Natural images
In September 1998, the UK Applied Vision Association held its first meeting on the topic of “Natural Images”. The meeting, which was held at University of Bristol attracted seven submitted papers, together with an invited lecture by David Tolhurst. In discussion, the idea arose that the topics of the papers might be suitable for a Special Issue of Perception. Here is the Special Issue, after a suitable gestation period. There was sustained interest in the topic, with a further meeting in 1999 attracting nine papers, and a third meeting in September 2000, as an adjunct to the British Machine Vision Conference, also attracting nine papers.

Why this interest in natural images? There seem to be several reasons. First, it is clear that the human visual system normally deals with images of greater complexity then are traditionally used in psychophysical or physiological studies. The reason for using simplified stimuli is (a) that they are easier to generate, and (b) that experimental control is easier. However, there has been considerable progress, both in computer technology/software, and also in the underlying mathematics. This means that such complex images can now be used as stimuli in a controlled manner and with a good idea of what the information content of the images actually is. The main aim of this issue is to give the reader an idea of how such experimentation can proceed. This also entails looking at the underlying theoretical issues, and these are dealt with in most of the papers in rather different ways.

There is, of course, a problem. By the time the reader receives this issue the work is somewhat out of date, and in a rapidly changing field recent developments are therefore not covered. It will be apparent, for example, that the majority of scenes in these papers are monochrome and static—which is clearly not the case for real-world scenes. These topics have been aired in later meetings, but the present issue presents a selection of topics which are still very much relevant. Interested readers are encouraged to submit material and to attend the Natural Images Symposium at ECVP 2001, which is the successor to the AVA meeting.

So—how can experiments be carried out in this field? The first question concerns the independent variable: the information content of the images. How can this be succinctly summarised? A common approach to this has been to look at the second-order statistics of the image. Typically, these are represented by the power spectrum and the phase spectrum. Until recently, the power spectrum has been the most popular measure of information content due primarily to its conceptual tractability. However, as some of the present papers demonstrate, this is changing. While the power spectrum remains an important consideration, the phase spectrum is receiving more attention. This phase spectrum is a complicated variable, and to tease out an understanding of what it is, there is an increasing tendency to tease it apart by considering ‘higher-order’ statistics.

The second problem in carrying out research in this field is to know what hypotheses to adopt and how to test them. A way of beginning to tackle this is to suppose that human vision is somehow optimised to encode the critically important information in natural scenes. All that is needed is a method of varying the information content, and a dependent measure which allows an assessment of the efficiency with which an observer can carry out a visual task. The first paper in this issue, by Willmore, Watters, and Tolhurst, provides a good introduction to the issue of what the independent
variable can be. They investigate the correlation between the statistical structure of natural images and the statistical structure of putative codes of cells in area V1. The problem here is to see how an ensemble of such cells can encode the incoming information in an efficient manner. For some time, it has been thought that such efficient coding involves developing a sparse representation of the input in the neuronal firing patterns. Here, “sparse” means that any given stimulus will stimulate a small subset of a large ensemble of cells. A different stimulus will stimulate a different small subset, and so on. The authors find that a sparse representation is strongly influenced by the pre-processing of the image prior to this encoding stage. An issue relevant to this topic, which has been much discussed previously, concerns the mathematics of how to generate a sparse code. The present paper compares two methods of achieving this: a nonlinear model originally proposed by Olshausen and Field (1997) and a common linear statistical procedure known as principal components analysis (PCA). The authors show that, if the ‘retinal’ pre-processing is applied, the difference between these two procedures is relatively small.

The second paper, by Brady and Field, again tackles the issue of coding efficiency. Cells in V1 are known to be tuned to orientation and spatial frequency. How are such cells stimulated by natural scenes? In particular, what are the contrasts of such scenes and how do the responses of the cells encode contrast information? The authors revisit an idea, originally proposed by Laughlin (1981), that output nonlinearities of cells maximise the information capacity of the neural code. The authors show that a contrast normalisation procedure both reduces the variability of the response, and the shape of the response distribution shifts to a form associated with a more efficient neural code. As in the first paper, this study demonstrates the importance of pre-processing the image before ‘cortical’ modelling is attempted.

