Computational approaches to human pattern recognition

TERRY CAELLI1,* and WALTER F. BISCHOF2

1 Department of Computer Science, Curtin University of Technology, GPO Box U1987, Perth, WA 6001, Australia
2 Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Received for publication 11 October 1993

Abstract—This paper consolidates recent findings on how humans detect and recognize patterns and considers computational procedures which reflect observed performance. A multi-level correlation model for spatial information processing is proposed and used to interpret past results on human psychophysical performance.

1. INTRODUCTION

The aim of this paper is to consolidate research completed over the past decade into how humans detect and recognize forms or shapes, and to present an overview of the underlying processes which capture various aspects of observed behaviour. Different types of recognition problems are considered and it is shown that a common set of computational procedures seems to underpin the known performance of the human visual system in interpreting images. However, before dealing with these issues in some detail, some consideration of what is understood by 'form' and 'shape' is necessary.

Pattern, form or structure are quite difficult to define in a succinct way. But one necessary condition for the presence of a structure in a signal is the existence of, at least, some degree of correlation within the signal. A signal which is perfectly uncorrelated in all dimensions (that is, white noise) has, by definition, no structure. Here, 'uncorrelated' means that we cannot predict the intensity of any pixel from any set of other pixels and so structures in textures, scenes or patterns are determined by the types of correlations—or pixel dependencies—present. Indeed, the study of these correlations—and the techniques to represent them formally—has been, for a long time, the main focus of both texture and form analysis (see Julesz, 1962).

An important example of how the existence of spatial correlations define patterns is the notion of an 'edge': the locus of the points corresponding to the boundary of patterns. They are the pixels whose intensities cannot be predicted from their neighbours. This can be clearly seen in the error image resultant from linear predictive coding (LPC, see Rosenfeld and Kak, 1982) and can also be applied to spatio-chromatic domains (Caelli and Reye, 1993). One of the early papers which proposed this connection between correlation and human form perception was that of Dodwell (1970) who argued that eye movements provide the basis for the autocorrelation of image data. This theme of autocorrelation was further developed by Uttal (1985), and the aim of this paper is to show how

* To whom reprint requests should be addressed.
Correlation has played a critical role in even defining the processing of spatial information in the vertebrate visual system, in particular for the notion of a 'receptive field profile' (RFP) that is used so often to model biological information processing. Neurophysiology and psychophysics assume the 'principle of maximum signal response' to measure what is being processed. That is, the response of a cell to different parametric states of a signal (for example, orientation) defines the neurone's RFP and the maximum response determines the underlying feature detector (see, for example, Hubel and Wiesel, 1968). All this assumes that analogues to such detectors exist within the visual system and that the response is reasonably modelled by the system correlating the internal detector with the external signal. Without this assumption it would not be possible to conclude anything like what has been concluded from the electrophysiology of vision over the past 50 years.

The reverse is assumed to occur in masking studies. That is, the very notion of 'fatiguing' spatially tuned neurones in the human visual cortex by repeated exposure to a given signal also assumes that the maximum masking effect will be induced by the signal which 'best matches' the given detector's profile. Again, correlation defines this match. In fact, in a direct study of this, it was possible to show that we could not reject the hypothesis that the masking effect was determined by the cross-correlation between signal and mask (Caelli and Moraglia, 1987a). The experiments used Gabor signals (Gaussian modulated sinusoidal gratings) which could be decorrelated by changes in frequency, orientation or phase between the masking and test patterns. A forward-masking task was used where a test signal was presented after a masking signal and subjects were required to indicate whether the test was present or not. The percentage correct detection was clearly predictable from the inverse of the peak of the cross-correlation between signal and mask. Notice that this cross-correlation was not 'in-place': the masking could occur within a neighbourhood of the test centre. These results argued for an adaptive channel model where the channel centre and bandwidth are determined by the correlation between signal and mask.

Such results have also been duplicated with more natural scenes (Caelli and Moraglia, 1987b) where the masking patterns had either the same or different amplitude or phase spectra as the test patterns. Here it was shown that similarity in power spectra was unrelated to the degree of masking (the power spectra of two images were identical or quite different) and that the correlation between the actual images (in the space domain) was the deciding factor.

This discussion demonstrates that there is some basis for the involvement of correlation in early visual encoding and broad support for channels—neuronal subsystems selectively sensitive to specific signal correlations—in biological vision. What is required, however, is an analysis of the extent to which such a computational procedure applies to higher-order or more natural visual tasks.

Image correlations—and lack thereof, as measured by edge coding—occur at many different scales and colour bands, as is illustrated in Fig. 1 using different isotropic filters (from Caelli and Reye, 1993). Here we have used standard zero-crossings of the $\nabla^2 G$ (Gaussian low-pass-filter followed by the Laplacian differential operator) with two different Gaussians to obtain the results over