SAND REPORT

SAND2003-3788
Unlimited Release
Printed January 2003

Final Report of LDRD Project Number 34693 "Building Conscious Machines Based Upon the Architecture of Visual Cortex in the Primate Brain"

Christof Koch

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

Sand2002-3788
Unlimited Release
Printed January 2003

Final Report of LDRD Project Number 34693
"Building Conscious Machines Based Upon the Architecture of Visual Cortex in the Primate Brain"

Prof. Christof Koch
California institute of Technology
Division of Biology
139-74 Caltech
Pasadena, CA 91125

Abstract

Our research plan is two-fold: first, we have extended our biological model of bottom-up visual attention with several recently characterized cortical interactions that are known to be responsible for human performance in certain visual tasks, and second, we have used an eyetracking system for collecting human eye movement data, from which we can calibrate the new additions to the model. We acquired an infrared video eyetracking system, which we are using to record observers' eye position with high temporal (120Hz) and spatial (+/- 0.25 deg visual angle) accuracy. We collected eye movement scan paths from observers as they view computer-generated fractals, rural and urban outdoor scenes, and overhead satellite imagery. We found that, with very high statistical significance (10 to12 z-scores), the saliency model accurately predicts locations that human observers will find interesting. We adopted our model of short-range interactions among overlapping spatial orientation channels to better predict bottom-up stimulus-driven attention in humans. This enhanced model is even more accurate in its predictions of human observers' eye movements. We are currently incorporating biologically plausible long-range interactions among orientation channels, which will aid in the detection of elongated contours such as rivers, roads, airstrips, and other man-made structures.

Final Report of LDRD project number 34693

"Building Conscious Machines Based Upon the Architecture of Visual Cortex in the Primate Brain"

Prof. Christof Koch
California Institute of Technology
Division of Biology, 139-74, Caltech, Pasadena, California 91125

Summary:

1. We acquired an infrared video eyetracking system, which we are using to record observers' eye position with high temporal (120Hz) and spatial (+/- 0.25 deg visual angle) accuracy.

2. We collected eye movement scan paths from observers as they view computer-generated fractals, rural and urban outdoor scenes, and overhead satellite imagery. We found that, with very high statistical significance (10 to 12 z-scores), the saliency model accurately predicts locations that human observers will find interesting.

3. We adopted our model of short-range interactions among overlapping spatial orientation channels to better predict bottom-up stimulus-driven attention in humans. This enhanced model is even more accurate in its predictions of human observers' eye movements.

4. We are currently incorporating biologically plausible long-range interactions among orientation channels, which will aid in the detection of elongated contours such as rivers, roads, airstrips, and other man-made structures.

Background: Our research plan is two-fold: first, we have extended our biological model of bottom-up visual attention with several recently characterized cortical interactions that are known to be responsible for human performance in certain visual tasks, and second, we have used an eyetracking system for collecting human eye movement data, from which we can calibrate the new additions to the model. The work reported here was carried out by Robert Peters, a CNS graduate students, under guidance of Prof. Christof Koch.

Figure 0. Eyetracking system from ISCAN. At left, infrared illuminator and camera; at right, experimenter's rig includes scene monitor, eye monitor, and control computer.

Eyetracking psychophysics: We are using a high-speed (120/240Hz) infrared eyetracking system from ISCAN, Inc. (www.iscaninc.com). This system is now set up (Figure 1) and operational for collecting eye position recordings while subjects view different types of images.

We are using several image databases for psychophysics experiments with the eyetracker (see Figure 2 for sample images). These include both artificially generated images (such as fractals), which have well-characterized low-level image properties, as well as natural images (such as outdoor photos and overhead satellite imagery) that have special domain significance. We are also planning to work with images selected by NIMA for their special relevance, or because the images have been pre-analyzed to help us understand which areas of the image are most important.