The third paper, by Thomson, Foster, and Summers, introduces the topic of higher-order statistics. The challenge is to find a correspondence between the highly complex properties of the phase spectrum and responses in human vision. The authors tackle this problem by randomisation and quantisation of the phase information. The aims here are to find a simple global image statistic which is sensitive to image phase spectra; and to assess the ability of these statistics to account for human sensitivity to phase perturbations. The paper provides both a demonstration of these effects and an experimental study with phase-perturbed stimuli.

As highlighted by the previous paper, people often talk about the Fourier representation of stimuli. The next paper, by Schofield, defines second-order (non-Fourier) stimuli as those requiring nonlinear processing and asks the question what such second-order vision ‘sees in an image’. The suggestion is that nonlinear mechanisms provide certain types of enhancement to image representation, particularly for providing input to texture segmentation tasks.

Having considered the neurophysiological and computational issues, we now turn to some experimental papers. The next two papers offer variants on the general theme whether vision is optimised to the properties of natural scenes, with psychophysical techniques used to measure some kind of discrimination in naturalistic stimuli. In the first of these papers, Tolhurst and Tadmor use a ‘spectral blending’ technique to measure scene discriminability. The parameters by which these scenes are made more or less ‘natural’ is the magnitude of the spectral slope. The spectral slope is obtained by plotting log Fourier amplitude (averaged across all orientations) against log spatial frequency. For many scenes, such a plot gives a straight line with a slope of about $-1.2$. Changing that slope makes the scenes less natural. Tolhurst and Tadmor show that discriminability of images is best when the spectral slope has its ‘natural’ value, thus providing experimental evidence for optimality.
In the next paper, Párraga and Tolhurst look at the question how these spectral slopes may be discriminated. It is possible to imagine this being done either by comparing the outputs of individual channels which respond to given spatial-frequency bands, or by comparing contrast across more than one band. For general pictures, the authors’ results suggest that the second explanation holds true. However, this was found to be changed by modifying the image so that it appears ‘whitened’ or edge-enhanced. This is a process of normalising to the local luminance in that part of the scene. The implication is that, if a model is to be developed in which responses of individual filters are compared, this model should run on whitened, rather than original, images.

So far, we have been assuming that vision is foveal. This is, of course, not true in general. The next paper, by Melmoth, Kukkonen, Mäkelä, and Rovamo, takes as its task the detection of distortion of a face or grating stimulus, and measures how this varies with eccentric fixation. The findings suggested that the same metric accounts for performance in both tasks, even though the stimuli are very different. This is taken to indicate that task simplicity, rather than the type of stimulus used, determines the type of scaling metric.

The final paper takes a rather different approach to natural images. So far, the experimenters have measured discrimination or detection of stimuli. However, our visual system is used to tell us what is where in a scene. Recently, systems using artificial neural networks (ANNs) have been developed to classify object types in a scene into categories such as ‘vehicle’, ‘vegetation’, etc. Do these systems work in a way similar to human vision? In either an ANN or in human vision, how important are certain kinds of information, such as colour, in performing such classification tasks?

In this paper, Clark, Troscianko, Campbell, and Thomas try to find out whether ANNs and humans use the same cues. In the process, they have had to develop a psychophysical scene-categorisation task that is as similar as possible to the task which the ANN is doing. The solution is to present individual segments of the scene, rather like small pieces of a jigsaw, to an observer and ask the observer to indicate what category these pieces belong to. This is a novel method of performing psychophysics on natural stimuli, and the results point to a similarity in processing between the ANN and the human visual system.

In general, then, the challenges in natural-scene research seem to be how to understand the relationship between scene structure and physiology, and to develop experiments which tap into these processes while retaining the rigour that one expects in vision science. The papers in this issue represent various attempts to come to grips with these issues. Time will tell how successful these are. The AVA meeting has certainly expanded in the years since these papers were presented and, in August 2001, will be a Symposium within the European Conference on Visual Perception. Natural images are increasingly being used in both theoretical and experimental studies in vision science, and such increased usage may lead to more links with other areas of image science such as computer graphics or machine vision.

Ian Moorhead, Tom Troscianko

References
Olshausen B A, Field D J, 1997 “Sparse coding with an overcomplete basis set: a strategy employed by VI” Vision Research 37 3311–3325
Laughlin S B, 1981 “A simple coding procedure enhances a neuron’s information capacity” Zeitschrift für Naturforschung C 36 910–912