The Selective Attention for Identification model (SAIM): Simulating visual search in natural colour images

Dietmar Heinke, Andreas Backhaus, Yarou Sun, and Glyn W. Humphreys

Behavioural and Brain Sciences Centre
University of Birmingham
Birmingham B15 2TT, United Kingdom
email: d.g.heinke@bham.ac.uk
Phone: +44-121-41-44920

Abstract. We recently presented a computational model of object recognition and attention: the Selective Attention for Identification model (SAIM) [1–7]. SAIM was developed to model normal attention and attentional disorders by implementing translation-invariant object recognition in multiple object scenes. SAIM can simulate a wide range of experimental evidence on normal and disordered attention. In its earlier form, SAIM could only process black and white images. The present paper tackles this important shortcoming by extending SAIM with a biologically plausible feature extraction, using Gabor filters and coding colour information in HSV-colour space. With this extension SAIM proved able to select and recognize objects in natural multiple-object colour scenes. Moreover, this new version still mimicked human data on visual search tasks. These results stem from the competitive parallel interactions that characterize processing in SAIM.

1 Introduction

Recently, we presented a computational model, termed SAIM (Selective Attention for Identification Model; [1–7]). SAIM was developed to model normal attention and attentional disorders by implementing translation-invariant object recognition in multiple object scenes. In order to do this, a translation-invariant representation of an object is formed in a "focus of attention" (FOA) through a selection process. The contents of the FOA are then processed with a simple template-matching process that implements object recognition. These processing stages are realized by non-linear differential equations often characterized as competitive and cooperative interactions between neurons (e.g. [8]). With these mechanisms SAIM can model a wide range of experimental evidence on attention and its disorders (see [9] for a discussion). These results include: costs on selection from having multiple objects present [10], the facilitatory effects of object familiarity on selection [11], global precedence [12], spatial cueing both within and between objects [13, 14], and inhibition of return [15]. When simulated lesions were conducted, SAIM also demonstrated both unilateral neglect

and spatial extinction, depending on the type and extent of the lesion. Different lesions also produced view-centred and object-centred neglect [16], and both forms of neglect could even be simulated within a single patient (see [17] for evidence). It is beyond the scope of this paper to describe in details these experimental findings and how SAIM simulated the data. In essence, though, SAIM suggested that attentional effects in human behaviour resulted from competitive interactions in visual selection for object recognition, whilst neurological disorders of selection can be due to imbalanced spatial competition following damage to areas of the brain modulating access to stored knowledge.

However, in these simulations SAIM only processed black and white input images. This limitation questions the viability of SAIM as a general model for the human visual system. The aim of this paper is to demonstrate that SAIM is capable of processing natural images by adding an appropriate feature extraction while maintaining the central concepts of the model, such as filtering images through a focus of attention and using competitive interactions between stimuli to generate selection and recognition (see Fig. 1 for examples of the natural image used here). As these elements are essential for SAIM's earlier successes, it is likely that the new version presented here will still model the effects captured by previous versions.

There are other attentional models being capable of processing natural scenes,

Fig. 1. Examples of the natural colour images used in this paper.

most notably the saliency-based model by Itti and Koch [18–20]. The Itti and Koch model focuses on modeling behavioral data from visual search tasks. Visual search task is a commonly-used paradigm in attention research in which participants are asked to report the absence or presence of a specified target item amongst irrelevant items (distractors). The performance of the participants is measured in terms of time until response (reaction time). The number of distractors is varied across trials. The typical outcome of such experiments is a linear relation between reaction time and number of distractors. The slope of this linear relation varies with characteristics of the items in the search display and is often interpreted as an indicator for the underlying search mechanism (see [21], for a recent review). To model experimental evidences from visual search tasks, the Itti and Koch model computes a saliency map from an input image in three stages: In the first stage, early visual feature extraction, seven

