Available online at www.sciencedirect.com
ScienceDirect
BioSystems 86 (2006) 91–99
Bio Systems
www.elsevier.com/locate/biosystems

Modeling feature-based attention as an active top-down inference process

Fred H. Hamker*

Allgemeine Psychologie, Psychologisches Institut II, Westf. Wilhelms-Universität, Fliednerstrasse 21, 48149 Münster, Germany

Received 15 December 2005; received in revised form 14 March 2006; accepted 17 March 2006

Abstract

Vision is a crucial sensor. It provides a very rich collection of information about our environment. The difficulty in vision arises, since this information is not obvious in the image, it has to be constructed. Wheres earlier approaches have favored a bottom-up approach, which maps the image onto an internal representation of the world, more recent approaches search for alternatives and develop frameworks which make use of top-down connections. In these approaches vision is inherently a constructive process which makes use of a priory information. Following this line of research, a model of primate object perception is presented and used to simulate an object detection task in natural scenes. The model predicts that early responses in extrastriate visual areas are modulated by the visual goal.

Keywords: Vision; Top-down; Bottom-up; Feature-based attention; Computational model; Gain control; V4

1. Introduction

Object recognition, generally implemented in a hierarchical bottom-up process (Fukushima, 1980; Perrett and Oram, 1993; Wallis and Rolls, 1997; Riesenhuber and Poggio, 1999) in which the complexity of representation along with the receptive field size increases, leads to a strong overlapping of populations encoding features belonging to different objects. These ambiguities in cell populations encoding features within the same receptive field limit the use of these approaches for non-segmented scenes like natural images.

The closely linked paradigms of active vision, purposive vision and animate vision (Aloimonos, 1993; Ballard, 1991) have proposed that bottom-up directed vision is an ill-posed-problem and suggested each task requires its own specific algorithm. In this regard, an universal, general vision is not possible. According to these paradigms, the fundamental problem of vision is the selection of the relevant information within the scene and the computation of an appropriate representation. An "active" vision system – in the sense of a visually selective device – is able to acquire the necessary information on demand by focusing on the relevant areas within the visual scene and taking different views from the same object.

The approach of "Deictic Codes for the Embodiment of Cognition" aims to provide a framework for describing the phenomena that appear at about one-third of a second in the perception–action process (Ballard et al., 1997). Deictic primitives dynamically refer to points in the world with respect to their crucial describing features (e.g., color or shape). The outcome of the processing after one-third second, which is the natural sequentiality of body movements can be matched to the natural computational economies of sequential decision systems through a system of implicit reference (called deictic) in which pointing movements are used to bind objects in the world

* Tel.: +49 251 8334171; fax: +49 251 8334180.
E-mail address: fhamker@uni-muenster.de (F.H. Hamker).

0303-2647/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2006.03.010

to cognitive programs. Ballard et al. (1997) suggested visual routines (Kosslyn, 1994; Ullman, 1984; Just and Carpenter, 1976) to divide one complex task into subtasks, such as selection and identification.

Selective perception has been addressed in attention related experimental frameworks such as visual search. The basic idea is that once an object is selected by a focus of attention it can be connected to an internal pointer and being processed in high-level areas. This view has its origin in the classical approach of perception that separates between a pre-attentive and attentive stage (Treisman and Gelade, 1980). Computer implementations of these types of models use a saliency map to indicate a location of interest (Koch and Ullman, 1985; Wolfe, 1994; Itti and Koch, 2000) and compute a focus of attention that selects an object (Olshausen et al., 1993). This focus could be guided by some rough knowledge about an object, such as its color. Feature-based attention is left to only guide the selection process by weighting the input into the saliency map (Wolfe, 1994; Milanese et al., 1995; Navalpakkam and Itti, 2005).

We have developed an alternative approach in which feature-based attention acts on the object representations itself. Spatially selective attentive binding, however, occurs through reentrant oculomotor loops. The search for an object or just parts of it produces top-down expectations, which meet the bottom-up processed stimulus features in the ventral pathway. This initiates a dynamic and distributed recognition process at different levels of the hierarchy by enhancing the features of interest. At higher areas these are typically complex patterns. At lower levels these complex patterns have to be decomposed into more simple patterns. Thus, top-down inference has to rely on reverse weights to decompose a pattern into its parts. By competitive interactions such a mechanism would allow to flexibly filter out the information which is inconsistent with the high-level goal description. However, the sensory evidence of the encoded items does not always allow to rule out all objects but one. This top-down inference only strengthens the expected features, which are not necessarily the to be reported ones, and guides goal-directed behavior. Thus, in parallel, areas responsible for oculomotor selection start to plan appropriate responses. Specifically, the target location of the planed eye movement is used for a location specific inference operation which in turn filters out objects at irrelevant locations. This spatial attention effect could be interpreted as a shortcut of the actual planned eye movement. It facilitates planning processes to evaluate the consequences of the planned action. As a result of both inference operations, the high-level goal description is bound to an object in the visual world.