Our experimental approach consists of collecting eye movement patterns from a number of subjects while they "free-view" the images for 3-5 seconds at a time. While it is well-known that high-level task demands (such as "look for faces" or "look for agricultural fields") can have a large influence on the patterns of eye movements, at this point we are using a task that is relatively free from top-down biases, so that we can make more direct comparisons with our attention model, which is based on bottom-up processes. It is also possible to limit top-down effects by limiting the duration of image presentation to only a few seconds, since top-down effects are likely to weakest when the image is first presented.

Computational modeling: In previous work we have developed a model of bottom-up attention, which processes an input image in several channels, including intensity, orientation, and color, and combines the output of these channels with a nonlinear interaction to form a saliency map. Figure 2 demonstrates a simple comparison between patterns of human fixations recorded with our eyetracker, and patterns of attentional shifts predicted by our existing bottom-up attention model. In many cases there is a clear qualitative correspondence between the regions identified as salient by both the human observer and the computer model.

The qualitative similarity is strongly supported by a more rigorous quantitative analysis as well. We computed the average predicted salience at locations visited by human observers' scanpaths, and compared these values with the distribution expected by chance. This showed that, with very strong statistical significance (12-14 *z*-scores), the saliency model makes accurate predictions about which locations are likely to be visited by observers' eye movements. This type of comparison allows us to carefully test new extensions to the model, and find the best way to tune the model's parameters for optimum performance in specific tasks. We are working on two specific extensions to the attention model that relate to the processing of orientation information; one relates to interactions among orientation channels at spatially adjacent locations, and the other relates to long-range interactions among orientation channels, which gives rise to contour integration.