types of feature maps are calculated: intensity contrast, red-green double opponency, blue-yellow opponency and four orientation selective maps based on Gabor-filters. In the second stage the maps of the three feature pathways (intensity, colour and orientation) are combined into three separate "conspicuity maps". The conspicuity maps show high activation at locations where untypical feature values are found, e.g. the image location of a tilted bar amongst vertical bars would receive a high activation value in the orientation conspicuity map. In the third stage the three conspicuity maps are linearly combined into the saliency map. The saliency map guides a serial search scan for the search target, with the scan starting at the location with the highest salient value and visiting locations in an order of descending saliency values. The content of each location is compared with the search target and if it matches, the search is terminated. In order to simulate experimental data from visual search [19] the number of serial search steps are related to human reaction times.
SAIM has also been shown to be able to simulate data from visual search tasks [2, 4]. In this case, search efficiency is determined by from interactions between SAIM's object recognition system with the competitive selection mechanisms. However, as with earlier versions of SAIM, these results were based on artificial black and white pixel images, lacking realistic properties of natural scenes. Moreover, these versions of SAIM did not possess biologically plausible feature extraction as is done by the saliency-based approach of Koch and Itti. The current paper presents an extension of SAIM which contains a biologically plausible feature extraction, and which uses a more flexible template matching process than before. We demonstrate that this extension is capable of mimicking results from visual search tasks with natural images as inputs.

2 SAIM

2.1 Overview

Figure 2 gives an overview of SAIM's architecture. In a first stage of processing, features are extracted from the input image. In an earlier version of SAIM only horizontal and vertical lines were used in the feature extraction [4]. The current version extends the feature extraction processes to include intensity, colour and orientation. This creates a more biologically plausible feature extraction, while, at the same time, it also allows SAIM to process successfully natural images, as we will show here. The contents network then maps a subset of the features into a smaller Focus of Attention (FOA). This mapping of the contents network into the FOA is translation-invariant and is gated by activity from all image locations competing through the selection network to gain control of units in the FOA. The selection network controls the contents network by competitive interactions between its processing units, so that input from only one (set of) locations is dominant and mapped into the FOA. At the top end of the model, the knowledge network identifies the contents of the FOA using template matching. The knowledge network also modulates the behaviour of the selection network with top-down activation, with known objects preferred over unknown objects.

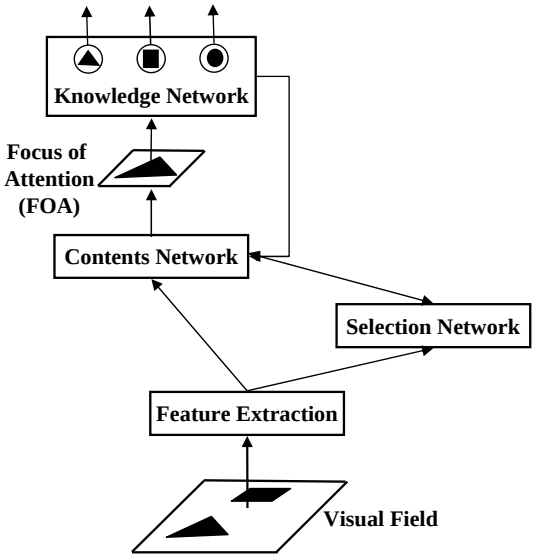


Fig. 2. Architecture of SAIM

The design of SAIM's network follows the idea of soft constraint satisfaction in neural networks that use "energy minimization" techniques [22]. In SAIM the "energy minimization" approach is applied in the following way: Each module in SAIM carries out a pre-defined task (e.g. the knowledge network has to identify the object in the FOA). In turn each task describes allowed states of activation in the network. These states then define the minima in an energy function. To ensure that the model as a whole satisfies each constraint, set by each network, the energy functions of each module are added together to form a global energy function for the whole system. The minima in the energy function are found via gradient descent, as proposed by [22]:

$$\tau \dot{x_i} = -\frac{\partial E(\mathbf{y})}{\partial y_i} \tag{1}$$

whereby y_i is the output activation of an unit and x_i the internal activation of an unit. The factor τ is antiproportional to the speed of the gradient descent. In the Hopfield approach x_i and y_i are linked together by the sigmoid function:

$$y_i = \frac{1}{1 + e^{-m \cdot (x_i - s)}}$$

and the energy function includes a leaky integrator, so that the descent turns into:

$$\tau \dot{x_i} = -x_i - \frac{\partial E(\mathbf{y})}{\partial y_i} \tag{2}$$

2.2 Contents network

The contents network aims at enabling translation-invariant mapping from input image to the FOA. This is implemented through the following energy function:

$$E\left(\mathbf{y^{CN}}, \mathbf{y^{SN}}\right) = \sum_{lmij} \left(\sum_{s,r} \sum_{n} \left(y^{CN}_{\substack{l+s,\\ m+r,n}} - f^{n}_{\substack{i+s,\\ j+r}} \right)^2 \right) \left(y^{SN}_{lmij}\right)^q \tag{3}$$