In this approach vision is an active, dynamic and constructive process. It allows a more close look onto the processes of binding objects in the world to cognitive programs that act within one-third of a second. Our proposed concept relies on top-down connections in vision, which have been discussed and its usefulness has been demonstrated for several times (Grossberg, 1980; Mumford, 1992; Ullman, 1995; Tononi et al., 1992; Tsotsos et al., 1995; Rao and Ballard, 1999; Rao, 1999; Hamker, 1999; Engel et al., 2001; Hamker and Worcester, 2002; Corchs and Deco, 2002; Hochstein and Ahissar, 2002; Rao, 2004; Hamker, 2004b). However, top-down connections have not been used in an unequivocal fashion. The generative approach (Mumford, 1992; Olshausen and Field, 1997; Rao, 1999) predicts that the top-down signal is subtracted from the bottom-up signal. Such models predict a reduction of activity when the predicted input matches with the actual input. Our model predicts an enhancement, as previously suggested by ART (Grossberg, 1980). We have shown that this is consistent with cell recordings in IT, V4 and FEF in visual search (Hamker, 2005a) and in other attentional experiments (Hamker, 2004a,b). Since these simulations have been done with artificial inputs, we have recently scaled up this model to simulate object detection (Hamker, 2005c,d) and change detection (Hamker, 2005b) tasks in natural scenes. Here, we will focus on feature-based attention in area V4/TEO with respect to the search task.

2. The model

2.1. Anatomical, pysiological and behavioral evidence

The brain has developed specific functional areas in the visual cortex, which can be divided into two major streams. Form and color travel from V1 to V2, V4 of the occipital lobe into TEO and TE of the inferior temporal lobe (Zeki, 1978; Livingstone and Hubel, 1988). This ventral pathway is known to encode object identity. It is generally accepted that the complexity of encoded features increases along the ventral pathway. V1 neurons can be driven by simple properties of a stimulus, such as the orientation of a bar. TE neurons, however, encode highly sophisticated shape properties. These "experts" have probably evolved to meet the statistics of stimuli we typically encounter. The receptive field size has also been suggested to increase along the ventral pathway as well. Most of the receptive field size mappings have been done with anestesized monkeys. The idea is that the increasing receptive field size supports location

invariance, since a cell in higher areas is less sensitive to the position of a stimulus. The ventral pathway has been often proposed of having a limited capacity. Referring to a bottleneck, the processing in early stages is supposed to operate in parallel, whereas further processing in higher areas has been proposed to require stimulus selection (Treisman and Gelade, 1980; Koch and Ullman, 1985; Wolfe, 1994). Recent findings suggest that these conclusions might depend on the artificial stimuli used in attentional experiments. In statistically richer data sets, such as natural scenes, the similarity between target and distractors is probably much lower than in most of the artificial stimuli used. For example, it has been shown that the detection of animals in natural scenes is easy even in dual-task conditions (Li et al., 2002; Rousselet et al., 2004). These findings suggest that a strong capacity limitation in the ventral pathway is an overestimation due to the used stimulus set. The findings, however, could alternatively be taken as evidence that we can do all processing of natural scenes within the feedforward sweep. Again, this might depend on the stimuli used: the detection of trained and thus familiar objects placed into natural scenes requires monkeys to serially search for the target (Sheinberg and Logothetis, 2001).

In order to give priority to one pattern over the other, e.g., by attending to the location of one stimulus, a top-down bias was proposed (Desimone and Duncan, 1995; Reynolds et al., 1999). Similarly, a feature-based mechanism could allow to emphasize one pattern over others (Chelazzi et al., 1998; Treue and Martínez Trujillo, 1999). Thus, more general, feedback allows to resolve ambiguities and to reveal visual details. Extending the idea of a mere "attentional" bias, I propose that a target template travels the ventral pathway downwards via massive feedback connections (Rockland and van Hoesen, 1994; Rockland et al., 1994) and enhances the firing rate of cells supporting the target template by the proposed inference approach. This mechanism implements a dynamic filter to compute the relevant properties within a general purpose machinery of localized experts.

The target template might be generated in ventrolateral prefrontal cortex. Dorsolateral PFC (areas 8, 9 and 46), which includes the frontal eye field (area 8), is often reported to code location and motor mapping and the ventrolateral part (areas 45 and 47) is more devoted to categories and features. However, these areas also contain cells that show a dependency on both the location and the feature (Rao et al., 1997). There is direct evidence that prefrontal cortex is involved in providing visual, top-down directed cues (Tomita et al., 1999). I hypothesize, that the target template is not limited to simple properties, such as color, it can be highly complex. Recent evidence suggests that the prefrontal cortex relies on an adaptive neural coding (Duncan, 2001) which could compute and provide a rich target template given the context of the present task to perform. Thus, I suggest that the prefrontal cortex guides visual perception by generating an appropriate target template in time, which is then used for an inference mechanism implemented by the visual system.

I have proposed that occulomotor areas responsible for planning an eye movement, such as the frontal eye field, influence perception prior to the eye movement (Hamker, 2003). The activity reflecting the planning of an eye movement reenters the ventral pathway and provides a spatially selective expectation signal. As one of many possible, I focus on the FEF as a putative source of this reentry signal (Hamker, 2005a). The FEF has connections to occipital, temporal and parietal areas, the thalamus, superior colliculus and prefrontal cortex (Schall et al., 1995). It can be subdivided into a lateral and medial part. The lateral FEF, which generates short and precise saccades is connected to the dorsal (LIP, MT, MST and V3) and ventral (TEO, V4 and V2) pathways and the ventrolateral prefrontal cortex (Schall et al., 1995). The projections from V2 and V3 are weak, while the one from V4 are intermediate. Strong projections from TEO, MT and MST suggest that the FEF uses features after several stages of processing for target selection (Schall et al., 1995). The neurons in the FEF can be categorized based on both their responses to visual stimuli and to saccade execution into visual, visuomovement, fixation and movement cells (Bruce and Goldberg, 1985; Schall et al., 1995). There is some recent experimental evidence that the source of the reentry signal is indeed the FEF (Moore and Armstrong, 2003), but other sources cannot be excluded.

