

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this report is estimated to be 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, completing and reviewing the collection of information, sending comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE		3. REPORT TYPE AND DATES COVERED FINAL 01 Sep 92 TO 31 Aug 95	
4. TITLE AND SUBTITLE TOWARD A NEUROBIOLOGICAL THEORY OF VISUAL ATTENTION				5. FUNDING NUMBERS F49620-92-J-0454 61102F 2313/AS	
6. AUTHOR(S) Dr Christoff Koch				7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Dept of Computation & Neural Systems Division of Biology 216-76 California Institute of Technology Pasadena CA 92215	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NL 110 Duncan Ave Suite B115 Bolling AFB DC 20332-8080 Dr John F. Tangney				10. SPONSORING / MONITORING AGENCY REPORT NUMBER AFOSR-TR-95 0656	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION / AVAILABILITY STATEMENT approved for public release distribution unlimited.				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The purpose of this effort was to study visual, focal selective attention and its implementation in the primate visual system from a computational point of view. It is known that at the neuronal level, two cortical pathways exist that are responsible for mediating attention: the where pathway that selects interesting or conspicuous locations and the what pathway that identifies and recognizes objects. In the effort, it was shown how neuronal networks based on those found in the cerebral cortex can implement these pathways using real images. In particular, the use of a saliency map, that encodes how "interesting" or "salient" locations are in the visual field (rather than what features are present at these locations) represents a powerful strategy to aid visual search. These algorithms are being ported onto Pentium-based machines for various machine-vision applications.					
DTIC QUALITY INSPECTED 8					
14. SUBJECT TERMS 19951017 035				15. NUMBER OF PAGES	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT (U)		18. SECURITY CLASSIFICATION OF THIS PAGE (U)		19. SECURITY CLASSIFICATION OF ABSTRACT (U)	
20. LIMITATION OF ABSTRACT (U)					

Final Report

Contract No. AFOSR F49620-92-J-0454

Towards a Neurobiological Theory of Visual Attention

Christof Koch

Professor

California Institute of Technology

September, 1. 1992 - August 31, 1995

**California Institute of Technology
Division of Biology, 139-74, Pasadena, CA 91125**

Technical Progress

The purpose of this effort was to study visual, focal selective attention and its implementation in the primate visual system from a computational point of view. It is known that at the neuronal level, two cortical pathways exist that are responsible for mediating attention: the *where* pathway that selects interesting or conspicuous locations and the *what* pathway that identifies and recognizes objects. In the effort, it was shown how neuronal networks based on those found in the cerebral cortex can implement these pathways using real images. In particular, the use of a *saliency map*, that encodes how "interesting" or "salient" locations are in the visual field (rather than what features are present at these locations) represents a powerful strategy to aid visual search. These algorithms are being ported onto Pentium-based machines for various machine-vision applications.

Introduction

The computations of early vision are essentially parallel operations, *i.e.*, they are applied in parallel to all parts of the visual field. This high degree of parallelism cannot be sustained in intermediate and higher vision because of the astronomical number of different possible combination of features. Therefore, it becomes necessary to select only a part of the instantaneous sensory input for more detailed processing and to discard the rest. This is the mechanism of visual selective attention which we set out to study by computational methods.

Primate vision is organized along two major anatomical pathways. One of them is concerned mainly with *object recognition*. For this reason, it has been called the *What* pathway; for anatomical reasons, it is also known as the ventral, or occipito-temporal, pathway. The principal task of the other major pathway is the determination of the *location* of objects and therefore it is called the *Where* pathway or, again for anatomical reasons, the dorsal, or occipito-parietal, pathway.

Most of our work during the first period of the grant period was devoted to the development of model for the implementation of the *What* pathway. The underlying mechanism is "temporal tagging:" it is assumed that the attended region of the visual field is distinguished from the unattended parts by the temporal fine-structure of the neuronal spike trains. We have shown that temporal tagging can be achieved by introducing oscillations (Niebur, Koch and Rosin, 1993) or moderate levels of correlation among groups of cells (Niebur and Koch, 1994) and that the tag thus generated may be available at all stages of the perceptual hierarchy.