y^{SN}_{lmij} is the activation of units in the selection network and y^{CN}_{lmn} is the activation of units in the contents network. Here and in all the following equations the indices i and j refer to retinal locations and the indices l and m refer to locations in the FOA. f^n_{ij} is the output of the feature extraction with n noting the featural dimension. The term $\left(\sum_{s,r} \sum_{n} \left(y^{CN}_{\substack{l+s,\\ m+r,n}} - f^{n}_{\substack{i+s,\\ j+r}} \right)^2 \right)^2$ ensures that the units in the contents network match the feature values in the input image. The term $\left(y^{SN}_{lmij}\right)^q$ ensures that the contents of the FOA only reflect the region selected by the selection network ($y^{SN}_{lmij} = 1$). Additionally, since setting an arbitrary choice of y^{SN}_{lmij}s to 1 allows any location to be routed from the feature level to the FOA level, the contents network enables a translation-invariant mapping. The gradient descent with respect to y^{CN}_{lmn} defines the feedforward connections from feature extraction to FOA:

$$\frac{\partial E(\mathbf{y^{SN}}, \mathbf{y^{CN}})}{\partial y^{CN}_{lmn}} = 2 \cdot \sum_{ij} \left(y^{CN}_{lmn} - f^n_{ij}\right) \left(y^{SN}_{lmij}\right)^q \tag{4}$$

The gradient descent with respect to y^{SN}_{lmij} defines the feedback connections from FOA to selection network:

$$\frac{\partial E(\mathbf{y^{SN}}, \mathbf{y^{CN}})}{\partial y^{SN}_{lmij}} = \sum_{s,r} \sum_{n} \left(y_{\substack{l+s,\\ m+r,n}} - f^{n}_{\substack{i+s,\\ j+r}} \right) \cdot 2 \cdot q \cdot \left(y^{SN}_{lmij}\right)^{q-1} \tag{5}$$

Note that these feedback connections implement a matching between FOA contents and the features extracted from the input image. In fact, the matching results form the input into the selection network which guides the selection network towards choosing location input image that match well features represented in the FOA.

2.3 Selection network

The mapping from the retina to the FOA is mediated by the selection network. In order to achieve successful object identification, the selection network has to fulfill certain constraints when it modulates the mapping process. These constraints are that: (i) units in the FOA should receive the activity from only one retinal unit; (ii) activity of retinal units should be mapped only once into the FOA; (iii) neighbourhood relations in the retinal input should be preserved in

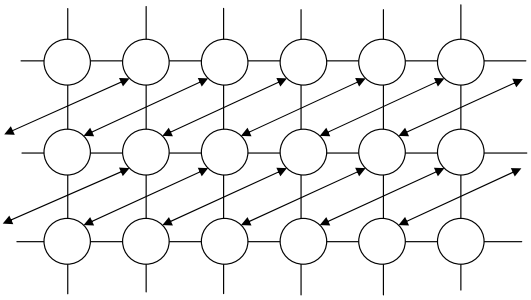


Fig. 3. Illustration of connections between units of the selection network. For simplicity the illustration depicts the selection network for a one-dimensional input image and one-dimensional FOA. A layer (row of units) controls the activation of one FOA unit via the contents network. A column in the selection network corresponds to a location in the input image. There are three types of connections within the selection network: two inhibitory connections depicted as lines within a layer and a column and excitatory connections depicted as slanted arrows. (see text for details).

mapping through to the FOA. As we will show, these three constraints implement three types of connections illustrated in Fig. 3. Now, to incorporate the first constraint, that units in the FOA should receive activity from only one location in input image unit, the equation of the WTA-equation suggested by [23] turns into:

$$E_{WTA}^{SN1}(\mathbf{y^{SN}}) = \sum_{ij}(\sum_{lm} y_{lmij}^{SN} - 1)^2 \tag{6}$$

The second term implements the second constraint:

$$E_{WTA}^{SN2}(\mathbf{y^{SN}}) = \sum_{lm}(\sum_{ij} y_{lmij}^{SN} - 1)^2 \tag{7}$$

In both terms the expression $(\sum y_{ikjl}^{SN} - 1)^2$ ensures that the activity of one location is mapped only once into the FOA.
For the neighborhood constraint the energy function was based on the Hopfield associative memory approach:

$$E(\mathbf{y}) = -\sum_{\substack{ij \\ i \neq j}} T_{ij} \cdot y_i \cdot y_j \tag{8}$$