2.2. Population-based inference

Population coding has been frequently used as a theoretical basis for describing computation in the brain. Much emphasis has been given to investigate how a population encodes a stimulus. Our population-based inference approach provides a framework to continuously update the conspicuity of an internal variable using prior knowledge in form of generated expectations. The population is represented by a set of cells. The selectivity of each cell is defined by its location $i \in \{1, \ldots, 20\}$ in the population and its activity r_i reflects the conspicuity of its preferred stimulus. Each cell is simulated by an ordinary differential equation, that governs its average firing rate over time. Thus, the model allows to describe the temporal change of activity induced by top-down inference.

In abstract terms, the top-down signal represents the expectation $\hat{r}$ to which the input (observation) $r^{\uparrow}$ is compared. If the observation is similar to the expectation, the conspicuity is increased. This increase is implemented as a gain control mechanism on the feedforward signal. As far as feature-based attention is concerned a cell's response in V4 $r_{d,i,\mathbf{x}}(t)$ at location $\mathbf{x}$, selective dimension d and preferred feature i can be computed over time by a differential equation (with a time constant τ):

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} r^{\mathrm{V4}}_{d,i,\mathbf{x}} = I^{\uparrow}_{d,i,\mathbf{x}} + I^{N}_{d,i,\mathbf{x}} + I^{A}_{d,i,\mathbf{x}} - I^{\mathrm{inh}}_{d,\mathbf{x}} \quad (1)$$

The activity of a V4 cell is primarily driven by its bottom-up input $I^{\uparrow}$. Activity-dependent inhibition $I^{\mathrm{inh}}_{d,\mathbf{x}}$ introduces competition among cells and normalizes the cell's response. $I^{N}_{d,i,\mathbf{x}}$ describes the lateral excitatory influence of other cells in the population. Feature-based attention is a result of the bottom-up signal $I^{\uparrow}_{d,i,\mathbf{x}}$ modulated by the feedback signal from TE $r^{\mathrm{TE}}_{d,j,\mathbf{x}'}$ with $w^{\mathrm{IT,V4}}_{i,j,\mathbf{x},\mathbf{x}'}$ as the strength of the feedback connection:

$$I^{A}_{d,i,\mathbf{x}} = I^{\uparrow}_{d,i,\mathbf{x}} [\alpha - r^{\mathrm{V4}}_{d,k,\mathbf{x}}]^{+} \cdot \max_{j,\mathbf{x}'} (w^{\mathrm{TE,V4}}_{i,j} \cdot r^{\mathrm{TE}}_{d,j,\mathbf{x}'}) \quad (2)$$

$[\alpha - r^{\mathrm{V4}}_{d,k,\mathbf{x}}]^{+}$ implements a saturation of the gain for salient stimuli, since the expression is zero for negative arguments (Hamker, 2005d). The proposed population-based inference mechanism has been developed to capture the essential observations of attention on the population level. It has been demonstrated on data of spatial (Hamker, 2004a) and feature-based attention effects (Hamker, 2004b). Consistent with the Feature-Similarity Theory (Treue and Martínez Trujillo, 1999), the enhancement of the gain depends on the similarity between the input and the feedback signal. A number of studies have investigated various effects of dynamic gain changes. Such effects might occur on the biophysical level of a single neuron or on the network level. Correlated activity of the input could be another mechanism. Observations revealed an enhanced correlated activity in the gamma band (roughly 30–80 Hz) prior to any gain increase (Fries et al., 2001), which in turn could increase the gain (Salinas and Sejnowski, 2001; Azouz and Gray, 2003). A model based on the idea of synchrony has been proposed by Tiesinga (2005).

2.3. Network model

We simulate the interactions between areas on the level of a population code. In this model, neural populations are defined in a space spanned by the feature selectivity i and spatial selectivity $\mathbf{x} \in (x_1, x_2)$ of the cells. The variable d refers to different channels computed from the image such as orientation (O), intensity (I) or red–green (RG), blue–yellow (BY), or spatial resolution (σ). The conspicuity of each encoded feature is altered by the target template. A target encoded in prefrontal cortex defines the expected features $r^{\mathrm{PF}}_{d,i}$ (Fig. 1). We infer the conspicuity of each feature in TE denoted as $r^{\mathrm{TE}}_{d,i,\mathbf{x}}$ by comparing the expected features $r^{\mathrm{PF}}_{d,i}$ with the observation, i.e. the bottom-up input $r^{\mathrm{TE}\uparrow}_{d,i,\mathbf{x}}$. If the observation is similar to the expectation we increase the conspicuity. Such a mechanism enhances in parallel the conspicuity of all features in TE which are similar to the target template. The same procedure is performed in V4 to compute the conspicuity $r^{\mathrm{V4}}_{d,i,\mathbf{x}}$ where the expected features are the ones encoded in TE. The model is simplified in two aspects. Firstly, the high-level goal description is not constructed by the model on its own but a target template is presented to the model. Secondly, the target template is defined only in low level feature space. This constraint occurs, since the complexity of the feature space does not increase along the models "what" pathway.