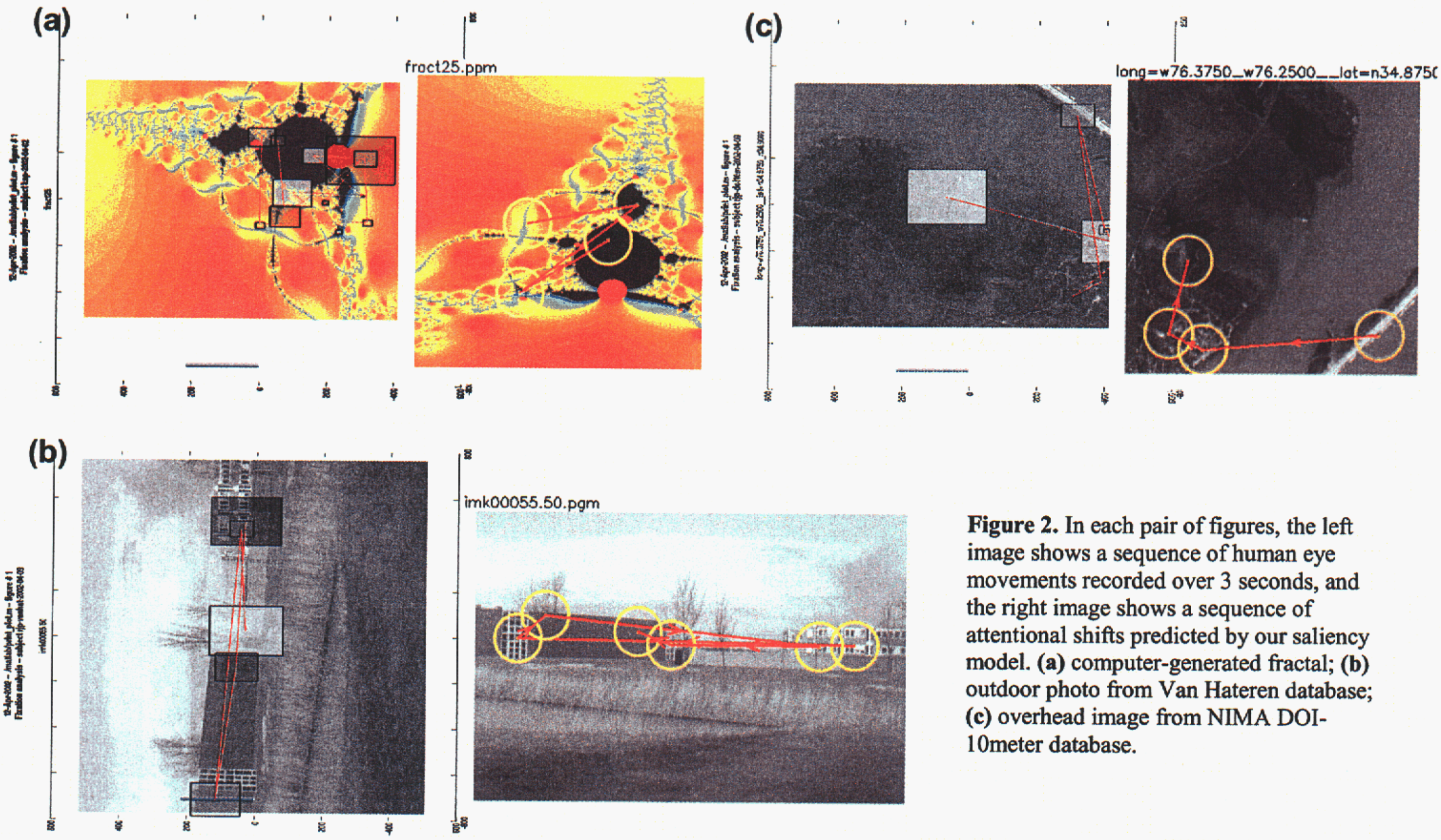

Figure 2. In each pair of figures, the left image shows a sequence of human eye movements recorded over 3 seconds, and the right image shows a sequence of attentional shifts predicted by our saliency model. **(a)** computer-generated fractal; **(b)** outdoor photo from Van Hateren database; **(c)** overhead image from NIMA DOI-10meter database.

Short-range orientation interactions. Previous psychophysics and modeling work by us and others has shown the importance of excitatory and inhibitory interactions among spatially overlapping units tuned to different orientations and spatial scales. In particular, the effect of attention was succinctly captured by a simple change in the strengths of both the excitatory and inhibitory interactions. In practice, these interactions are useful in figure/ground segregation, and in detecting features among noise; Figure 3 shows an example of such effects. We found that when such interactions are included in the saliency model, the model's predictive ability increases significantly (by ~2 *z*-scores).

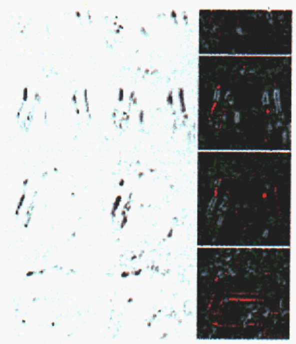

Figure 3. Short-range orientation interactions. Output of orientation channels before (left) and after (center) interactions. At right, the effect of the interactions on activity (green=increase, red=decrease). Note the effect on the central horizontal stripes.

Long-range orientation interactions. We are now introducing excitatory and inhibitory interactions among distant orientation-tuned units. These types of interactions are known to facilitate contour integration—a process in which contours in the visual image lead to neural activity that increases with the smoothness, length, continuity, and closure of the contour. However, it is less well understood how contour integration interacts with other information (such as including local orientation interactions, image contrast, or color contrast) in the computation of a feature-independent saliency map. Our model will be able to address this issue, and answer questions about how learning or training can affect the relative weights of the different visual features.

Distribution:

SPECIFIED DISSEMINATION:

1 - MS-0839 Gerold Yonas, 16000
1 - MS-1153 M.T. Buttram, 15330
1 - MS-0188 LDRD Office, 1010
1 - MS-9018 Central Technical Files, 8945-1
2 - MS-0899 Technical Library, 9616
1 – MS-0612 Review & Approval Desk, 9612

Prof. Christof Koch
California Institute of Technology
Division of Biology, 139-74
Caltech
Pasadena, California 91125