Modeling Attention: From Computational Neuroscience to Computer Vision

Fred H. Hamker

[1] Allgemeine Psychologie, Psychologisches Institut II
Westf. Wilhelms-Universität,
48149 Münster, Germany
fhamker@uni-muenster.de
http://wwwpsy.uni-muenster.de/inst2/lappe/Fred/FredHamker.html
[2] California Institute of Technology,
Division of Biology 139-74,
Pasadena, CA 91125, USA

Abstract. We present an approach to modeling attention which originates in computational neuroscience. We aim at elaborating the underlying mechanisms of attention by fitting the model with data from electrophysiology. Our strategy is to either confirm, reject, modify or extend the model to accumulate knowledge in a single model across various experiments. Here, we demonstrate the present state of the art and show that the model allows for a goal-directed search for an object in natural scenes.

1 Introduction

Visual Search and other experimental approaches have demonstrated that attention plays a crucial role in human perception. Understanding attention and human vision in general could be beneficial to computer vision, especially in vision tasks that are not limited to specific and constrained environments. We discuss recent findings and hypotheses in the neurosciences that have been modeled by approaches from computational neuroscience. Neuroscience gives an insight into the brain which allows to further constrain algorithms of attention. In computational neuroscience the topics of interest are usually focused on a specific mechanism and networks often comprise only a relatively low number of cells and artificial inputs are used. Thus, scaleability becomes an important issue. A transfer of knowledge from computational neuroscience to computer vision requires at least the solution of three constraints: i) Does the number of cells influence the convergence of the algorithm? ii) Can the preconditions of the proposed solution be embedded into the systems level? iii) Can the model be demonstrated to operate on natural scenes?

In this contribution we derive a computational principle that allows to model large scale systems and vision in natural scenes. We demonstrate an approach for object detection in natural scenes. We suggest that goal directed attention and object detection are necessarily coupled, since an efficient deployment of attention benefits from an at least partial match of the encoded objects with the target.

L. Paletta et al. (Eds.): WAPCV 2004, LNCS 3368, pp. 118–132, 2005.

2 Spatial Attention

2.1 Gain Control

Single cell recordings have revealed that the neural response is enhanced in a multiplicative fashion when attention is directed to a single stimulus location [23], [15]. The neural correlate of such a multiplicative effect is still under discussion. It has been suggested that the gain of a neuronal response to excitatory drive is decreased by increasing the level of both, excitatory and inhibitory, background firing rates in a balanced manner [1]. On a more abstract level a feedback signal could increase the gain of the feedforward pathway in a multiplicative fashion [18], [9], [22]. We investigated such a gain control mechanism by simulating a V4 layer which receives input from a V2 population. We consider feedforward, lateral excitory and inhibitory input and spatial bias.

Given a neural population in V4 and a feedforward input $I^{\uparrow}$ we have proposed that a cell's response over time $r(t)$ can be computed by a differential equation:

$$\tau \frac{d}{dt} r_k(t) = I_k^{\uparrow} + I_k^N + I_k^A - I_k^{inh} \tag{1}$$

Inhibition I_k^{inh} introduces competition among cells and normalizes the cell's response by a shunting term. I_k^N describes the lateral influence of other cells in the population. Spatial attention is proposed to emerge from the modulation of the feedforward signal by feedback A_x prior to spatial pooling:

$$I_k^A = f(w_{i,x} a_{i,x}^A); \quad a_{i,x}^A = I_{i,x}^{\uparrow} \cdot A_x \qquad f = \max_{i,x} \tag{2}$$

We presented a stimulus to a population of 11 orientation selective cells for 150 ms and computed the average activity of each cell with and without a spatial bias. Consistent with the findings, we observed that the response on the population level is close to a multiplicative increase of the gain (Fig. 1). If we consider neural cells as feature detectors indicating the probability that the encoded feature is present in the scene, the function of gain control is increasing the probability of a feature being detected.

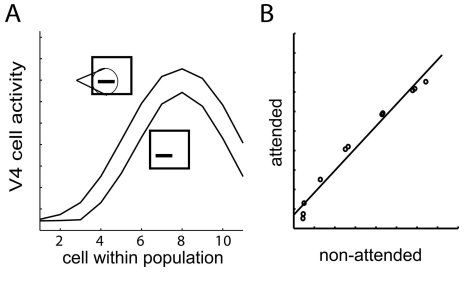

Fig. 1. Gain control. (A) Population responses to a single horizontal bar 100 ms after stimulus onset with and without a spatial bias. Each cell encodes a different orientation. (B) The firing rate of each cell in the non-attended case is plotted against the attended case (see [15]). The model shows approximately a multiplicative gain increase (CorrCoeff=0.99) of 13% (slope=1.13)

2.2 Contrast Dependence

A simple multiplicative gain control model (response gain model) predicts that the effect of attention increases with stimulus contrast. However, this is not a very useful strategy, since a high contrast stimulus is already salient. If we assume at least some parallel processing, a too high gain to an already salient stimulus could suppress other potentially relevant responses. Indeed the brain uses a strategy in which the magnitude of the attentional modulation decreases with increasing contrast [21], [13]. Attention results rather in a shift of the contrast response function (contrast gain model). This was experimentally tested by presenting a single luminance-modulated grating within the receptive field of a V4 cell. The monkey was then instructed to either attend towards the stimulus location or towards a location far outside of the receptive field. An increase of the stimulus contrast resulted in an increase of the neural response and in a decrease of the difference between both conditions (Fig. 2A).

