{ikjl} \cdot y_{i+r,k+s,j+r,l+s} \tag{4}$$

The coefficient g_{sr} introduces Gaussian weighting into the equation. In addition, this implements a simple form of perceptual grouping determined by the proximity of the elements.

4. For simplicity, the constraint for considering the visual input is implemented in the same way as the inputs in Eqn. 2.

2.3 Knowledge Network

In order to introduce knowledge into SAIM a simple template matching approach was used. Here, the match between templates and contents of the FOA was determined using a scalar product as a similarity measure:

$$I^{temp}_m = \sum_{I=1}^{M}\sum_{j=1}^{M} y^{FOA*}_{ij} \cdot w^m_{ij} \tag{5}$$

Where w^m_{ij}'s are the templates and M is the size of the FOA.
In the knowledge network the templates are formed from the connecting weights into template units (y^{temp}_m). A WTA is used to detect the best matching template. The same energy function as in Eqn. 2 was used with I^{temp}_m as input.

In order to get the complete energy function of SAIM which satisfies all constraints, one simply sums the different energy functions:

$$\begin{aligned} E_{total}(y^{temp}_m, y_{ikjl}) = {} & a_1 \cdot E^{(1)}_{WTA2}(y_{ikjl}) + a_2 \cdot E^{(2)}_{WTA2}(y_{ikjl}) + b_2 \cdot E_{input}(y_{ikjl}) \\ & + b_1 \cdot E_{neighbour}(y_{ikjl}) + b_3 \cdot E_{knowledge}(y^{temp}_m, y_{ikjl}) \end{aligned} \tag{6}$$

The coefficients of the different energy functions weight the different constraints against each other.

2.4 Topology and Lesion of Selection Network

The energy function introduced in the previous section defines minima at certain values of y_{ikjl} and y^{temp}_m, where they satisfy the constraints for a correct functioning of SAIM. To find these values a gradient descent was used as in [3]. This leads to the following differential equation system (for corresponding topology see Fig. 2):

$$\dot{x}_{ikjl} = -x_{ikjl} + d^{intr}_{ikjl} \cdot (-I^-_{ikjl} + b_1 \cdot I^+_{ikjl}) + b_2 \cdot y^{VF}_{kl} + b_3 \cdot I^{td}_{ikjl} \tag{7}$$

Where I^-_{ikjl} provides global inhibition, which comes from the WTA2 behaviour. I^+_{ikjl} is derived from the neighbourhood preserving behaviour. y^{VF}_{kl} comes from the visual field and I^{td}_{ikjl} is the top-down modulation from the knowledge network.

In order to simulate visual neglect, an additional factor d^{intr}_{ikjl} is introduced in order to lesion the model. This type of lesioning is called intrinsic lesioning, because it only involves intrinsic inputs from within the selection network. We report here effects of lesioning units in the selection network corresponding to inputs from the left visual field. [5] and [2] contain fuller reports in which several other forms of lesioning were examined.

3 Results

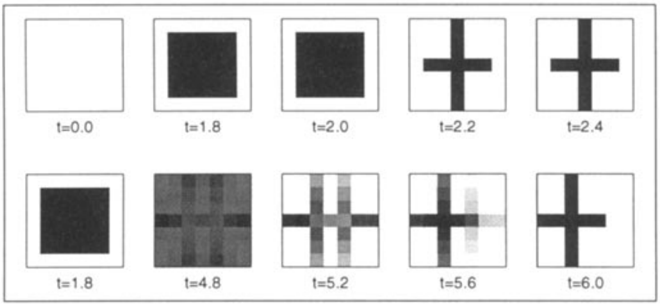

Fig. 3. Simulation of Posner Task. In the Posner Task [8] the reaction times of subjects vary as a function of whether a spatial precue matches the position of a subsequent target (with valid cues, relative to invalid cues where positions of the cue and the target do not match). This result is mimicked by SAIM. The time course of the FOA activity on the top shows the behaviour in the valid condition with a square as precue and a cross as target. The time course on the bottom shows the result in the non valid condition. Reaction times are faster with a valid cue.