How can such temporal modulation be obtained? Periodicity can be generated by subthreshold additional input. To generate synchronous

Codes

activity, we have suggested a simple mechanism, namely a very brief common input to all cells which respond to attended stimuli. Such (excitatory) input will increase the propensity of postsynaptic cells to fire for a very short time after receiving this input, and thereby increase the correlation between spike trains without necessarily increasing the average firing rate.

Over the last year, we have developed a model of the control system which generates such modulating input. We have shown that it is possible to construct an integrated system of attentional control based on neuronally plausible elements and which is compatible with the anatomy and physiology of the primate visual system. The system scans a visual scene and identifies its most salient parts.

We have also pursued the question of the neuronal implementation of *feature-based* attention. We showed (Usher and Niebur, 1995) that the dynamical response of cortical IT neurons in a delayed-match-to-sample task can be understood as an attentional modulation which is mediated by a simple feedback mechanism involving prefrontal cortex. This study complements our work on location-based attention and generalizes our model to incorporate those attentional phenomena which do not depend on the activity in a saliency map.

Finally, in collaboration with Dr. Francis Crick at the Salk Institute we have continued to pursue the question of the neuronal correlate of awareness (Crick and Koch, 1995). More specifically, in what neurons in what brain areas does attention and the content of visual awareness arise. We conclude that neurons that mediate visual awareness must directly project to the planning stages of the brain, that is to pre-frontal cortical areas. We therefore conclude that the firing of neurons in primary visual cortex, while necessary for most of conscious vision, do not directly cause visual awareness. This proposal has caused a lot of controversy, with a number of laboratories actively working on evaluating its experimental status. Indeed, one such psychophysical study, originating in our laboratory, has just appeared in *Nature* (Kolb and Braun, 1995).

Because our model of the dorsal *What* pathway has interesting implications for the manner in which computer-vision systems allocate their computational resources (see also our "Future outlook" section), we will describe it in more detail.

A Simple Model of The Dorsal Pathway

Overall Structure

Figure~1 shows an overview of the model *Where* pathway. Input is provided in the form of digitized images from an NTSC camera which is then analyzed in various feature maps. These maps are organized around the known operations in early visual cortices. They are implemented at

different spatial scales, and each feature is computed in a center-surround structure akin to visual receptive fields (Abelson *et al.*, 1984). The features implemented so far are the three principal components of primate color vision (intensity, red-green, blue-yellow), four orientations, and temporal change.

Target Selection and the Saliency Map

The task of the saliency map is the computation of the salience at every location in the visual field and the subsequent selection of the most salient areas or objects. At any time, only one such area is selected. The feature maps provide current input to the saliency map. The output of the saliency map consists of a spike train from neurons corresponding to this selected area in the topographic map which project to the ventral ("What") pathway. By this mechanism, they are "tagged" by modulating the temporal structure of the neuronal signals corresponding to attended stimuli.

Once all relevant features have been computed in the various feature maps, they have to be combined to yield the salience, *i.e.* a scalar quantity. In our model, we solve this task by simply adding the activities in the different feature maps (reduced to the size of the saliency map using a Gaussian pyramid).

At any given time, the maximum of this map is therefore the most salient stimulus. As a consequence, this is the stimulus to which the focus of attention should be directed next to allow more detailed inspection by the more powerful "higher" process which are not available to the massively parallel feature maps. This maximum is selected by application of a winner-take-all mechanism.

For a static image, the system would so far attend continuously the most conspicuous stimulus. This is neither observed in biological vision nor desirable from a functional point of view; instead, after inspection of any point, there is usually no reason to dwell on it any longer and the next-most salient point should be attended.

We achieve this behavior by introducing feedback from the winner-take-all array. When a spike occurs in the WTA network, the integrators in the saliency map receive additional input with the spatial structure of an inverted Mexican hat, *ie.* a difference of *i.e.* The (inhibitory) center is at the location of the winner which becomes thus inhibited in the saliency map and, consequently, attention switches to the next-most conspicuous location. At the same time, the system avoids to return to the same location it has just visited, a phenomenon which is well-known in psychophysics under the term "inhibition of return."