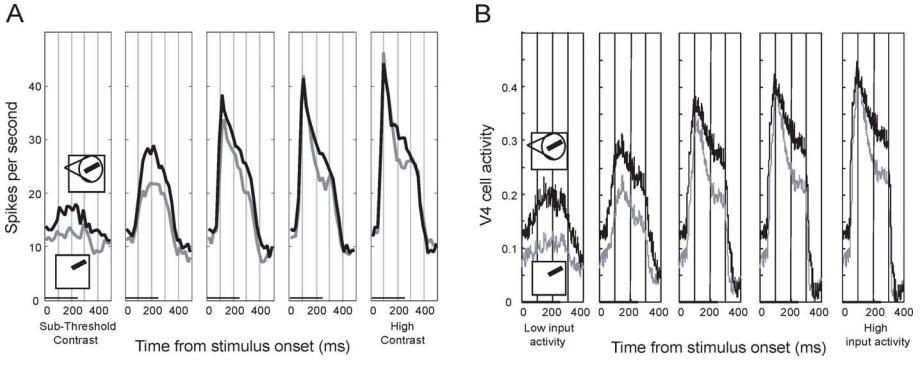

Fig. 2. Contrast dependent effect of attention. (A) Averaged single cell responses over time, with increasing contrast. The data was provided by J. Reynolds [21]. (B) Simulation results of our model. The initial burst at high contrast levels occurs due do the delayed inhibitory response I^{inh}. Similar to the data, the timing of the attention effect shifts with increasing contrast from early to late

In the model of Reynolds and Desimone [19] spatial attention affects the weight of feedforward excitation and feedforward inhibition. As the activation of the input increases the inhibition increases as well and the cell's response will saturate at a level where excitation and inhibition are balanced. However, a re-implementation of this model shows that the model replicates their finding on the level of the mean response over the whole presentation time, but it does not account very well for the observed temporal course of activity and the timing of the attention effect [22]. The model of Spratling and Johnson [22] accounts better for the temporal course of activity, but different as indicated by the data, the decrease in the magnitude of the attentional modulation occurs only at high contrast. A potential problem of this model is, that it explains the decrease of the attentional modulation by a saturation effect, which occurs only at high activity.

We postulate that the efficiency of the feedback signal decreases with increasing strength of the cell. Thus, the feedback signal A_x (eq. 2) which determines the gain factor $1 + A_x$ is combined with an efficiency term using the activity of the output cell k.

$$A_{k,x} = \sigma(\alpha - r_k) \cdot A_x \tag{3}$$

with $\sigma(a) = max(a, 0)$. We applied the extended model to simulate the effect of contrast dependence. Increasing contrast was simulated by increasing the stimulus strength. The model accounts quite well for the findings, even in the temporal course of activity (Fig. 2B). The magnitude of the attention effect is not explained by the saturation of the cell. Thus, the contrast dependency of attention is consistent with an effective modulation of the input gain by the activity of the cell. An answer towards the underlying exact neural correlate, however, requires more research.

To further demonstrate that the model is consistent with the contrast gain model, the magnitude of attention on the time averaged response with varying stimulus strength is shown (Fig. 3). We computed the mean response beginning from stimulus onset (Fig. 3A) and the mean over the first initial response (Fig. 3B) to show the timing of attention. For high contrast stimuli attention is most prominent in the late response and almost diminishes in the early response. Please note, the spatial feedback signal itself is constant.

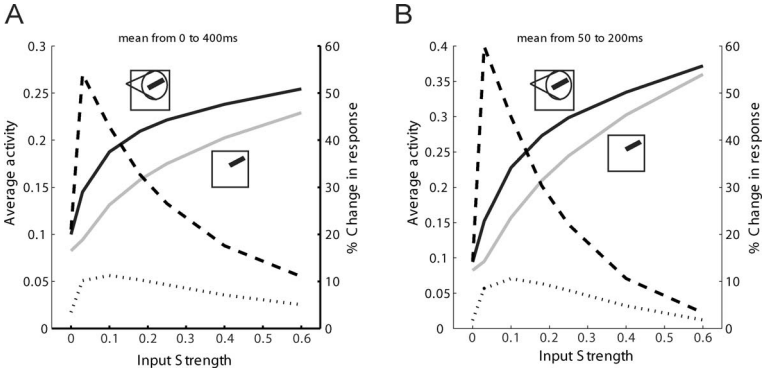

Fig. 3. Simulation of the attention effect on the mean response by varying the strength of the input stimulus. The dotted line shows the absolute difference between the attended and non-attended condition, and the dashed line the difference in percent. Consistent with the contrast gain model [21], the primary effect of attention occurs with low input activty. (A) Time average over the whole neural response after stimulus onset. (B) Time average over the initial burst

2.3 Biased Competition

It has been observed that neuronal populations compete with each other when more than a single stimulus are presented within a receptive field. Such competition can be biased by top-down signals [4]. As a result, the irrelevant stimulus is suppressed as if only the attended one had been presented. Numerous experiments have supported this framework.

Reynolds, Chelazzi and Desimone observed competitive interactions by placing two stimuli (reference and probe) within the receptive field of a V4 neuron [20]. They found that when spatial attention was directed away from the receptive field, the response to both stimuli was a weighted average of the responses to the stimuli presented in isolation. If the reference elicits a high firing rate and the probe a low firing rate, then the response to both is in between. Attending to the location of one of the stimuli biases the competition towards the attended stimulus.