In general, SAIM selects an object from the visual field by the criteria of size and the focus of attention falls at the centre of gravity of the object. This behaviour is mainly due to the interplay between the neighbourhood constraint and the WTA2 behaviour.

The Posner Task [8] is a classic experiment associated with selective attention and, its experimental data can be mimicked by SAIM (Fig. 3). In the Posner Task, patients with damage to the parietal lobe show particular problems in responding to invalid targets presented on the side of space contralateral to their lesion. This can be mimicked by SAIM as well [5]. The main reason in SAIM for these results is the storage of the activity in the dynamics of SAIM caused by the precue. In the lesioned version, the stored activity cannot be overcome by the activity of the target.

Fig. 4 shows the performance of the model after lesioning. When the cross is in the far left (lesioned) side, only activation from the right-side of the cross is mapped through to the FOA. Because of the lesion, the computed centre of gravity of the cross is shifted and the shape is project into the left side of the FOA. This result suggests that one of the reason for visual neglect is a distorted computation of the centre of gravity. This finding has to be tested in

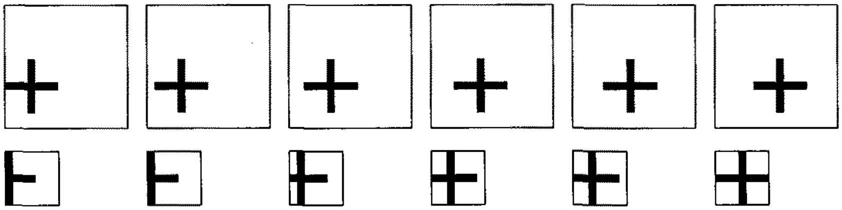

Fig. 4. Visual Neglect patients tend to neglect left parts of objects if they appear on the left[4]. This can be mimicked by SAIM. The row on the top shows the contents of visual field and the row on the bottom shows the resulting contents of the FOA. The lesion in SAIM was "vertical" (see text).

appropriate experiments. This neglect effect can be overcome by increasing the top down modulation of the knowledge network [2]. This result is well-known in neuropsychology, where additional knowledge about objects can compensate perceptual deficits [4].

4 Conclusion and Outlook

SAIM successfully simulates aspects of visual attention and neglect in human subjects. There remains some problems in scaling up the model, because the number of necessary units in the selection network increase quadratically with the number of input units and units in the FOA. [7] suggested a dynamic routing circuit in order to overcome the problem. This might be worth integrating into our approach.

References

1. J. Driver. Object Segmentation and visual neglect. *Behavioural Brain Research*, 71:135–146, 1995.
2. D. Heinke and G. W. Humphreys. SAIM: A connectionist model of selection for identification, in normality and pathology. in preparation.
3. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA*, 81:3088–3092, 1984.
4. Glyn W. Humphreys and M. Jane Riddoch. Attention to Within-object and Between-object Spatial Representations: Multiple Sites for Visual Selection. *Cognitive Neuropsychology*, 11(2):207–241, 1994.
5. G.W. Humphreys and D. Heinke. Spatial representation and selection in the brain: Neuropsychological and computational constraints. *Cognitive Neuroscience of Visual Attention, a special issue of Visual Cognition*, in press.
6. E. Mjolsness and Ch. Garrett. Algebraic Transformation of Objective Functions. *Neural Networks*, 3:651–669, 1990.
7. B. A. Olshausen and D. C. Andersen, C. H.and Van Essen. A multiscale dynamic routing circuit for forming size– and position–invariant object representations. *J. of Computational Neuroscience*, 2:45–62, 1995.
8. M.I. Posner, C.R.R. Snyder, and B.J. Davidson. Attention and the detection of signals. *J. Exp. Psychol.*, 109:160–174, 1980.