In order to detect an object in space the conspicuities $r^{\mathrm{V4}}_{d,i,\mathbf{x}}$ and $r^{\mathrm{TE}}_{d,i,\mathbf{x}}$ are combined across all channels d and encoded in the frontal eye field visuomovement cells. The projection from the visuomovement cells to the movement cells generates an expectation in space $r^{\mathrm{FEFm}}_{\mathbf{x}}$. Thus, a location with high conspicuity in different channels d tends to have a high expectation in space $r^{\mathrm{FEFm}}_{\mathbf{x}}$. Analogous to the inference in feature space the expected location $r^{\mathrm{FEFm}}_{\mathbf{x}}$ is iteratively compared with the observation $r^{V4\uparrow}_{d,i,\mathbf{x}}$ in $\mathbf{x}$ and the conspicuity of a feature with a similarity between expectation and observation is enhanced. 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.

We now briefly explain the simulated areas in the model. A detailed description can be found in (Hamker, 2005c).

Early visual processing: Feature maps for red–green opponency (RG), blue–yellow opponency (BY), intensity (I), orientation (O), and spatial resolution (σ) are computed. The initital conspicuity is determined by center-surround operations (Itti and Koch, 2000). Center-surround operations calculate the difference of feature values in maps with a fine scale and a coarse scale and thus, the obtained conspicuity value is a measure of stimulus-driven saliency. The feature information and the

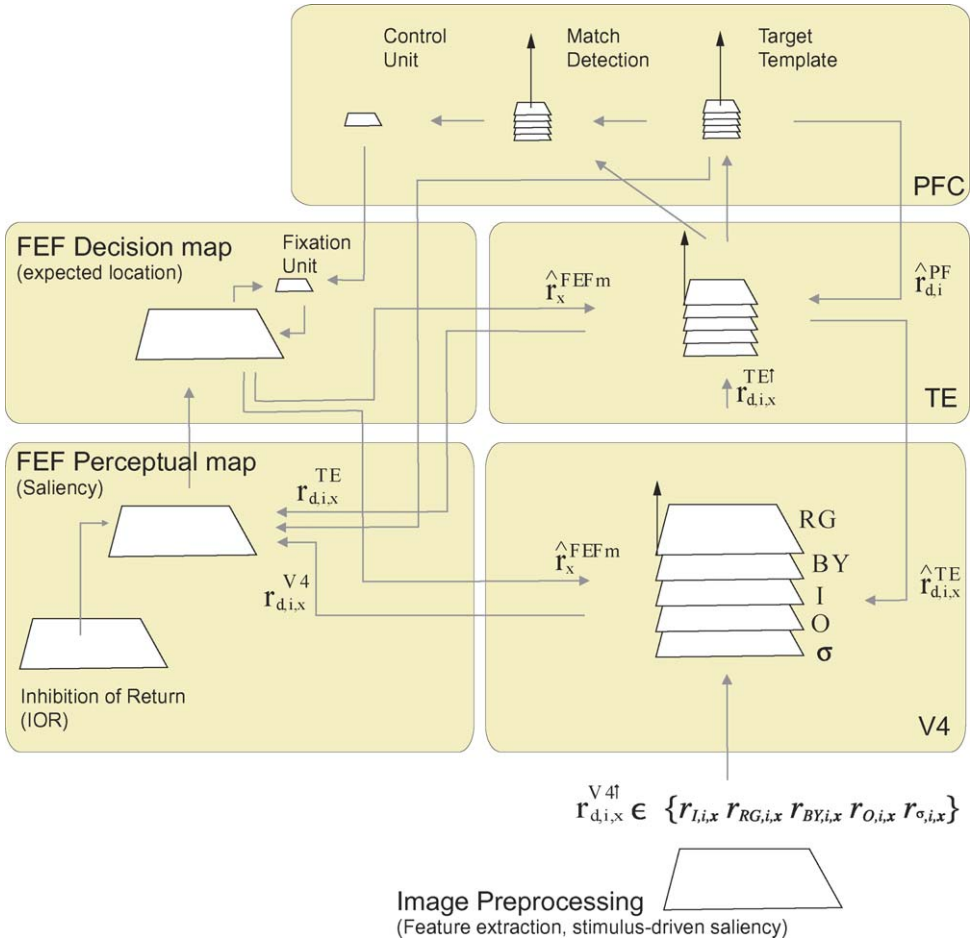

Fig. 1. Model for object detection in natural scenes. From the image, the features of five channels (RG, BY, I, O, σ) are obtained. For each feature we also compute its conspicuity as determined by the spatial arrangement of the stimuli in the scene and represent both aspects within a population code, so that at each location a feature and its related conspicuity is encoded. This initial, stimulus-driven conspicuity is now dynamically updated within a hierarchy of levels. From V4 to TE, a pooling across space is performed to obtain a representation of features with a coarse coding of location. The target template encodes features of the target object by a population of sustained activated cells. It represents the expected features $\hat{r}^{\mathrm{PF}}_{d,i}$ which are used to compute the (posterior) conspicuity in TE. Similarly, TE represents the expectation for V4. As a result, the conspicuity of all features of interest is enhanced regardless of their location in the scene. In order to identify candidate objects by their saliency the activity across all five channels is integrated in the FEF perceptual map. The saliency is then used to compute the target location of an eye movement in the FEF decision map. The activity in this map $\hat{r}^{\mathrm{FEFm}}_{\mathbf{x}}$ is fed back, which in turn enhances the conspicuity of all features in V4 and TE at the activated areas in the FEF decision map. Thus, objects at expected locations are preferably represented. By comparing the conspicious features in TE with the target template in the match detection units it is possible to continuously track if the object of interest is encoded in TE. Visited locations are being tagged by an inhibition of return. This allows the model to make repeated "fixations" while searching for an object.

conspicuity are used to determine a population code, so that at each location the features and their related conspicuities are encoded.