The minimum of the function is determined by the matrix T_{ij}. For T_{ij}s greater than zero the corresponding y_is and y_js should either stay zero or become active in order to minimize the energy function. In the associative memory approach T_{ij} is determined by a learning rule. Here, we chose the T_{ij} so that the selection network fulfills the neighborhood constraint. The neighborhood constraint is fulfilled when units in the selection network which receive input from the adjacent

units in the visual field, and control adjacent units in the FOA are active at the same time. Hence, the T_{ij} for these units in Equation 8 should be greater than zero and for all other units T_{ij} should be less than or equal zero. This leads to the following equation:

$$E^{SN}_{neighbor}(\mathbf{y^{SN}}) = -\sum_{i,j,l,m}\sum_{\substack{s=-L\\s\neq 0}}^{L}\sum_{\substack{r=-L\\r\neq 0}}^{L} g_{sr}\cdot y^{SN}_{lmij}\cdot y^{SN}_{i+\Delta\cdot r,k+\Delta\cdot s,j+\Delta\cdot r,l+\Delta\cdot s} \quad (9)$$

with g_{sr} being defined by a Gaussian function:

$$g_{sr} = \frac{1}{A}\cdot e^{-\frac{s^2+r^2}{\sigma^2}} \quad (10)$$

where A was set, so that the sum over all g_{sr} is 1. When units linked via g_{sr} are activated to $y^{SN}_{lmij} = 1$, the energy is smaller than when these units have different values, e.g. zero and one. In the versions of SAIM g_{sr} connected units that relate to adjacent locations in both the FOA and the input image, implementing the neighbourhood constraint. In the current version the neighbourhood relationship with respect to the input image is modulated by the parameter Δ. With this modification SAIM maps every Δth-pixel from the input image into the FOA. Initially, this subsampling was introduced for practical reasons, as the objects used here span a region of around 30 by 30 pixels and it would have not been feasible to use a FOA of 30x30 pixels for computer time. Interestingly, this type of subsampling also introduces some kind of robustness into SAIM's processing of natural images, as we will discuss at the end of this paper.

2.4 Knowledge network

The knowledge network aims at recognizing the object in the FOA by matching the contents of the FOA with the templates stored in the knowledge network. Due to the subsampling introduced in the selection network, the template consists of a grid placed over an object (see Fig. 4 for examples). At each grid point the features of the location in the object are stored in the template weights. The features are generated by the feature extraction specified in the next section. Importantly, not every grid point is located on a object pixel. For these grid points, subsequently termed "background" (bg), the template weights are set to zero to mark them as non-object grid points. In order to distinguish these grid points from object template pixels, the range of feature values was set to be larger than 0.
The energy function of the knowledge network is defined as

$$\begin{aligned} E^{KN}(\mathbf{y^{KN}},\mathbf{y^{CN}}) &= a^{KN}(\sum_k y^{KN}-1)^2 - \\ & b^{KN}\cdot\sum_{klmn}\begin{cases} 0 & \text{If } \mathbf{w^k_{lmn}} \text{is "bg"} \\ (y^{CN}_{lmn}-w^k_{lmn})^2\cdot\frac{y^{KN}_k}{N_k} & \text{otherwise}\end{cases} \end{aligned} \quad (11)$$

The index k refers to template units whose templates are stored in their weights (w^k_{lmn}). The term $(\sum_k y^{KN} - 1)^2$ restricts the knowledge network to activate only one template unit. The term $\sum_{lmn}(y^{CN}_{lmn} - w^k_{lmn})^2 \cdot \frac{y^{KN}_k}{N_k}$ ensures that the best-matching template unit is activated whereby N_k is the number of object grid points. The normalization ensures that the matching value is independent of the number of object pixels. a^{KN} and b^{KN} weight these constraints against each other. The exclusion of the background pixels from the matching function takes into account the fact that feature values at those locations do not belong to objects and are therefore not relevant for the quality of the fit. As will be explained below, the introduction of the background pixels also required to modify the original gradient descent with respect to $\mathbf{y^{CN}}$:

$$\frac{\partial E(\mathbf{y^{KN}}, \mathbf{y^{CN}})}{\partial y^{CN}_{lmn}} = b^{KN} \cdot \sum_{klmn} \begin{cases} 0 & \text{If } \mathbf{w^k_{lmn}} \text{is "bg"} \\ 2 \cdot (y^{CN}_{lmn} - w^k_{lmn}) \cdot y^{KN}_k & \text{otherwise} \end{cases}$$