Simulation Results

We have studied the system with inputs constructed analogously to typical visual psychophysical stimuli. For instance, bright spots in dark backgrounds (or dark spots in bright backgrounds) are reliably detected, and the focus of attention immediately jumps to such stimuli. If there is more than one such stimulus, the system scans them one-by-one, in the order of decreasing contrast from the background. The same is true for stimuli which have a color or orientation different from that of the background, or for moving stimuli in front of a static background. In all cases, the elements can have different sizes (*i.e.* different spatial dimensions). The system was also applied to images of natural and artificial environments (recorded with a commercial NTSC camcorder) and was in general successful in selecting the most salient portions of the scenes (see Fig.~2).

Space limitations prevent a detailed presentation of these results in this report. Some of the results have been published already (Niebur and Koch, 1995) others are contained in a forthcoming publication.

Conclusion And Outlook

We present in this final technical report a prototype for an integrated system mimicking the control of visual selective attention. Our model is compatible with the known anatomy and physiology of the primate visual system, and its different parts communicate by signals which are neurally plausible. The model identifies the most salient points in a visual scenes one-by-one and scans the scene autonomously in the order of decreasing saliency. This allows the control of a subsequently activated processor which is specialized for detailed object recognition. At present, saliency is determined by combining the input from a set of feature maps with fixed weights. In future work, we will generalize our approach by introducing plasticity in these weights and thus adapting the system to the task at hand.

Partially funded by a Multidisciplinary Research Program (MURI) via the Office of Naval Research (ONR), we are now porting this suite of programs onto pentium-based machines for studying machine vision application. In particular, we are focusing on the problem of very rapid identification of particular features (e.g. faces or weapons) in video sequences via a saliency-map based procedure. This involves the need to learn how the various features map can be mapped onto the saliency map (in a supervised or unsupervised manner) to assure optimal detection.

Figure 1: Overview of the model *Where* pathway. Features are computed as center-surround differences at 4 different spatial scales (only 3 feature maps shown). They are combined and integrated in the saliency map ("SM") which provides input to an array of integrate-and-fire neurons with global inhibition. This array ("WTA") has the functionality of a winner-take-all network and provides the output to the ventral pathway ("V2") as well as feedback to the saliency map (bold arrow).

Figure 2: Example of the performance of the attentional control system. Input to the system is a picture of the Caltech bookstore, here reproduced in gray-scale. Top left: The most salient point in the image is a red banner on the wall of the bookstore (in the center of the image). Therefore, this becomes the location where the focus of attention is directed to first (shown as white square). Top left, top right, bottom left: Trajectory of the focus of attention, shown as dark line, at the simulated times listed above the respective images. The focus of attention shifts about every 35 msec to the next-most salient location. Note that at t=280 msec (top right), the continuously high input leads to a re-focusing on the location attended first (at t=140 msec). However, the system does not enter in a loop as is seen at t=315 msec (bottom left). Bottom right: After t=540 msec, the system has visited 11 locations (shown by white squares; one location visited twice) and scanned a significant portion of the image. Bottom center: Instantaneous activity in the saliency map at the times corresponding to the images pointed to by the arrows. Note the dark "blobs" caused by the inhibitory feedback at the just-visited location. The center-bottom map (without arrow) shows the input to the saliency map, *i.e.* the sum of all feature maps.

Personnel

Number of researchers working on this project:

Christof Koch, PhD, Professor

Ernst Niebur, PhD, Senior Research Fellow

Marius Usher, PhD, Post-Doctoral Fellow

Bartlet Mel, PhD, Senior Research Fellow

William Softky, PhD, Research Fellow

Mike Harville, CNS graduate student

Jiajun Wei, Physics Graduate Student

Chris Waite, Undergraduate student

Publications

All of the following publications result from the effort summarized above and explicitly acknowledge the support of the Air Force Office of Scientific Research.

Papers published in refereed journal

Crick, F. and Koch, C. The problem of consciousness. *Scientific American*, **267**(3): 153-159, 1992.

Koch, C. and Schuster, H. A simple network showing burst synchronization without frequency-locking. *Neural Computation*, **4**: 211-223, 1992.

Koch, C. Computational approaches to cognition: the bottom-up view. *Current Opinion in Neurobiology*, **3**: 203-208, 1993.

Niebur, E., Koch, C. and Rosin, C. An oscillation-based model for the neuronal basis of attention. *Vision Research* **33**: 2789-2802, 1993.