V4: V4 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, $r_{\mathrm{RG},i,\mathbf{x}}$ for red–green opponency, $r_{\mathrm{BY},i,\mathbf{x}}$ for blue–yellow opponency and $r_{\sigma,i,\mathbf{x}}$ for spatial frequency (Fig. 1). The expectation of features in V4 originates in TE $\hat{r}^{\mathrm{V4}_F}_{d,i,\mathbf{x}'} = r^{\mathrm{TE}}_{d,i,\mathbf{x}}$ and the expected location in the FEF decision map $\hat{r}^{V4_L}_{\mathbf{x}'} = r^{\mathrm{FEFm}}_{\mathbf{x}'}$. Please note that even TE has a coarse dependency on location.

TE: The features with their respective conspicuity and location in V4 project to TE, but only within the same dimension d, so that the conspicuity of features at several locations in V4 converges onto one location in TE. A map containing nine populations with overlapping receptive fields is simu-

lated. The complexity of features from V4 to TE is not increased. The expected features in TE originate in the target template $\hat{r}^{\mathrm{TE}_F}_{d,i,\mathbf{x}} = w \cdot r^{\mathrm{PF}}_{d,i}$ and the expected location in the FEF decision map $\hat{r}^{\mathrm{TE}_L}_{\mathbf{x}} = w \cdot r^{\mathrm{FEFm}}_{\mathbf{x}}$

FEF perceptual map: The FEF perceptual map indicates salient locations by integrating the conspicuity of V4 and TE across all channels. Its cells show a response which fits into the category of FEF visuomovement cells (FEFv). In addition to the conspicuity in V4 and TE the match of the target template with the features encoded in V4 is considered by computing the product $\prod_d \max_i r^{\mathrm{PF}}_{d,i} \cdot r^{\mathrm{V4}}_{d,i,\mathbf{x}}$. This implements a bias to locations with a high joint probability of encoding all searched features in a certain area.

FEF decision map: The projection of the perceptual map to the decision map transforms the salient locations into a few candidate locations, which dynamically compete for determining the target location of an eye movement. This is achieved by subtracting the average saliency from the saliency at each location $w^{\mathrm{FEFv}} r^{\mathrm{FEFv}}_{\mathbf{x}} - w^{\mathrm{FEFv}}_{\mathrm{inh}} \sum_{\mathbf{x}} r^{\mathrm{FEFv}}_{\mathbf{x}}$. Thus, the cells in the decision map show none or only little response to the onset of a stimulus, such that their response fits into the category of the FEF movement cells (FEFm). Their activity provides the expected location for V4 and TE units.

3. Results

I now demonstrate the predictions of the model on the early response of cells in extrastriate areas (specifically V4) in a visual search task using natural scenes.

An object is presented to the model for 100 ms and the model memorizes some of its features as a target template. We do not give the model any hints which feature to memorize. The model's task is to make an eye movement towards the target (Fig. 2(A and B)). When presenting the search scene, TE cells that match the target template quickly increase their activity to guide perception on the level of V4 cells. Thus, the features of the object of interest are enhanced prior to any spatial focus of attention. This feature-based attention effect allows for a goal-directed planning of a saccade in the FEF. The planning of an eye movement provides a spatially organized reentry signal, which enhances the gain of all cells around the target location of the intended eye movement. As a result of these inference operations, the high-level goal description in PFC is bound to an object in the visual world. Further simulation results are discussed in (Hamker, 2005c).

We now take a close view on the feature-based attention effects of the model. In this respect we compare two conditions: attend towards the visual properties of the lighter (Fig. 2A) and attend towards the cigarettes (Fig. 2B). Fig. 2C shows the difference activity of both conditions in V4 prior to any spatial selection as determined by a low FEFm activity ($\max r^{\mathrm{FEFm}}_{\mathbf{x}}(t) < 0.05$). Our analysis clearly shows that the activity is selectively modulated according to the task at hand. Thus, the model predicts feature-based attention effects independent of focused attention. Although the effect is global in space it can guide gaze towards the object of interest since it depends on the content encoded at each location.

To illustrate the effects of feature-based attention on the cell level we show their time course of activity. Fig. 3A shows the activity of cells with their receptive field centered on the lighter. A difference in activity between the attend lighter and attend cigarettes condition reflects the relative effect of feature-based attention. In the orientation channel (O), cell 01 shows an enhancement in the attend cigarettes condition whereas cell 08 an enhancement in the attend lighter condition. Thus, even cells with their receptive field on the lighter can be enhanced in the attend cigarettes condition. The target template for orientation in the attend lighter condition was close to horizontal and thus increased the activity of cell 08, whereas target template for orientation in the attend cigarettes condition was vertical and thus enhanced the sensitivity of cell 01 and adjacent cells. The blue color of the lighter primarily increased the activity of cells around cell 14 of the BY channel in the attend lighter condition. The white color of the cigarette box increased cell 18 of the intensity channel in the attend cigarettes condition. We observe also differences in the timing of the feature-based attention effect, which are based on recurrent interactions between V4 and TE as well as TE and PFC.

4. Discussion

We predict that goal directed, feature-based search first selectively modulates feature-sensitive cells prior to any spatial selection.

