—————— NEWS AND VIEWS ——————

Perception

Vision in humans and computers

from Oliver Braddick

HUMANS extract information about objects and events from optical images with effortless efficiency. It is surprisingly difficult to specify exact algorithms that can do the same, but this is the task that the field of 'computational vision' sets itself. A recent symposium* reviewed advances in this area and their impact on our understanding of human vision.

The interaction between computational and experimental studies of vision is now taking place over a wide range of topics, albeit in rather different ways. In colour and motion perception, detailed modelling of initial detection processes has been well established for some time and is closely integrated with quantitative experimental work. For motion detection, G. Sperling (New York University) discussed the elaborated Reichardt model, which multiplies filtered versions of the signal from adjacent regions, whereas E. Adelson (RCA Laboratories) argued for a model that combined signals from receptive fields in quadrature phase to derive the 'spatio-temporal energy' present in a particular direction of motion. Both agreed that these models are, at least potentially, formally equivalent in their input–output relations, but that one or other would turn out more appropriate as a description of the biological system.

It was striking that, in motion as in other topics, the neurophysiological work now attracting most computational interest is not that on the primary visual cortex (V1), but on the multiplicity of higher cortical visual areas. The combination of motion signals from different parts of a contour is necessary to provide unambiguous directional information (E. Hildreth, MIT), and this may be a function of cortical area MT, known to be specialized for motion processing (J.A. Movshon, New York University). Cells in this same areas respond to differences in motion between the receptive field and the surrounding region (J. Allman, California Institute of Technology), which may be important for segmentation of the visual scene.

In colour as well as motion, broader computational issues arise from how initial detector measurements are combined to yield information more closely related to properties of objects in the outside world. An important goal of such combination is to disentangle information about surface properties from the effects of the illuminating light (colour constancy). As in many computational analyses, a key element is to discover the

Computational Models in Human Vision Center for Visual Science, University of Rochester, 19–21 June 1986.

physical constraints which the visual system can exploit to limit an otherwise intractable problem. Although both illuminants and surface reflectances can in principle have an infinite variety of spectral functions, in the real world 99 per cent of the variance in illuminants can be described by three variables, and a similar reduction is possible for the description of

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

a–c, Fractal brownian surfaces can be constructed by recursively adding smaller and smaller 'steps' to a smooth surface. The fractal dimension of a surface depends on the ratio between the altitude of the larger 'bumps' to that of the small bumps. *d*, Six different ratios, giving fractal dimensions between 2.0 and 2.5. Increasing fractal dimension corresponds very well to the perceptual notion of 'roughness'. (Courtesy of Alex Pentland.)

surfaces (B. Wandell, Stanford University). The problem of colour constancy still exists, for these six variables have to be computed from the three signals of trichromatic colour vision, and the process requires integration of information from surfaces at different locations. The neural basis for this process may be in cortical area V4 (S. Zeki, University College, London).

Area V4 is probably not exclusively concerned with colour processing. Neural responses in this area are modulated by the monkey's selective attention (R. Desimone, NIMH, Bethesda; J. Maunsell, University of Rochester). Cells in V4 have receptive fields large enough to contain several stimuli, but Desimone finds that attention to one of these narrows the field to that region only. Interestingly, this does not act to suppress the response of a cell entirely when its field contains only the single unattended stimulus; the effect seems to be intrinsically competitive. Work on selective attention might seem to be a long way from problems of purely visual computation, but S. Ullman (MIT) pointed out the important connection. Certain apparently simple visual properties (for instance, whether two marks lie on the same

continuous contour) cannot be readily computed by uniform parallel operations over the visual field but are handled much more naturally by serial 'visual routines' built up out of a few basic operations. One of the most important of these operations is 'indexing', or spotlighting a particular location within the visual field as the site of the current operation. Ullman's group is now gathering response-time data, which allow experimental tests of the sequences of basic operations used by human perceivers making simple spatial judgments.

The example of selective attention shows that computational models are certainly not restricted to the lower levels of visual processing, but can now be applied to problems such as texture, shape and space. Many visual textures can be modelled by brownian fractals (see figure), in which an exponent (the fractal dimension) relates the amplitude of features to their scale (A. Pentland, SRI International, Menlo Park). Pentland suggested that the visual system extracts this exponent by detecting linear trends in the spatial frequency spectrum, and that regions which differ in this exponent can be perceptually segmented. These interesting ideas have not yet been tested in experimental studies. The same appears to be true of work on how object shapes are subdivided into parts (D. Hoffman, University of California, Irvine). Hoffman's computational model, which locates part boundaries at negative extrema along lines of greatest surface curvature, seems to fit in many cases with intuition, but we lack psychophysical methods for systematically locating the perceptual boundaries of object parts. On the other hand, the novel work of R. Andersen (Salk Institute), on how neurones in a parietal visual area jointly represent retinotopic location and direction of gaze, highlights an issue of longstanding psychophysical and neurophysiological interest that has received little computational analysis.

The subjects of the symposium were diverse, but several themes recurred. Images contain information at various scales, and can often be dealt with efficiently by a 'pyramid' of scale-invariant processes whose capacity increases proportionably with the greater detail of the finer scales of analysis. Segmentation is an essential process, but one which must preserve the hierarchical relationship between a whole object and its component parts. Computational thinking is now intimately embedded in the study of the early processes of vision; in pursuing these principles and their implications it is making increasing contact with our rapidly expanding knowledge of the neural basis of higher visual processes. □

Oliver Braddick is in the Department of Experimental Psychology, University of Cambridge, Cambridge CB3 2EB, UK.