We modeled this experiment by presenting now two stimuli to our neural population. In the attended case the gain of the input from one location is increased. The simulation results of our population approach fit with the experimental data (Fig. 4). A slope of 0.5 indicates that reference and probe are equally well represented by the population. The small positive y-intercept signifies a slight overall increase in activity when presenting a second stimulus along with the first. Attending to the probe increases the slope (not shown), indicating the greater influence of the attended probe over the population. Attending to the reference reduces the slope, signifying the greater influence of the attended reference stimulus. Attention in general enhances the overall response within the population, which is observed by the greater upward shift of the sensory interaction index as compared to the attend away condition.

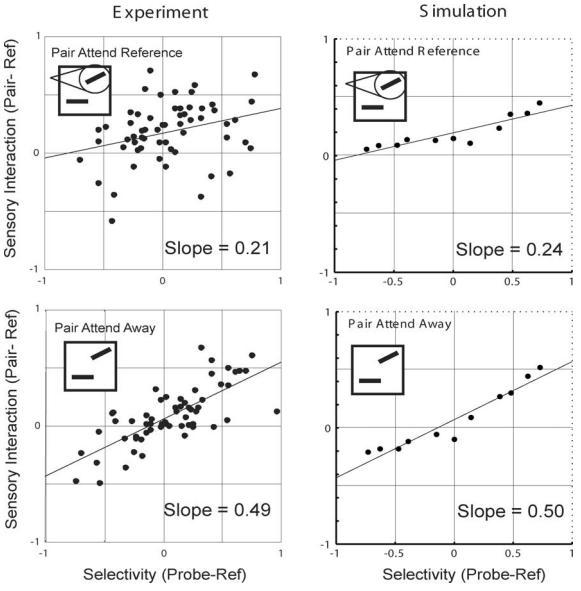

Fig. 4. Comparison of the simulation results with the experimental data (modified from [20]) to investigate the influence of attention on the sensory interaction. For each cell, its selectivity index is plotted over its sensory interaction. A selectivity value of 0 indicates identical responses to reference and probe in isolation, a positive value a preference towards the probe and a negative a preference towards the reference. An interaction index of 0 signifies that the cell is unaffected by adding a probe. Positive values indicate that the cell's response to the reference is increased by adding a probe and negative values signify a suppression by the probe

Along with the experimental data, Reynolds, Chelazzi and Desimone [20] demonstrated that a feedforward shunting model [5] can account for their findings. In their model competition among cells occurs due to feedforward inhibition from V2 cells onto V4 cells. In our approach competition occurs after pooling. It is based on lateral short range excitory and long range inhibitory connections within the population, which is in accordance with findings in V4. Other models [6] [3] [22] have referred to the findings of Reynolds, Chelazzi and Desimone [20] as well, but no quantitative comparison with the experimental data (Fig. 4) has been given.

3 Feature-Based Attention

3.1 Feature-Similarity

With reference to feature-based attention Treue and Martínez Trujillo [24] have proposed the Feature-Similarity Theory of attention. Their single cell recordings in area MT revealed that directing attention to a feature influences the encoding of a stimulus even when the second stimulus is presented outside of the receptive field. They proposed that attending towards a feature could provide a global feedback signal which affects other locations than the attended one as well. Feedback can be a very useful mechanism for a feature-based selection, as already demonstrated in early computational models of attention [25].

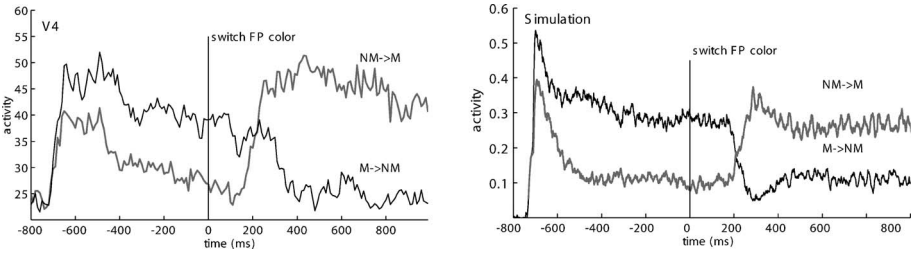

Fig. 5. Temporal course of activity in the match and non-match condition of V4 cells and simulated cells. The scene is presented at -800 ms. Cells representing the potential target object show an enhanced activity. If the fixation point color switches to another color at 0 ms, the activity follows the definition of the target. Neurons previously representing potential targets change into distractors and vice versa. The left figure shows the original data provided by B. Motter (the published data [17] does not show the activity after stimulus onset)

In an earlier experiment that presumably revealed feature-based attention effects the knowledge of a target feature increased the activity of V4 cells [17]. The task required to report the orientation of an item that matches the color of the fixation point. Since the display during the stimulus presentation period contains several possible targets, the monkey had to wait until the display contained only one target. Even during this stimulus presentation period, V4 neurons showed an enhanced activity if the presented colour or luminance items matched the target (Fig. 5 left). This dynamic effect is thought to occur

in parallel across the visual field, segmenting the scene into possible candidates and background.