This prediction is consistent with cell recordings in visual search (Bichot et al., 2005; Ogawa and Komatsu, 2004) and recent findings in which the learning of degraded natural scenes resulted in a selective enhancement of V4 cells (Rainer et al., 2004). According to this study, V4 plays a crucial role in resolving an indeterminate level of visual processing by a coordinated interaction between bottom-up and top-down streams.

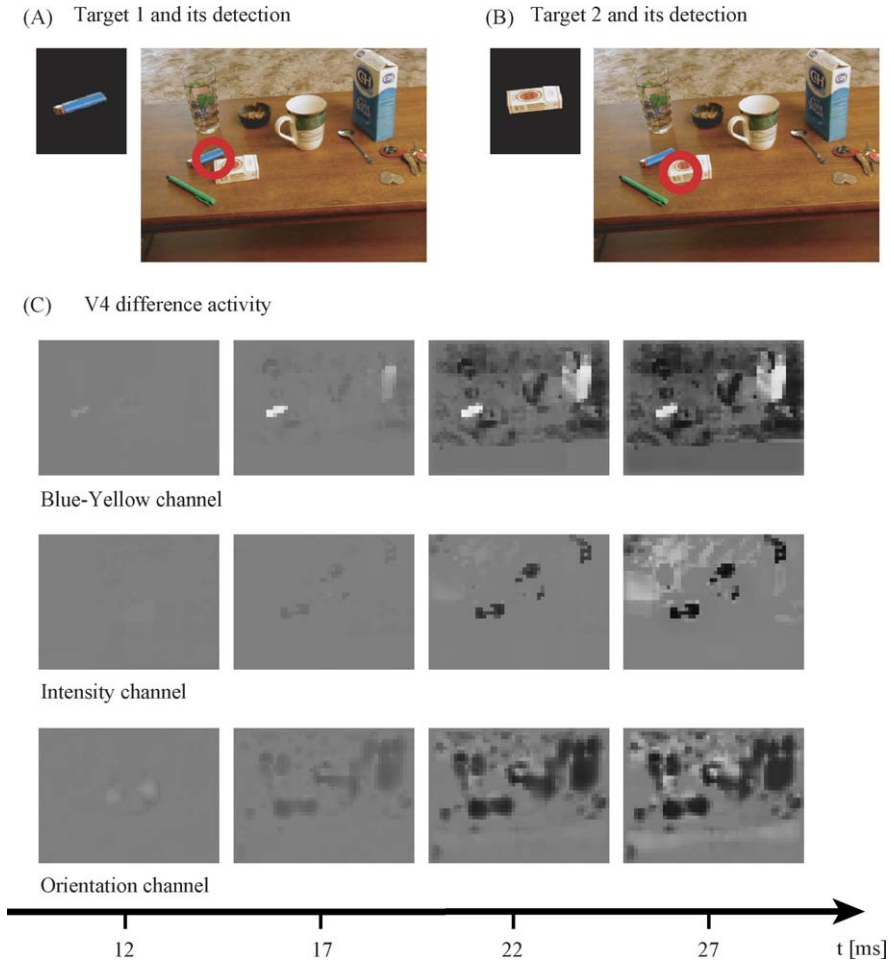

Fig. 2. Illustration of feature-based attention. (A) Target object 1 and its detection in the visual scene; (B) target object 2 and its detection in the visual scene; (C) difference activity in V4 in three channels over time. For a comparison with cell recordings a latency of about 60–80 ms has to be added to the time axis. Only the difference of the maximal activity at each location is shown irrespective of the feature selectivity. Gray areas indicate equal (maximal) activity, light areas more activity in the first condition and dark areas more activity in the second condition. We can observe that parts of the scene are relatively enhanced or reduced according to the target template.

Our model further predicts that saliency is encoded as part of the variable itself through the dual coding property of a population code. Saliency is not encoded in a single map. Thus, attentional effects can be found throughout the visual system. The observation of an attentional modulation does therefore not allow to conclude that a stimulus has been selected by a spatially directed focus. For example, V4 also provides a spatially organized map encoding saliency (Fig. 2C), which is consistent with recent findings (Mazer and Gallant, 2003). However, V4 cells are selective for location and specific features. Consistent with recordings in the FEF (Schall, 2002), the FEF visuomovement cells in our model are more related to the classical idea of a saliency map (Itti and Koch, 2001), since they solely encode location by integrating the activity across all channels and features. We assume that this information needs an additional, decisional stage of processing before it is feed back such that the saliency information is transformed into a dynamic, competitive representation of a few candidate regions.

The present model uses only simple features as detectors of visual properties. However, object recognition requires a much richer set of detectors. If we want to incorporate those into the model we have to ensure that the feedforward and feedback connections are consistent with each other. Learning appropriate feedforward and feedback connections by the statistics of the visual scene would allow to generate consistent complex feature detectors. With such an