This feedback from the knowledge network leads into the contents network leads to the effect that FOA-pixels are not affected by background pixels. However, in order for the selected region to fit the object shape, the selection network via the contents network need to be influenced by background pixels. This can be achieved by the following, modified feedback:

$$fb = b^{KN} \cdot \sum_{klmn} \begin{cases} 2 \cdot (y^{CN}_{lmn} - w^k_{lmn}) \cdot y^{KN}_k & \text{If } \mathbf{w^k_{lmn}} \text{is "bg" } \& \; y^{KN}_k > \Theta \\ 0 & \text{If } \mathbf{w^k_{lmn}} \text{is "bg" } \& \; y^{KN}_k \leq \Theta \\ 2 \cdot (y^{CN}_{lmn} - w^k_{lmn}) \cdot y^{KN}_k & \text{otherwise} \end{cases} \quad (12)$$

With this modification the feedback behaves so long as the activation of the template unit (y^{KN}_k)is smaller than Θ. However, as soon as y^{KN}_k is larger than Θ is forced to converge to zero, since w^k_{lmn} as a background pixel is zero. The passing of Θ is interpreted as the knowledge network having recognized a selected object. The convergent of features value in the contents network towards zero leads to a suppression of activation in the corresponding layers of the selection network. This results from the fact that all feature values are larger than zero (see Feature extraction) and, therefore, the input of the selection network is highly negative, suppressing the activation in the layers of the selection network. Consequently, the selected locations form the shape of the object.

2.5 Feature extraction

The feature extraction extracted from an input image are intensity, colour and orientation, similar to the feature extraction in Itti and Koch's model. However, different from their model, no conspicuity map or saliency map is calculated from the feature maps. Instead, the feature extraction feeds into the selection network and the contents network.

The input image to the feature extraction is a RGB-image (r_{ij}, g_{ij}, b_{ij}) and the output are feature vectors noted as f^n_{ij}, whereby the indices i and j refere to image locations and n to the feature dimension. A constant (*const*) is added to

each feature dimension in order to allow the knowledge network to distinguish between deselected pixels and selected pixels (see Knowledge Network). The first feature dimension is intensity:

$$f_{ij}^{(1)} = (r_{ij} + g_{ij} + b_{ij})/3 + const \tag{13}$$

For the feature dimension "colour", the RGB-image is transformed into the HSV colour space (hue-saturation-value), as the HSV space is a good approximation of the way humans perceive colour [24]:

$$h = \begin{cases} 60 \cdot \frac{g-b}{Max-Min} + 0, \text{if } Max = r \text{ and } g \geq b \\ 60 \cdot \frac{g-b}{Max-Min} + 360, \text{if } Max = r \text{ and } g < b \\ 60 \cdot \frac{b-r}{Max-Min} + 120, \text{if } Max = g \\ 60 \cdot \frac{r-g}{Max-Min} + 240, \text{if } Max = b \end{cases} \tag{14}$$

$$s = \frac{Max - Min}{Max} \tag{15}$$

$$v = Max \tag{16}$$

whereby Max and Min are the maximum and minimum of the RGB-values, respectively. HSV-values represent positions in the HSV-space in cylindric coordinates with h being an angle and s and v a length ranging from 0.0 to 1.0. SAIM uses euclidian distances for template matching, thus, cartesian coordinates are more suitable for representing colour. The following equation transforms the cylindric coordinate into cartesian coordinates:

$$f_{ij}^{(2)} = v_{ij} + const \tag{17}$$

$$f_{ij}^{(3)} = s_{ij} \cdot \sin(h_{ij}) + 1 + const \tag{18}$$

$$f_{ij}^{(4)} = s_{ij} \cdot \cos(h_{ij}) + 1 + const \tag{19}$$

In order to extract orientation information from the input image Gabor filters are used:

$$G(x, y, \lambda, \theta) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp(-\frac{x^2 + y^2}{2\sigma^2}) \exp(2\pi\lambda i(x\cos\theta + y\sin\theta)) \tag{20}$$

where θ is the orientation. σ is the standard deviation of Gaussian envelope and λ is the spatial frequency of the filter. This filter is used as it is generally accepted that Gabor-filter are good approximation of receptive fields in V1. We filter 8 orientations (0, 22.5, 45, 67.5, 90, 112.5, 135, 157.5), as these are the orientation of the receptive fields in V1. Therefore, the last 8 feature dimensions are the intensity image filtered with the Gabor-filters in 8 orientations and to each filter result the positive constant *const* is added to ensure that those feature dimensions are larger than zero as well.