We simulated this experiment by presenting input to six $x \in \{1 \dots 6\}$ V4 populations containing 11 cells i in each dimension $d \in \{color, orientation\}$ (eq. 4).

$$\tau \frac{d}{dt} r^{V4}_{d,i,x} = I^{\uparrow}_{d,i,x} + I^{N}_{d,i,x} + I^{A}_{d,i,x} - I^{inh}_{d,x} \tag{4}$$

We also model one IT population, whose receptive field covers all V4 receptive fields (Fig. 6). Let us assume the model is supposed to look for red items. This is implemented by generating a population of active prefrontal cells (PF) representing a red target template. At $t = -900ms$ we activate the target template in PF and present the inputs at $t = -800ms$. The input activity travels up from V4 to IT. Once the activity from V4 enters IT, competition gets biased by feedback from prefrontal cells. They in turn project back to V4 and enhance the gain of the V4 input. Thus, the term $I^{A}_{d,i,x}$ is a result of the bottom-up signal $I^{\uparrow}_{d,i,x}$ modulated by the feedback signal $r^{\text{IT}}_{d,j}$ with $w^{\text{IT,V4}}_{i,j}$ as the strength of the feedback connection:

$$I^{A}_{d,i,x} = f\left(I^{\uparrow}_{d,i,x} \sigma(\alpha - r^{\text{V4}}_{d,i,x}) \cdot \max_{j} (w^{\text{IT,V4}}_{i,j} \cdot r^{\text{IT}}_{d,j}) \right) \tag{5}$$

Feedback in the "object pathway" operates feature specific and largely location unspecific. By changing the target template at $t = 150ms$ the model now switches into a state where again all items of the target color in V4 are represented by a higher firing rate than those with a non-matching color (Fig. 5 right). Consistent with the Feature-Similarity Theory, the enhancement of the gain depends on the similarity of the input population with the feedback population.

The computationally challenging task of this experiment is to enable the model to switch its internal representation. Our gain control mechanism supports rapid switches, because feedback acts on the excitatory input of a cell and not on its output activity. Due to the switch in the PF activity, the population in IT encoding red looses its feedback signal whereas the one for green receives support. As a result, the prioritized encoding in IT changes, and the whole system switches to a state where populations encoding the new target feature are represented by a higher activity.

4 Attention on the Systems Level

A top-down feature-specific signal has also been revealed in IT cells during visual search [2]. In this experiment an object was presented to a monkey, which after a brief delay, had to be detected in a visual search scene. The monkey was trained to indicate the detection by shifting its gaze from the fixation point towards the target. Chelazzi found that the initial activation of IT neurons is largely stimulus driven and cells encoding target and non-target become activated. Since different populations compete for representation, typically the cells encoding the non-target get suppressed. A computational approach by Usher and Niebur [26] has shown that a parallel competition based on lateral interactions and a top-down bias is sufficient to qualitatively replicate some of those

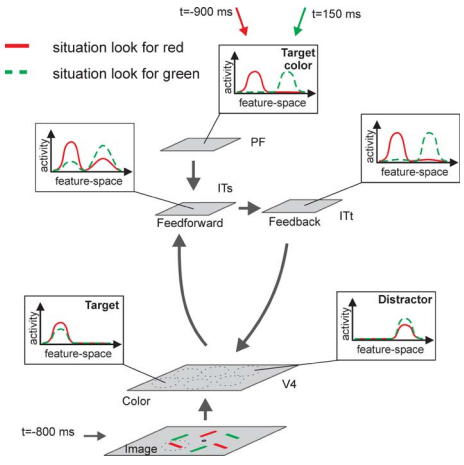

Fig. 6. Illustration of the pathway for "object recognition". At each V4 receptive field we model an arbitrary color space. We only show one target and one distractor. Due to the gain control by feedback the population encoding the target color gets enhanced and the one for the distractor is suppressed. The network settles in a state where all items of the target color in V4 are represented by a higher firing rate than those with a non-matching color. The second situation is "looking for green items", indicated by the dashed activity curves

findings. However, we have argued that the planning of an eye movement towards the target should produce a spatial reentry signal directed to the target location [8]. This prediction recently received further evidence by a study in which a microstimulation in the frontal eye field resulted in a modulation of the gain in V4 cells [16].

We have modeled the visual search experiment on the systems level by a model consisting of areas V4, IT, FEF and PFC (Fig. 7A,B). Thus, spatial and feature-based attention are now brought together in a single model. V4 cells receive a top-down signal from IT and the FEF, which both add up:

$$I^A_{d,i,x} = f\left(I^{\uparrow}_{d,i,x} \cdot \sigma() \cdot \max_j w^{\mathrm{ITt,V4}}_{i,j} \cdot r^{\mathrm{IT}}_{d,j}\right) + f\left(I^{\uparrow}_{d,i,x} \cdot \sigma() w^{\mathrm{FEFm,V4}} r^{\mathrm{FEFm}}_{x}\right) \quad (6)$$

with $\sigma() = \sigma(\alpha - y^{V4}_{d,k,x})$ and $w^{\mathrm{FEFm,V4}}$ defines the weight of the feedback from the FEF. For implementation details please refer to [8].

Our simulation result matches even the temporal course of activity of the experimental data (Fig. 7C). The model predicts that the firing rate of V4 and IT cells show an early feature-based effect and a late spatial selectivity (after 120 ms). In the 'Target Absent' condition where the cue stimulus is different from the stimuli in the choice array no spatial reentry signal emerges since in this case a saccade has to be withheld. The model does not contain any control units or specific maps that implement attention. The proposed gain control and competition allows higher areas to influence processing in lower areas. As a result, suppressive and facilitatory effects occur, commonly referred to as "attention". Thus, attention can emerge on the network level and does not have to be explicitly implemented.