Bernander, Ö., Koch, C. and Usher, M. The effect of synchronized inputs at the single neuron level. *Neural Computation* **6**: 622-641, 1994.

Bair, W., Koch, C., Newsome, W. and Britten, K. Power spectrum analysis of bursting cells in area MT in the behaving monkey. *Journal of Neuroscience*, **14**: 2870-2892, 1994.

Niebur, E. and Koch, C. A Model for the Neuronal Implementation of Selective Visual Attention Based on Temporal Correlation Among Neurons, *Journal of Computational Neuroscience* **1**: 141-158, 1994.

Niebur, E. and Wörgötter, F. Design principles of columnar organization in visual cortex. *Neural Computation* **6**: 602-614, 1994.

Brettell, D.W. and Niebur, E. A detailed simulator of a biological neural network implemented on a massively parallel supercomputer. *IEEE Computational Science and Engineering* **1** (4): 31-43, 1994.

Usher, M., Stemmler, M., Koch, C. and Olami, Z. Network amplification of local fluctuations causes high spike rate variability, fractal firing patterns and oscillatory local field potentials. *Neural Computation* **6**: 795-836, 1994.

Fahle, M. and Koch, C. Spatial displacement, but not temporal asynchrony, destroys figural binding, *Vision Research* **35**: 491-494, 1995.

Crick, F. and Koch, C. Are we aware of neural activity in primary visual cortex? *Nature*, **375**: 121-123, 1995.

Koch, C., Bernander, Ö. and Douglas, R.J. Do neurons have a voltage or a current threshold for action potential initiation? *Journal of Computational Neuroscience* **2**: 63-82, 1995.

Douglas, R., Koch, C., Mahowald, M., Martin, K. and Suarez, H. Recurrent excitation in neocortical circuits. *Science* **269**: 981-985, 1995.

Usher, M. and Niebur, E. A neural model for parallel, expectation-driven attention for objects. *Journal of Cognitive Neuroscience*, in press.

Stemmler, M., Usher, M. and Niebur, E. Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics. *Science*, in press.

Books or book chapters published

Schuster, H. and Koch, C.. Burst synchronization without frequency-locking in a completely solvable network model. In: *Advances in Neural Information Processing Systems 4*, Moody, J.E., Hanson, S.J. and Lippman, R., eds., pp. 117-124, Morgan Kaufmann, San Mateo, 1992.

Niebur, E. and Olshausen, B. Neurophysiology, Psychophysics and Computational Models of Visual Attention. In: *Advances in Neural Information Processing Systems 6*, Hanson, S.J., Cowan, J. and Giles, L., pp. 1167-1168, Morgan Kaufmann, San Mateo, 1994.

Stemmler, M., Usher, M., Koch, C. and Olami, Z. Synchronization, oscillations and 1/f noise in networks of spiking neurons. In: *Advances in*

Neural Information Processing Systems 6, Hanson, S.J. , Cowan, J. and Giles, L., pp. 629-638, Morgan Kaufmann, San Mateo, 1994.

Niebur, E. and Brett, D.W. Efficient Simulation of Biological Neural Networks on Massively Parallel Supercomputers with Hypercube Architecture. In: *Advances in Neural Information Processing Systems 6*, Hanson, S.J. , Cowan, J. and Giles, L., pp. 904-910, Morgan Kaufmann, San Mateo, 1994.

Niebur, E. and Koch, C. Models for the neurophysiological basis of attention based on the temporal structure of neural signals. In: *Oscillatory event related brain dynamics*. Elbert, Th., Lütkenhöner, B. and Pantev, C., eds., pp. 295-309. Plenum Press: London, 1994.

