Some task and signal dependent rules for spatial vision

TERRY CAELLI¹ and M. NAMIK OGUZTORELI²

Departments of Psychology¹ and Mathematics², The University of Alberta, Edmonton, AB, T6G 2E9, Canada, and Institute for Medical Psychology, University of Munich, Munich, FRG

Received 7 April 1986; revised and accepted 11 August 1987

Abstract—In this paper we consider the types of computational processes which may be involved in solving a variety of perceptual problems from the detection of signals in the presence of others, to texture discrimination, and some aspects of pattern recognition. These processes centre around the involvement of correlational computations, the transduction of their input/output values, and the apparent involvement of selective filtering mechanisms. Our results suggest that even if fixed detectors (in tuning characteristics) are involved in low-level vision, the human observer apparently employs much more adaptive (variable tuning characteristic) filters and nonlinear mechanisms in more complex spatial tasks.

INTRODUCTION

One of the more central problems for spatial vision research is that of determining the ways in which the input signals are 'coded' along the visual pathways and how such codes are utilized to solve specific perceptual problems. Whether it be with the study of the responses of individual neurones to selected signals, or the investigation of human responses to relatively simple light sources, the prevailing epistemology has been that 'breaking the spatial code' is accomplished by describing a finite set of fixed feature analysers which are consistent with behaviour on a given task and a restricted set of signals.

Even limiting attention to achromatic (grey-scaled) images, a number of different 'coding models' for spatial vision have been proposed over the past 40 y—and precisely developed for the analysis of specific signal types. Examples include:

(a) Line, edge, intersection angle detectors for edge or line images (Attneave and Arnoult, 1956); blurred edges (Watt and Morgan, 1984); bars and slits (Hubel and Wiesel, 1962, 1965);
(b) Autocorrelation and nth-order statistics for textures (Julesz, 1962, 1981) and dot patterns (Caelli et al., 1978; Uttal, 1983);
(c) Spatial frequencies (channels or filters) for apertured gratings (Schade, 1956; Daugman, 1980; De Valois and De Valois, 1980).

Of particular popularity, over the past two decades, has been Fourier transform-based definitions of visual detectors or 'channels' (Schade, 1956; Kabrisky, 1966; De Valois and De Valois, 1980) where the visual system is proposed to directly code elementary spatial features by filtering the input signal with sets of parallel acting two-dimensional detectors whose responses are depicted by modulations of the image power (energy) spectrum.
In these examples of spatial coding theories the underlying epistemology entails the assumption that a singular type of code embodies the key to understanding visual function and that, once defined for a specific paradigm, provides the format for predicting a variety of spatial information processing behaviours. The fact remains, however, that this approach has never been successful since counterexamples have always been generated whether it be with texture codes (Caelli et al., 1978), or spatial-frequency energy coding models (Caelli and Moraglia, 1987a,b) as only two examples in psychophysics. In the study of individual cell neural activities, similar problems exist. That is, the determination of a given cell’s receptive field profile is gauged from the signal parameter states which elicit the best response. This assumes that the relationship between the receptive field and the input signal is one of matching, or cross-correlation, whereby the peak of the matching process occurs when signal and the receptive field profiles are equal, up to a scalar multiple (the Cauchy-Schwarz inequality). Further, the same cell obviously has many different receptive field profiles as a function of the signal being processed. Since the causal relationship(s) between receptive field profiles and underlying axo-dendritic activities have not been established, those above correlational and multivariate response properties can only be implied from consistent behaviours. It is for this reason that neurophysiological bases for sensory coding are as phenomenological as psychophysical ones.

To pre-empt slightly, the approach proposed here is not based on the existence of special image codes (parameters) but, rather, on the enumeration of processing rules which generate specific signal decompositions (channels or filters) or coding strategies according to the perceptual problem to be solved. That is, image frequencies, orientations, pixel histograms, grey-level quantizations, edges, etc., all may be selectively employed in different visual problems—yet none actually provide the ‘coding rules’ for spatial vision. The aim of this paper is to explore some of these rules in a number of different perceptual contexts, including masking, texture discrimination and pattern-recognition tasks.

An adaptive information processing system involves the use of filters or detectors whose parameter states vary as a function of the signal and task demands. This is to be contrasted with a classical control system (Ogata, 1970) where all detector states, transmission characteristics, are fixed for all signals and tasks. It is usually the case that both types of control systems are capable of simulating the same behaviour. However, adaptive control systems typically attain equal performance using lower numbers of processes, filters, etc. (Ogata, 1970). For these reasons, this paper is more concerned with presenting another interpretation of current observations and extending their application, rather than clear counter-examples. However, the underlying mechanisms of adaptive control systems are quite different from classical control. Similarly, full computational solutions to the types of perceptual problems of interest here involve more processes than the enumeration of filters or detector profiles.

VISUAL DETECTORS AND THE NOTION OF PERCEPTUAL EQUIVALENCE CLASSES

Whether it be the processing of spatio-temporal information in motion perception (Adelson and Bergen, 1985), texture discrimination (Harvey and Gervais, 1978; Caelli, 1982), or the discrimination of gratings (Wilson and Gelb, 1984), the fixed, invariant detector theory proposes that the perceived luminance value of an input image is determined by the (nonlinear) sum of the outputs of $n$ detectors. That is, the perceptual