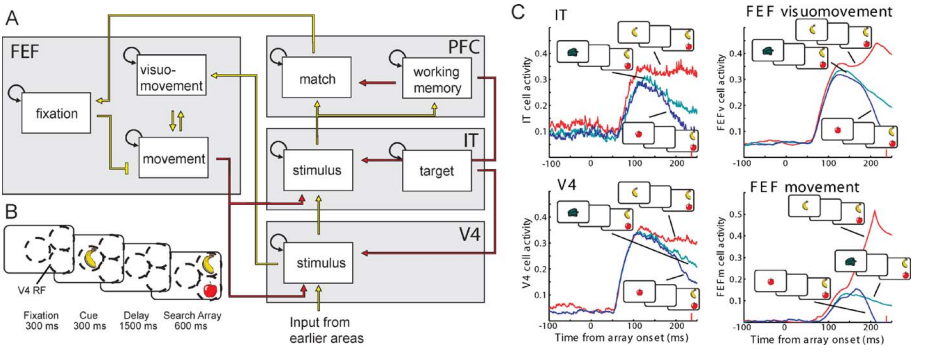

Fig. 7. (A) Sketch of the simulated areas. Each box represents a population of cells. The activation of those populations is a temporal dynamical process. Bottom-up (driving) connections are indicated by a bright arrow and top-down (modulating) connections are shown as a dark arrow. (B) Simulation of the experiment. The objects are represented by a noisy population input. RF's without an object just have noise as input. Each object is encoded within a separate RF, illustrated by the dashed circle, of V4 cells. All V4 cells are within the RF of the IT cell population. (C) Activity within the model areas aligned to the onset of the search array in the different conditions

5 Large Scale Approach for Modeling Attention

Our earlier simulations have shown that competition among feature representations could be a useful mechanism to filter out irrelevant stimuli for object recognition. A spatial focus of attention can reduce the influence of features outside the focus, whereas a competition among features could have the potential to select objects without the need of a segmentation on the image level. We now demonstrate how attention emerges in the process of detecting an object in a natural scene [7]. In extension to a mere biased competition we show that top-down signals can be modeled as an expectation, which alters the gain of the feedforward signal.

5.1 Overview

The idea is that all mechanisms act directly on the processed variables and modify their conspicuity. Each feature set is modeled as a continuous space with $i \in N$ cells at location $\mathbf{x} = (x_1, x_2)$ by assigning each cell a conspicuity $r_{d,i,\mathbf{x}}$. From the feature maps we determine contrast maps according to a measure of stimulus-driven saliency (Fig. 8). Feature and contrast maps are then combined into feature conspicuity maps which encode the feature and its initial conspicuity by means of a population code.

The conspicuity of each feature is altered by the target template. A target object is defined by the expected features $\hat{r}^F_{d,i}$. We infer the conspicuity of each feature $r_{d,i,\mathbf{x}}$ by comparing the expected features $\hat{r}^F_{d,i}$ with the bottom-up signal $r^{\uparrow}_{d,i,\mathbf{x}}$. If the bottom-up signal is similar to the expectation we increase the conspicuity. Such a mechanism enhances in parallel the conspicuity of all features at level II which are similar to the target template. We perform the same procedure on level I where the expected features are those from level II. In order to detect an object in space we combine the conspicuity

across all d channels in the perceptual map and generate an expectation in space $\hat{r}^L_{\mathbf{x}}$ in the movement map. The higher the individual conspicuity $r_{d,i,\mathbf{x}}$ across d at one location relative to all other locations the higher is the expectation in space $\hat{r}^L_{\mathbf{x}}$ at this location. Thus, a location with high conspicuity in different channels d tends to have a high expectation in space $\hat{r}^L_{\mathbf{x}}$. Analogous to the inference in feature space we iteratively compare the expected location $\hat{r}^L_{\mathbf{x}}$ with the bottom-up signal $r^{\uparrow}_{d,i,\mathbf{x}}$ in $\mathbf{x}$ and enhance the conspicuity of all features with a similarity of expectation and bottom-up signal. The conspicuity is normalized across each map by competitive interactions. Such interative mechanisms finally lead to a preferred encoding of the features and space of interest. Thus, attention emerges by the dynamics of vision.

Preprocessing: We compute feature maps for Red-Green opponency (RG), Blue-Yellow opponency (BY), Intensity (I), Orientation (O), and Spatial Resolution (σ). We determine the initital conspicuity by center-surround operations [11] from the feature maps which gives us the contrast maps. The feature-conspicuity maps combine the feature and conspicuity into a population code, so that at each location we encode each feature and its related conspicuity.

Level I: Level I has d channels which receive input from the feature conspicuity maps: $r_{\theta,i,\mathbf{x}}$ for orientation, $r_{I,i,\mathbf{x}}$ for intensity, $r_{RG,i,\mathbf{x}}$ for red-green opponency, $r_{BY,i,\mathbf{x}}$ for blue-yellow opponency and $r_{\sigma,i,\mathbf{x}}$ for spatial frequency (Fig. 8). The expectation of features at level I originates in level II $\hat{r}^{\mathrm{I}_F}_{d,i,\mathbf{x}'} = r^{\mathrm{II}}_{d,i,\mathbf{x}}$ and the expected location in the movement map $\hat{r}^{\mathrm{I}_L}_{\mathbf{x}'} = r^{m}_{\mathbf{x}'}$. Please note that even level II has a coarse dependency on location.