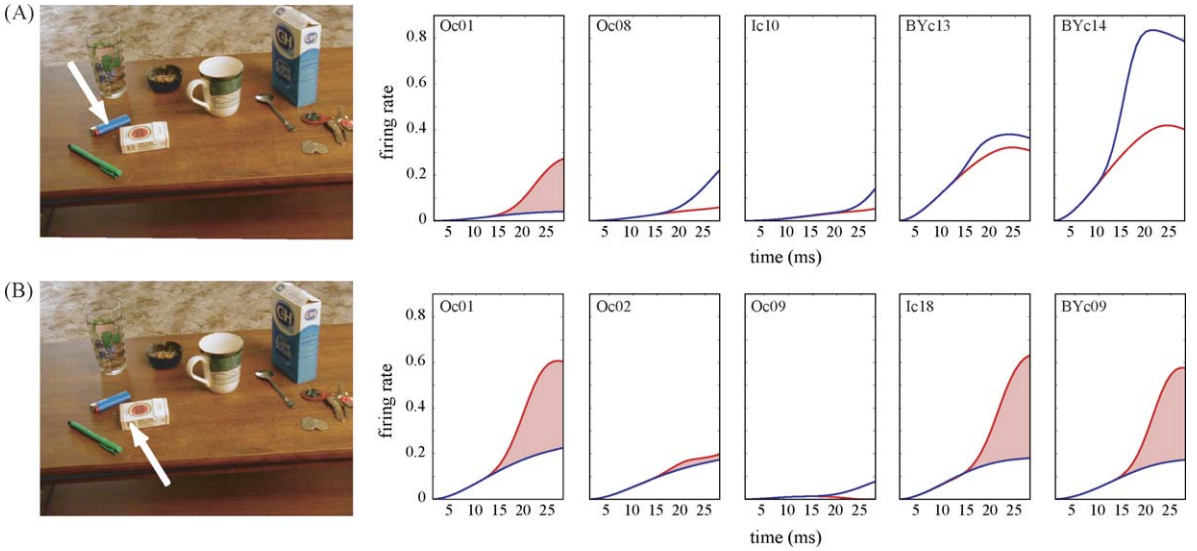

Fig. 3. Illustration of feature-based attention effects on the single cell level. The activity is shown in two conditions with time relative to search array onset (0 ms): attend towards the lighter (blue) and attend towards the cigarettes (red). The red shaded area between the curves appears when the activity in the second condition is higher. (A) Selected cells in the orientation (O), intensity (*I*) and blue–yellow (BY) channel with the receptive field center located on the lighter. (B) Selected cells in the orientation (O), intensity (*I*) and blue–yellow (BY) channel with the receptive field center located on the cigarette box.

extension the model would be able to shed more light on the puzzling issues of object recognition in natural scenes.

References

Aloimonos, Y., 1993. Introduction: active vision revisited. In: Aloimonos (Ed.), Acitve Perception. Lauwrence Erlabaum Associates, 1–18.

Azouz, R., Gray, C.M., 2003. Adaptive coincidence detection and dynamic gain control in visual cortical neurons in vivo. Neuron 37, 513–523.

Ballard, D., 1991. Animate vision. Artif. Intell. 48, 57–86.

Ballard, D., Hayhoe, M.M., Pook, P.K., 1997. Deictic codes for the embodiment of cognition. Behav. Brain Sci. 20, 723–742.

Bichot, N.P., Rossi, A.F., Desimone, R., 2005. Parallel and serial neural mechanisms for visual search in macaque area V4. Science 308, 529–534.

Bruce, C.J., Goldberg, M.E., 1985. Primate frontal eye fields. I. Single neurons discharging before saccades. J. Neurophysiol. 53, 603–635.

Chelazzi, L., Duncan, J., Miller, E.K., Desimone, R., 1998. Responses of neurons in inferior temporal cortex during memory-guided visual search. J. Neurophysiol. 80, 2918–2940.

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.

Desimone, R., Duncan, J., 1995. Neural mechanisms of selective attention. Annu. Rev. Neurosci. 18, 193–222.

Duncan, J., 2001. An adaptive coding model of neural function in prefrontal cortex. Nat. Rev. Neurosci. 2, 820–829.

Engel, A.K., Fries, P., Singer, W., 2001. Dynamic predictions: oscillations and synchrony in top-down processing. Nat. Rev. Neurosci. 2, 704–716.

Fries, P., Reynolds, J.H., Rorie, A.E., Desimone, R., 2001. Modulation of oscillatory neuronal synchronization by selective visual attention. Science 291, 1560–1563.

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36, 193–202.

Grossberg, S., 1980. How does the brain build a cognitive code?. Psychol. Rev. 87, 1–51.

Hamker, F.H., 1999. The role of feedback connections in task-driven visual search. In: Heinke, D., Humphreys, G.W., Olson, A. (Eds.), Connectionist Models in Cognitive Neuroscience. Springer-Verlag, London, 252–261.

Hamker, F. H., Worcester, J., 2002. Object detection in natural scenes by feedback. In: Bülthoff, H.H., et al., (Eds.), Biologically Motivated Computer Vision. Lecture Notes in Computer Science. Springer-Verlag, Berlin, Heidelberg, New York, 398–407.

Hamker, F.H., 2003. The reentry hypothesis: linking eye movements to visual perception. J. Vis. 11, 808–816.

Hamker, F.H., 2004a. Predictions of a model of spatial attention using sum- and max-pooling functions. Neurocomputing 56C, 329–343.

Hamker, F.H., 2004b. A dynamic model of how feature cues guide spatial attention. Vis. Res. 44, 501–521.

Hamker, F.H., 2005a. The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. Cereb. Cortex 15, 431–447.

Hamker, F.H., 2005b. A computational model of visual stability and change detection during eye movements in real world scenes. Vis. Cogn. 12, 1161–1176.

Hamker, F.H., 2005c. The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision. J. Comput. Vis. Image Understand. 100, 64–106.

Hamker, F.H., 2005d. Modeling attention: from computational neuroscience to computer vision. In: Paletta, L., et al., (Eds.), Attention and Performance in Computational Vision. Second International Workshop on Attention and Performance in Computer Vision (WAPCV 2004), LNCS 3368. Springer-Verlag, Berlin, Heidelberg, 118–132.