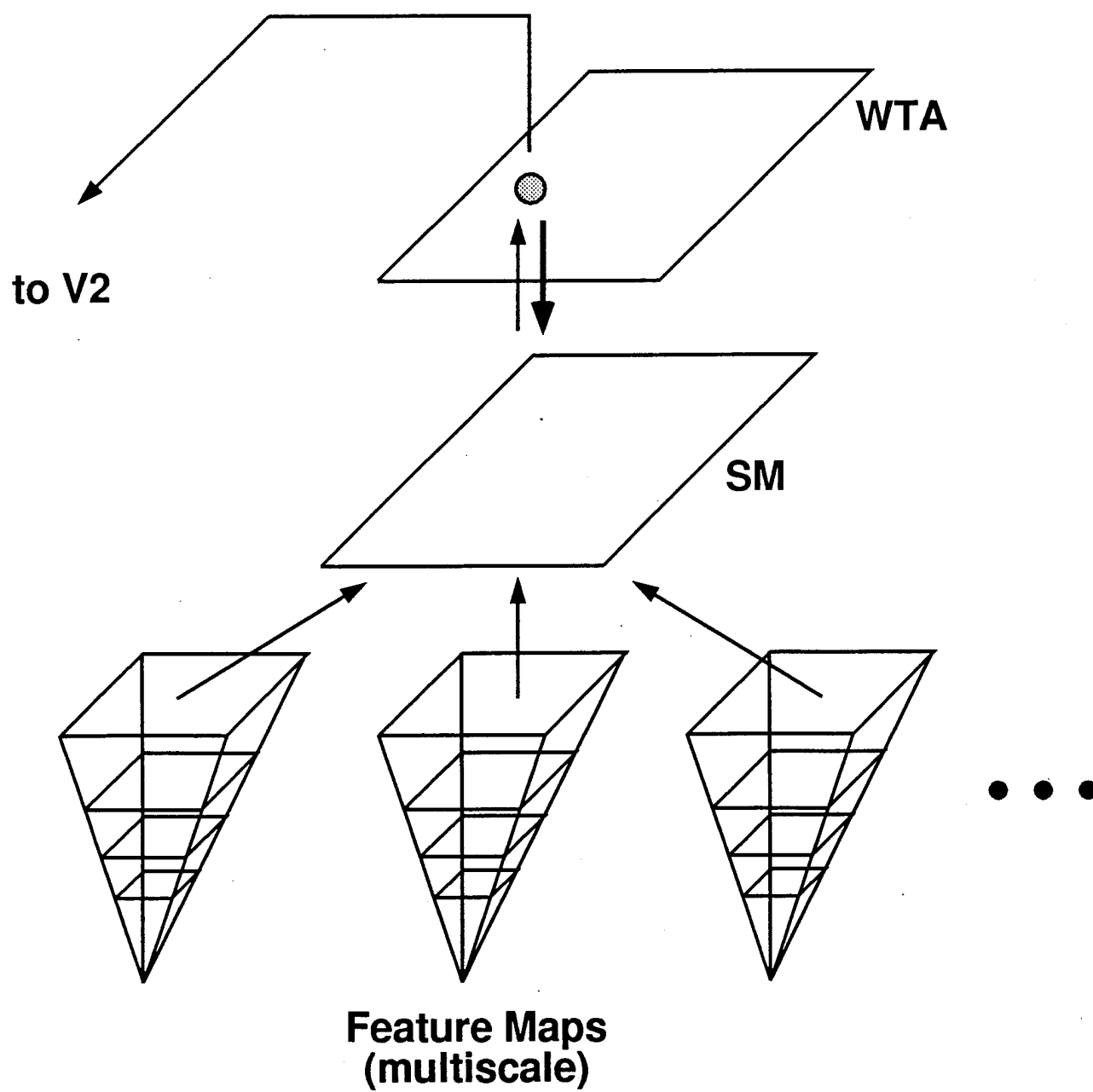
Usher, M. and Niebur, E. Stimulus-dependent synchronization in a model of orientation-tuned spiking neurons. In: *Computation and Neural Systems*, Bower, J., eds., pp. 341-346, Volume 3. Kluwer: Norwell, MA, 1995.

Niebur, E. and Koch, C. Computational architectures for attention. In: *The Attentive Brain*. Raruraman, R., ed., MIT Press: Cambridge, MA, in press.

Technical Reports

Usher, M. and Niebur, E. A Neural Model for Top-Down Selective Attention. *Carnegie-Mellon University Technical Memo*. No. PDP.CNS.94.6, Pittsburgh, PA, 1994.

Niebur, E. and Koch, C. Modeling the "Where" visual pathway. In: *Proc. 2. Joint Symp. on Neural Computation, Caltech-UCSD*. Sejnowski, T. ed., 26-35. Volume 5, Institute of Neural Computation, La Jolla, 1995.



400ms Of Attention