Level II: The features with their respective conspicuity and location in layer I project to layer II, but only within the same dimension d, so that the conspicuity of features at several locations in level I converges onto one location in level II. We simulate a map containing 9 populations with overlapping receptive fields. We do not increase the complexity of features from level I to level II. The expected features at level II originate in the target template $r^{II_F}_{d,i,\mathbf{x}} = w \cdot r^{\mathrm{T}}_{d,i}$ and the expected location in the movement map $\hat{r}^{II_L}_{\mathbf{x}} = w \cdot r^{\mathrm{m}}_{\mathbf{x}}$

Perceptual Map: The perceptual map (v) indicates salient locations by integrating the conspicuity of level I and II across all channels. In addition to the the conspicuity in level I and II we consider the match of the target template with the features encoded in level I by the product $\prod_d \max_{i,\mathbf{x}' \in RF(\mathbf{x})} r^{\mathrm{T}}_{d,i} \cdot r^{\mathrm{I}}_{d,i,\mathbf{x}'}$. This implements a bias to locations with a high joint probability of encoding all searched features in a certain area.

Movement Map: The projection of the perceptual map onto the movement map (m) transforms the salient locations into a few candidate locations which provide the expected location for level I and level II units. We achieve this by subtracting the average saliency from the saliency at each location $w^{\mathrm{v}} r^{\mathrm{v}}_{\mathbf{x}} - w^{\mathrm{v}}_{inh} \sum_{\mathbf{x}} r^{\mathrm{v}}_{\mathbf{x}}$. Simultaneously, the movement units indicate the target location of an eye movement.

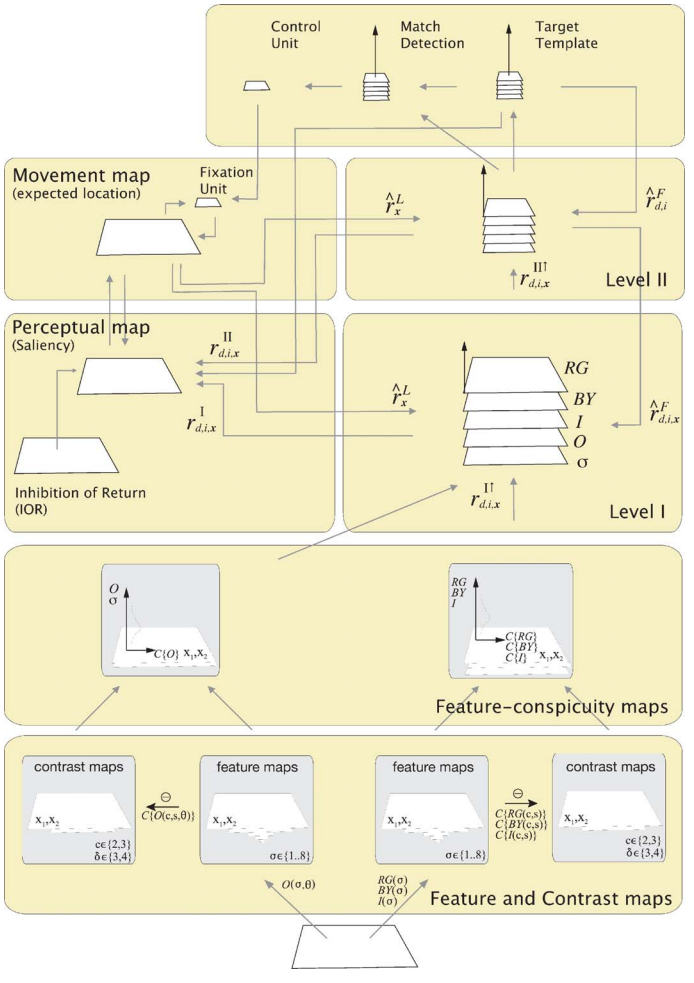

Fig. 8. Model of attentive vision. From the image we obtain 5 feature maps. For each feature at each location $\mathbf{x}$ we compute its conspicuity in the contrast maps and then combine feature and conspicuity into feature-conspicuity maps. This initial, stimulus-driven conspicuity is now dynamically updated within a hierarchy of levels. From level I to level II we pool across space to achive a representation of features with a coarse coding of location. The target template $\hat{r}^F_{d,i}$ holds the to be searched pattern regardless of its location and enhances the gain of level II cells which match the pattern of the template. $\hat{r}^F_{d,i,x}$ sends the information about relevant features further downwards to level I cells to localize objects with the relevant features. In oder to identify candidate objects the perceptual map integrates across all 5 channels to determine the saliency. The saliency is then used to compute the expected locations of an object $\hat{r}^L_{\mathbf{x}}$ in the movement map, which in turn enhances the conspicuity of all features at level I and II at these locations. Match detection cells fire, if the encoded features in level II match with the target template. This information can be used to control the fixation unit

5.2 Results

We now demonstrate the performance of our approach on an object detection task (Fig. 9). We present an object to the model for 100 ms and let it memorize some of its features as a target template. We do not give the model any hints which feature to memorize. As in the experiment done by Chelazzi, the model's task is to make an eye movement towards the target. When presenting the search scene, level II cells that match the target template quickly increase their activity to guide level I cells. In the blue-yellow channel at level I the target template is initially not dominant but the modulation by the expectation from level II overwrites the initial conspicuity. Thus, the features of the object of interest are enhanced prior to any spatial focus of attention which allows to guide the planning of the saccade in the perceptual and movement map sufficiently well. Saliency is not encoded in a single map. Given that level I cells have a spatially localized receptive field and show an enhanced response to relevant stimuli, they could be interpreted to encode a saliency map as well, which is consistent with recent findings [14]. A feature independent saliency map is achieved by the integration across all channels. The process of planning an eye movement provides a spatially organized reentry signal, which enhances the gain of all cells at the target location of the intended eye movement. Thus, spatial attention could be interpreted as a shortcut of the actual planned eye movement. Under natural viewing conditions spatial attention and eye movement selection are automatically coordinated such that prior to the eye movement the amount of reentry is maximized at the endpoint and minimized elsewhere. This would facilitate planning processes to evaluate the consequences of the planned action.