Hochstein, S., Ahissar, M., 2002. View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36, 791–804.

Itti, L., Koch, C., 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. 40, 1489–1506.

Itti, L., Koch, C., 2001. Computational modelling of visual attention. Nat. Rev. Neurosci. 2, 194–203.

Just, M.A., Carpenter, P.A., 1976. Eye fixations and cognitive processes. Cogn. Psychol. 8, 441–480.

Koch, C., Ullman, S., 1985. Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Psychol. 4, 219–227.

Kosslyn, S.M., 1994. Image and Brain. MIT Press (A Bradford Book), Cambridge, MA.

Li, F.-F., VanRullen, R., Koch, C., Perona, P., 2002. Rapid natural scene categorization in the near absence of attention. Proc. Natl. Acad. Sci. U.S.A. 99, 9596–9601.

Livingstone, M.S., Hubel, D.H., 1988. Segregation of form, color, movement and depth. J. Neurosci. 7, 3416–3468.

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.

Moore, T., Armstrong, K.M., 2003. Selective gating of visual signals by microstimulation of frontal cortex. Nature 421, 370–373.

Milanese, R., Gil, S., Pun, T., 1995. Attentive mechanisms for dynamic and static scene analysis. Opt. Eng. 34, 2428–2434.

Mumford, D., 1992. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol. Cybern. 66, 241–251.

Navalpakkam, V., Itti, L., 2005. Modeling the influence of task on attention. Vis. Res. 45, 205–231.

Ogawa, T., Komatsu, H., 2004. Target selection in area V4 during a multidimensional visual search task. J. Neurosci. 24, 6371–6382.

Olshausen, B.A., Anderson, C., van Essen, D., 1993. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J. Neurosci. 13, 4700–4719.

Olshausen, B.A., Field, D.J., 1997. Sparse coding with an overcomplete basis set: a strategy employed by V1?. Vis. Res. 37, 3311–3325.

Perrett, D.I., Oram, M.W., 1993. The neurophysiology of shape processing. Image Vis. Comp. 11, 317–333.

Rainer, G., Lee, H., Logothetis, N.K., 2004. The effect of learning on the function of monkey extrastriate visual cortex. PLoS Biol. 2, 275–283.

Rao, S.C., Rainer, G., Miller, E.K., 1997. Integration of what and where in the primate prefrontal cortex. Science 276, 821–824.

Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2, 79–87.

Rao, R.P., 1999. An optimal estimation approach to visual perception and learning. Vis. Res. 39, 1963–1989.

Rao, R.P., 2004. Bayesian computation in recurrent neural circuits. Neural Comput. 16, 1–38.

Reynolds, J.H., Chelazzi, L., Desimone, R., 1999. Competetive mechanism subserve attention in macaque areas V2 and V4. J. Neurosci. 19, 1736–1753.

Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nat. Neurosci. 2, 1019–1025.

Rockland, K.S., van Hoesen, G.W., 1994. Direct temporal-occipital feedback connections to striate cortex (V1) in the macaque monkey. Cereb. Cortex 4, 300–313.

Rockland, K.S., Saleem, K.S., Tanaka, K., 1994. Divergent feedback connections from areas V4 and TEO in the macaque. Vis. Neurosci. 11, 579–600.

Rousselet, G.A., Thorpe, S.J., Fabre-Thorpe, M., 2004. How parallel is visual processing in the ventral pathway?. Trends Cogn. Sci. 8, 363–370.

Salinas, E., Sejnowski, T.J., 2001. Correlated neuronal activity and the flow of neural information. Nat. Rev. Neurosci. 2, 539–50.

Schall, J.D., Morel, A., King, D.J., Bullier, J., 1995. Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. J. Neurosci. 15, 4464–4487.

Schall, J.D., 2002. The neural selection and control of saccades by the frontal eye field. Phil. Trans. R. Soc. Lond. B 357, 1073–1082.

Sheinberg, D.L., Logothetis, N.K., 2001. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. J. Neurosci. 21 1340–1350.

Tiesinga, P.H., 2005. Stimulus competition by inhibitory interference. Neural Comput. 17, 2421–53.

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., Miyashita, Y., 1999. Top-down signal from prefrontal cortex in executive control of memory retrieval. Nature 401, 699–703.

Tononi, G., Sporns, O., Edelman, G., 1992. Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. Cereb. Cortex 2, 310–335.

Treisman, A., Gelade, G., 1980. A feature integration theory of attention. Cogn. Psychol. 12, 97–136.

Treue, S., Martínez Trujillo, J.C., 1999. Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399, 575–579.

Tsotsos, J.K., Culhane, S.M., Wai, W., Lai, Y., Davis, N., Nuflo, F., 1995. Modeling visual attention via selective tuning. Artif. Intell. 78, 507–545.

Ullman, S., 1984. Cognition 18, 97–157.

Ullman, S., 1995. Sequence seeking and counter streams: a computational model for bidirectional flow in the visual cortex. Cereb. Cortex 5, 1–11.

Wallis, G., Rolls, E.T., 1997. Invariant face and object recognition in the visual system. Prog. Neurobiol. 51, 167–194.

Wolfe, J.M., 1994. Guided search 2.0: a revised model of visual search. Psychonom. Bull. Rev. 1, 202–238.

Zeki, S., 1978. Functional specialization in the visual cortex of the rhesus monkey. Nature 274, 423–428.