Fig. 4. Templates: Object 1 (left) and Object 2 (right). The crosses mark locations in the object that are used to extract feature values for the templates. The dots are "background pixels".

3 Simulation results

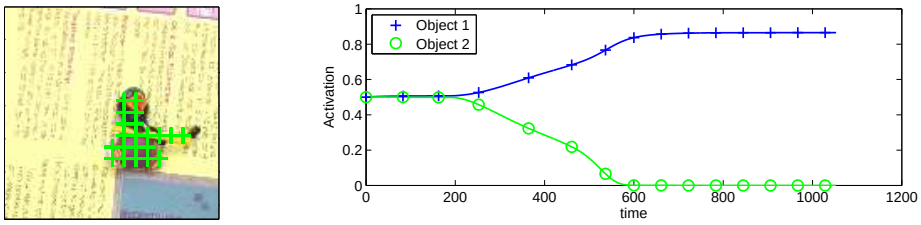
3.1 Input images and templates

Fig. 1 shows examples of the pictures used in the simulations. The pictures were used by one of the authors in earlier work (e.g. [25]). Two objects of similar size were placed by hand onto different backgrounds at different locations. Even though an effort was made to keep the orientation, lighting, colour and size of the objects constant, as can be seen from the examples, variations occurred and images exhibited natural noise. To simulate visual search the original images were not suitable, because the images did not contain the same object more than once, which is necessary to simulate visual search. Therefore, additional objects were inserted with the help of a graphics tool (see Fig. 6 for examples). We generated scenes with 2, 3 and 4 object, with the number of objects limited by the image size. The aim of this paper is to show that, in principle, SAIM is capable of simulating results of visual search tasks with natural images, so this limitation is not crucial to testing the proof of principle.
Fig. 4 shows the two templates (Object 1 and Object 2) used in simulations in this paper. The templates were cropped from two images and were used through out the simulations described in this paper.

3.2 Results

Figure 5 shows two examples of simulation results. The examples demonstrate that, in principle, the new version of SAIM is capable of processing natural images. In Figure 5a SAIM successfully selected Object 1 from a textured background. Fig. 5b shows that for a scene with two known objects in front of a textured background SAIM successfully selected one of the two objects, Object 2. SAIM selected Object 2, because it matched better the corresponding template than the template of Object 1. In both simulations SAIM's knowledge network correctly identified the selected object. Figure 5a also illustrates that SAIM appears to be robust against variations in these scenes, as Object 1 is slightly titled to the right and SAIM still successfully identifies the object. This

a.)



b.)

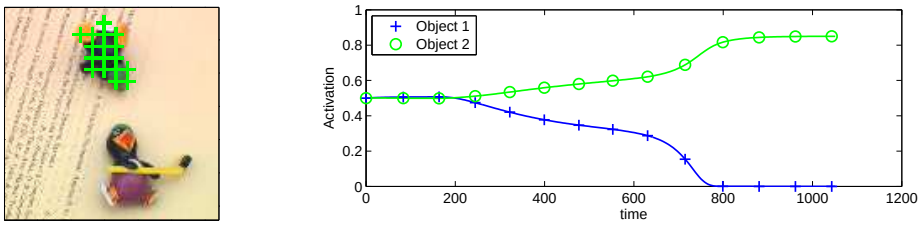


Fig. 5. Two examples of simulation results. The crosses in the left images indicate the image locations SAIM selected. The plots on the right show the time course of the activation of the templates in the knowledge network. The time scale is arbitrary, but can be interpreted as milliseconds.

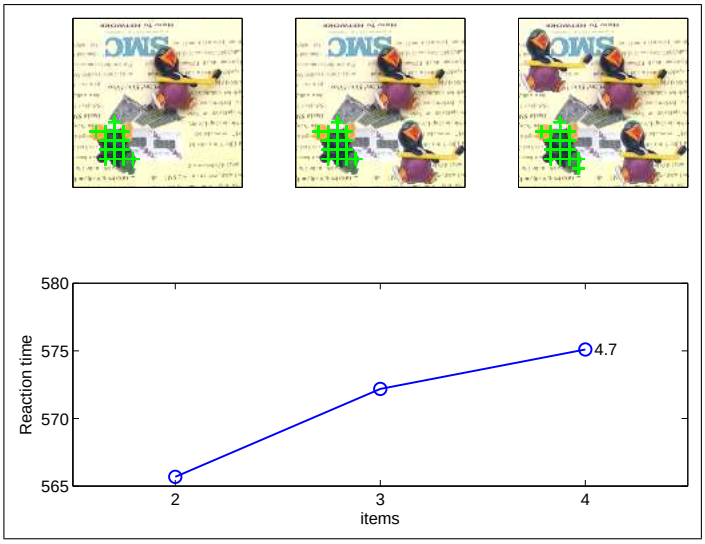
is due to the subsampling introduced in the selection network. For instance, the template grid points are still positioned on the (yellow) bat matching its colour, even though the exact position in the original template slightly different.