6 Discussion

We have modeled several attention experiments to derive the basic mechanisms of visual processing and attention. Initially we have focused on the gain control mechanism and demonstrated that an input gain model allows for a quantitative match with existing data. If we further assume that the gain factor decreases with the activity of the cell, the model is consistent with a contrast gain model. We then extended the model to simulate more complex tasks and to model the behavioral response as well. Again, we have been able to achieve a good match with the data. So far the model provides a comprehensive account of attention, specifically on the population averaged neural firing rate. Certainly attention is still more complicated than covered by the present model. However, the good fit with many existing data makes us believe the model contains at least several relevant local mechanisms that determine attention in the brain. Almost 20 years after the influential computational model of Koch and Ullman [12] was published, single cell recordings and computational modeling have now discovered a more fine graded model in which attention is explained on the systems level rather than by a selection within a single area.

In regard of this emerging new view on attention we investigated if the derived principles of the distributed nature of attention can be demonstrated to provide something useful beyond fitting experimental data. Thus, we tested an extension of the model on a goal-directed object detection task in natural scenes. We are confident that this joint approach gives the model a high potential for future computer vision tasks. The present

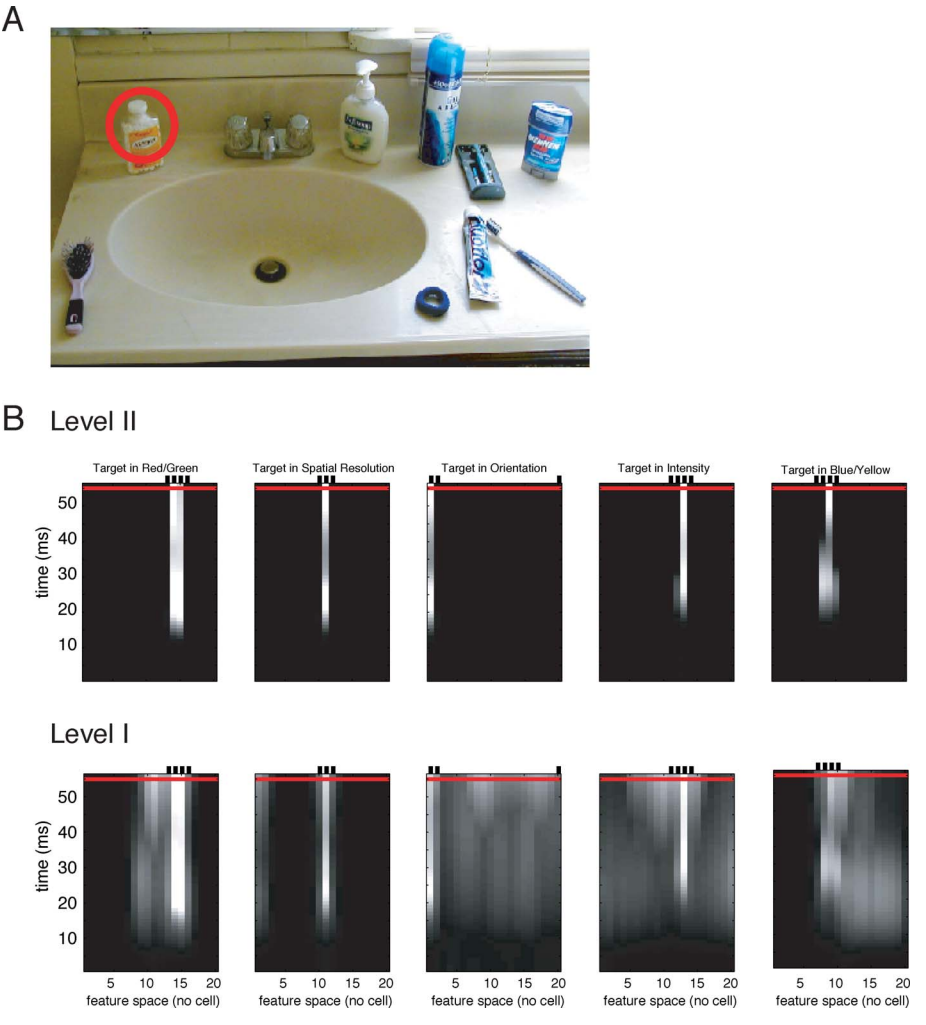

Fig. 9. Visual search in natural scenes. The asprin bottle in the upper left corner was presented to the model before the scene appeared, and in each dimension the most conspicious feature was memorized in order to generate a target template. Then the model searched for the target. A) Indication of the first eye movement, which directly selects the target. B) Conspicuity values of level II and level I cells in all channels over time. At each level the maximum response for each feature is shown, regardless of the receptive field of the cell. The strength of conspicuity is indicated by brightness. The target template is indicated by the bars at the top of each figure. The conspicuity of each feature occurs first in level I and then travels upwards to level II. Level II, however, first follows the target template, which then travels downwards to level I. This top-down inference is clearly visible in the blue-yellow channel (most right), where initially other features than the target feature are conspicious. The effect of the spatially localized reentry signal is best visible in the Intensity channel (second right). Prior to the eye movement several cells gain in activity, independent to their similarity of the encoded feature to the target template

demonstration of an object detection in natural scenes is very valuable from the viewpoint of attention, since it demonstrates that the derived principles even hold for the postulated three constraints: i) large number of cells ii) systems level and iii) natural scenes.

From the viewpoint of computer vision, we are aware that such an object detection task can be solved by classical methods. The advantage of our approach, however, lies in the integration of recognition and attention into a common framework. Attention improves object recognition, specifically in cluttered scenes, but only if attention can be properly guided to the object of interest. Feature-specific feedback within the object recognition pathway, gain control and competitive interactions directly enhance the features of interest and guide spatial attention to the object of interest. Partial attention improves further analysis which in turn helps to direct attention. We propose that the direction of attention and recognition must be an iterative process to be effective. In the present version we only used simple cues. Thus, future work has to focus on the learning of effective feedforward and feedback filters for shape recognition and object grouping.

Acknowledgements

The main part of the presented research has been done at Caltech. In this respect, I thank Christof Koch for his support and Rufin VanRullen for helpful discussions. I am grateful to John Reynolds and Brad Motter for providing data showing attention effects on V4 cells. Most of this research was supported by DFG HA2630/2-1 and in part by the ERC Program of the NSF (EEC-9402726).

References

1. Chance, F.S., Abbott, L.F., Reyes, A.D. (2002) Gain modulation from background synaptic input. Neuron, 35, 773-782.
2. Chelazzi, L., Miller, E.K., Duncan, J., Desimone, R. (1993) A neural basis for visual search in inferior temporal cortex. Nature, 363, 345-347.
3. Corchs, S., Deco, G. (2002) Large-scale neural model for visual attention: integration of experimental single-cell and fMRI data. Cereb. Cortex, 12, 339-348.
4. Desimone, R., Duncan, J. (1995) Neural mechanisms of selective attention. Anu. Rev. of Neurosc., 18, 193-222.
5. Grossberg, S. (1973) Contour enhancement short term memory, and constancies in reverberating neural networks, Studies in Applied Mathematics, 52, 217-257.
6. Grossberg, S., Raizada, R. (2000) Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. Vis. Research, 40, 1413-1432.
7. Hamker, F.H., Worcester, J. (2002) Object detection in natural scenes by feedback. In: H.H. Bülthoff et al. (Eds.), Biologically Motivated Computer Vision. Lecture Notes in Computer Science. Berlin, Heidelberg, New York: Springer Verlag, 398-407.
8. Hamker, F.H. (2003) The reentry hypothesis: linking eye movements to visual perception. Journal of Vision, 11, 808-816.
9. Hamker, F.H. (2004) Predictions of a model of spatial attention using sum- and max-pooling functions. Neurocomputing, 56C, 329-343.
10. Hamker, F.H. (2004) A dynamic model of how feature cues guide spatial attention. Vision Research, 44, 501-521.

11. Itti, L., Koch, C., Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 20, 1254-1259.
12. Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Human Psychology 4:219-227.
13. Martínez Trujillo, J.C., Treue, S., (2002) Attentional modulation strength in cortical area MT depends on stimulus contrast. Neuron, 35:365-370.
14. Mazer, J.A., Gallant ,J.L. (2003) Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. Neuron, 40, 1241-1250.
15. McAdams, C.J., Maunsell, J.H. (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4, J. Neurosci. 19, 431-441.
16. Moore, T., Armstrong, K.M. (2003) Selective gating of visual signals by microstimulation of frontal cortex. Nature, 421, 370-373.
17. Motter, B.C. (1994) Neural correlates of feature selective memory and pop-out in extrastriate area V4. J. Neurosci., 14, 2190-2199.
18. Nakahara, H., Wu, S., Amari, S. (2001) Attention modulation of neural tuning through peak and base rate, Neural Comput., 13, 2031-2047.
19. Reynolds, J.H., and Desimone, R. (1999) The Role of Neural Mechanisms of Attention in Solving the Binding Problem, Neuron, 24:19-29.
20. Reynolds, J.H., Chelazzi, L., Desimone R. (1999) Competetive mechanism subserve attention in macaque areas V2 and V4, J. Neurosci., 19, 1736-1753.
21. Reynolds, J.H., Pasternak, T., Desimone, R. (2000) Attention increases sensitivity of V4 neurons. Neuron, 26, 703-714.
22. Spratling MW, Johnson MH. (2004) A feedback model of visual attention. J Cogn Neurosci. 16:219-237.
23. Treue, S., Maunsell, J.H. (1999) Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas, J. Neurosci., 19, 7591-7602.
24. Treue, S., Martínez Trujillo, J.C. (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature, 399, 575-579.
25. Tsotsos JK, Culhane SM, Wai W, Lai Y, Davis N, Nuflo F (1995) Modeling visual attention via selective tuning. Artificial Intelligence, 78:507-545.
26. Usher, M., Niebur, E. (1996) Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. J. Cog. Neurosci., 8, 311-327.