Figure 6 shows that SAIM captures the essential aspect of experimental findings in visual search tasks that the slope of the search function (reaction time over number of items) varies with the content of the searched scenes. SAIM's reaction time is measured by the amount of simulated time it takes until the activation of a template unit passes a threshold of 0.8. The increase of reaction time results from the competitive interactions within the selection network. These competitive interactions are driven by the comparison between the output of the feature extraction and the top-down activation from the contents network (see Eq. 5). The more objects are in the scene the more activation competes in the selection network, thus SAIM's reaction time increases. The different slopes result from the different degree of matching.

4 Discussion

This paper extended the model of attention and object recognition in order to process natural images with SAIM. The extension were two fold: First SAIM received a biologically plausible feature extraction including Gabor filtering and encoding in HSV-colour space. Second, the templates in new SAIM were more

a.)



b.)

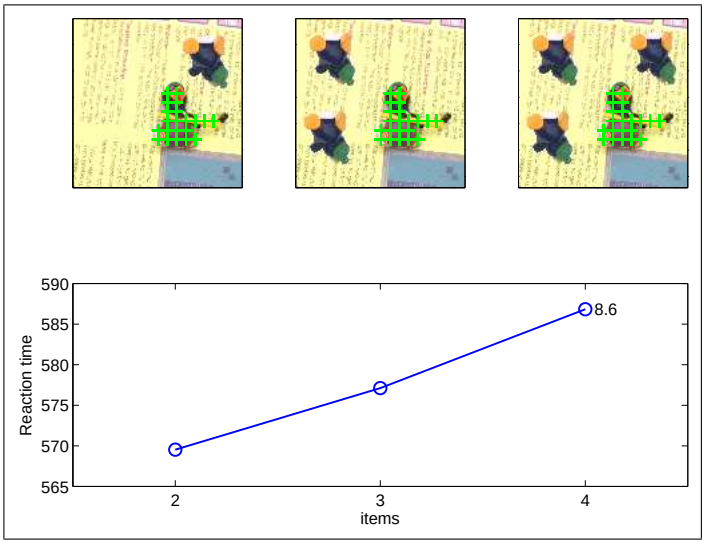


Fig. 6. Simulation results of visual search. The images show the search displays and indicate the selected image locations with crosses. The graphs illustrate how reaction time depends on the number of items in the scenes (search function). The figures at the right side of the graphs indicate the slope of the search function. The time scale can be interpreted as milliseconds.

flexible than in earlier versions, allowing to represent object shapes. The simulation results demonstrated that the new version of SAIM successfully selects

and recognizes object in natural colour images. Moreover, it can mimic experimental results from visual search tasks in which reaction times increase with the number of objects in the scene. This effect results from the influence of parallel, competitive interactions in SAIM. There are only a few models on human attention that are capable of processing natural images, the saliency-based model by Itti and Koch being the most prominent example [18–20]. In contrast to SAIM, Itti and Koch's model utilizes a serial search scan guided by a saliency-map to model visual search data. Also note that SAIM's selection process is not only driven by featural information (top-down and bottom-up), but also by proximity-based grouping implemented by the excitatory connections in the selection network (see [3, 6] for a detailed discussion). Such a behaviorally plausible grouping process is not integrated in Itti and Koch's model. Moreover, there is an interesting difference in the way object recognition is treated by the two models. The saliency-based model is often used as a front-end for an object recognition system (see [26] for an recent example). Thus there is little integration between the two processes, selection and recognition. In contrast to this, in SAIM the recognition system is an integral part of the whole architecture and acts to modulate selection. In further work we aim to test whether SAIM can serve as a useful framework for object recognition applications in computer vision.

Acknowledgment

This work was supported by grants from the European Union, the BBSRC and the EPSRC (UK) to Dietmar Heinke and Glyn W. Humphreys and from the EPSRC (UK) to Yarou Sun and Andreas Backhaus.

References

1. Heinke, D., Humphreys, G.W.: SAIM: A Model of Visual Attention and Neglect. In: Proc. of the 7th International Conference on Artificial Neural Networks–ICANN'97, Lausanne, Switzerland, Springer Verlag (1997) 913–918
2. Heinke, D., Humphreys, G.W., diVirgilo, G.: Modeling visual search experiments: Selective Attention for Identification Model (SAIM). In Bower, J.M., ed.: Neurocomputing. Volume 44 of Computational Neuroscience Meeting'01. (2002) 817–822
3. Heinke, D., Humphreys, G.W.: Attention, spatial representation and visual neglect: Simulating emergent attention and spatial memory in the Selective Attention for Identification Model (SAIM). Psychological Review **110**(1) (2003) 29–87
4. Heinke, D., Humphreys, G.W., Tweed, C.L.: Top-down guidance of visual search: A computational account. Visual Cognition **14**(4/5/6/7/8) (2006) 985–1005
5. Heinke, D., Humphreys, G.W.: Selective attention for identification model: Simulating visual neglect. Computer vision and image understanding **100**(1-2) (2005) 172–197
6. Heinke, D., Sun, Y.R., Humphreys, G.W.: Modeling grouping through interactions between top-down and bottom-up processes: The grouping and selective attention for identification model (G-SAIM). In: Lecture notes in computer science. Volume 3368. (2005) 148–158

7. Backhaus, A., Heinke, D., Humphreys, G.W.: Contextual Learning in the Selective Attention for Identification model (CL-SAIM): Modeling contextual cueing in visual search tasks. In: Proceedings of the 3rd international workshop on attention and performance in computer vision (WAPCV). (2005)
8. Amari, S.I.: Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. Biological Cybernetics **27** (1977) 77–87
9. Heinke, D., Humphreys, G.W.: Computational Models of Visual Selective Attention: A Review. In Houghton, G., ed.: Connectionist Models in Psychology. Psychology Press (2005) 273–312
10. Duncan, J.: The locus of interference in the perception of simultaneous stimuli. Psychological Review **87** (1980) 272–300
11. Kumada, T., Humphreys, G.W.: Lexical recovery on extinction: Interactions between visual form and stored knowledge modulate visual selection. Cognitive Neuropsychology **18**(5) (2001) 465–478
12. Navon, D.: Forest before trees: The precedence of global features in visual perception. Cognitive Psychology **9** (1977)
13. Egly, R., Driver, J., Rafal, R.D.: Shifting visual attention between objects and locations: Evidence from normal and parietal subjects. Journal of Experimental Psychology: Human Perception and Performance **123** (1994) 161–177
14. Posner, M.I., Snyder, C.R.R., Davidson, B.J.: Attention and the Detection of Signals. Journal of Experimental Psychology: General **109**(2) (1980) 160–174
15. Posner, M.I., Cohen, Y.: Components of Visual Orienting. Attention and Performance (1984) 531–556
16. Humphreys, G.W., Heinke, D.: Spatial representation and selection in the brain: Neuropsychological and computational constraints. Visual Cognition **5**(1/2) (1998) 9–47
17. Humphreys, G.W., Riddoch, M.J.: Separate Coding of Space Within and Between Perceptual Objects: Evidence from Unilateral Visual Neglect. Cognitive Neuropsychology **12**(3) (1995) 283–311
18. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence **20**(11) (1998) 1254–1259
19. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vison Research **40** (2000) 1489–1506
20. Koch, C., Itti, L.: Computational Modelling of Visual Attention. Nature Reviews: Neuroscience **2** (2001) 194–203
21. Wolfe, J.M.: Visual Search. In Pashler, H., ed.: Attention. Psychology Press (1998) 13–74
22. Hopfield, J.J., Tank, D.: "Neural" Computation of Decisions in Optimazation Problems. Biological Cybernetics **52** (1985) 141–152
23. Mjolsness, E., Garrett, C.: Algebraic Transformations of Objective Functions. Neural Networks **3** (1990) 651–669
24. Gonzalez, R., Woods, R.E.: Digital Image Processing. Prentice Hall Press, Upper Saddle River, New Jersey (2002)
25. Heinke, D., Gross, H.M.: A Simple Selforganizing Neural Network Archiecture for Selective Visual Attention. In: Proc. of the International Conference on Artificial Neural Network – ICANN'93, Amsterdam, The Netherlands, Springer Verlag (1993) 63–66
26. Walther, D., Itti, L., M., R., Poggio, T., Koch, C.: Attentional Selection for Object Recognition – a Gentle Way. In: Lecture notes in Computer Science. Volume 2525., Springer Verlag (2002) 472–479